

باسمه تعالی



مجموعه مقالات فارسی

پنجمین سمینار

نظریه مفصل و کاربردهای آن

گروه آمار، دانشگاه فردوسی مشهد

با همکاری

قطب علمی داده‌های ترتیبی و فضایی

۱۰ و ۱۱ بهمن ماه ۱۳۹۷

عنوان: مجموعه مقالات فارسی پنجمین سمینار نظریه مفصل و کاربردهای آن

تدوین کننده: محمد امینی

صفحه آرا: عادل احمدی نادی و مرتضی محمدی

طرح جلد: حسینعلی محتشمی برزادران

ناشر: کمیته برگزاری سمینار

شمارگان: ۱۲۰

تاریخ انتشار: اسفند ۱۳۹۷

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 برنامه های قطب علمی داده های ترتیبی و فضایی می باشد، 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 داده های ترتیبی و فضایی، مرکز اطلاع رسانی و پایگاه استنادی جهان اسلام، همکاری انجمن آمار ایران، و انجمن بیمه و مالی با هدف فراهم نمودن امکانات مناسب برای تبادل اطلاعات پژوهش های اعضای هیأت علمی، دانشجویان تحصیلات تکمیلی و کارشناسان و کاربران در محورهای سمینار، روزهای چهارشنبه و پنجشنبه ۱۰ و ۱۱ بهمن ماه ۱۳۹۷ در دانشگاه فردوسی مشهد برگزار می شود. امیدواریم که تلاش کمیته های مختلف برگزاری سمینار بتواند رضایت خاطر میهمانان و شرکت کنندگان 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 ۲۰ مقاله به صورت ارائه شفاهی و ۷ مقاله به صورت پوستر مورد پذیرش قرار گرفت. لازم به ذکر است که شرکت کنندگان این سمینار ۴۰ نفر از اعضای هیئت علمی و دانشجویان تحصیلات تکمیلی دانشگاه های معتبر کشور هستند. در پایان لازم می دانیم از کمیته های 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، قطب علمی داده های ترتیبی و فضایی، 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 شرکت کنندگان محترم در این سمینار را داریم. امیدواریم اقامت دلپذیری در شهر مشهد داشته باشند.

محمد امینی (دبیر)

بهمن ۱۳۹۷

محورهای سمینار:

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

اعضای کمیته علمی (به ترتیب حروف الفبا):

۱. دکتر محمد امینی، دانشگاه فردوسی مشهد (دبیر سمینار)
۲. دکتر غلامعلی پرهام، دانشگاه شهید چمران اهواز
۳. دکتر هادی جبّاری نوقابی، دانشگاه فردوسی مشهد (دبیر اجرایی سمینار)
۴. دکتر امین حسن‌زاده، دانشگاه شهید بهشتی
۵. دکتر سلیمان خیری، دانشگاه علوم پزشکی شهرکرد
۶. دکتر علی دولتی، دانشگاه یزد
۷. دکتر محمد ذونعمت کرمانی، دانشگاه شهید باهنر کرمان
۸. دکتر صدیقه شمس، دانشگاه الزهرا
۹. دکتر زهره شیشه‌بر، دانشگاه شیراز
۱۰. دکتر محمد حسین علامت‌ساز، دانشگاه اصفهان
۱۱. دکتر غلامرضا محتشمی برزادران، دانشگاه فردوسی مشهد
۱۲. دکتر محسن محمدزاده، دانشگاه تربیت مدرس
۱۳. دکتر سید محسن میرحسینی، دانشگاه یزد

اعضای کمیته برگزار کننده (به ترتیب حروف الفبا):

۱. دکتر محمد امینی، دانشگاه فردوسی مشهد (دبیر سمینار)
۲. دکتر جعفر احمدی، دانشگاه فردوسی مشهد
۳. دکتر هادی جباری نوقابی، دانشگاه فردوسی مشهد (دبیر اجرایی سمینار)
۴. دکتر مهدی جباری نوقابی، دانشگاه فردوسی مشهد
۵. دکتر مهدی دوست پرست، دانشگاه فردوسی مشهد
۶. دکتر علی دولتی، دانشگاه یزد
۷. دکتر بهرام صادقیپور گیلده، دانشگاه فردوسی مشهد (مدیر گروه آمار)
۹. دکتر وحید فکور، دانشگاه فردوسی مشهد
۱۰. دکتر غلامرضا محتشمی برزادران، دانشگاه فردوسی مشهد

کادر اجرایی سمینار (به ترتیب حروف الفبا):

۱. عادل احمدی نادى، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۲. مجتبی اصفهانی، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۳. احسان امینی، دانشجوی کارشناسی مهندسی کامپیوتر، دانشگاه فردوسی مشهد
۴. نازنین حوتی، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۵. ندا دودمان، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۶. جابر کاظم پور، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد
۷. حسینعلی محتشمی برزادران، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد (مسئول دبیرخانه)
۸. مرتضی محمدی، دانشجوی دکتری آمار، دانشگاه فردوسی مشهد

فهرست مقالات فارسی

تحلیل اندازه‌های ریسک رقابتی وابسته با ساختار وابستگی <i>FGM</i> تعمیم‌یافته	
توکلی، پ.، امینی، م.، محتشمی، غ.	۱۰
نتایج درباره‌ی معیار نادرستی در نظریه‌ی اطلاع براساس تابع مفصل	
حسینی، ت.، احمدی، ج.	۲۲
ارزش در معرض ریسک شرطی برای مفصل‌های با وابستگی دهی	
سالاری‌نژاد، ز.، میرحسینی، س.م.	۳۴
مدل‌بندی وابستگی در داده‌های طولی نامتعادل با استفاده از مفصل‌های جفتی	
سفیدی، س.، گنجعلی، م.، باغفلکی، ت.	۴۲
تعیین اندازه ریسک در سهام شرکت‌های بیمه بوسیله تابع مفصل	
سیف‌اللهی، س.، بیورانی، ح.	۵۰
انتخاب مفصل با استفاده از معیارهای نظریه اطلاع	
شیبانی، ا.، دهقان، م.ح.، احمدزاده، ح.	۶۱
ساخت مفصل دومتغیره با وابستگی دهی و کاربرد آن در مدل‌بندی داده‌های هزینه و درآمد	
عسگری، پ.، میرحسینی، س.م.	۷۴
مفصل‌هایی برای ساخت حرکت براونی دو بعدی	
بیک‌مرادی، ف.، دولتی، ع.، زمانی، م.ص.	۸۲

تحلیل دو متغیره خشکسالی با استفاده از شاخص SEPI بر مبنای توابع مفصل

وحدتی، ر.، داوری، ک.، قهرمان، ب.، رضایی‌پژند، ح. ۸۹

مدل سازی وابستگی ریسک‌های بیمه‌گری توسط کاپولا و کاربرد آن در محاسبه توانگری مالی

آراء، ش.، ملک‌پور، س.ا.، جناتی کاشانی، ر. ۱۰۱



تحلیل اندازه‌های ریسک رقابتی وابسته با ساختار وابستگی FGM تعمیم یافته

توکلی، پ^۱ امینی، م^۲ محتشمی، غ^۳

گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده

یک واحد را در نظر بگیرید که بتواند هر یک از k نوع شکست رقابتی را تجربه کند. فرض کنید که برای هر واحد زمان شکست T و نوع یا علت شکست $C \in \{1, 2, \dots, k\}$ را مشاهده کنیم، در این حالت زوج (T, C) که در متون علمی آماری ریسک‌های رقابتی نامیده می‌شوند، مشاهده می‌کنیم. اندازه‌های وابستگی مفید را برای تحلیل مدل ریسک‌های رقابتی دومتغیره و همچنین توابع زیر توزیع را تحت تعمیم‌های مفصل FGM که شامل تعمیم بایراموف و کاتز (۲۰۰۲)، لی و ژی (۲۰۰۰) است با حاشیه‌های بر نوع ۳ به دست آورده و اندازه‌گیری قابلیت اطمینان را برای این تعمیم‌ها ارائه می‌دهیم.

کلمات کلیدی: ریسک رقابتی، تابع زیر توزیع، مفصل FGM تعمیم یافته، اندازه قابلیت اطمینان

^۱ parisa.tavakoli@mail.um.ac.ir

^۲ m-amini@um.ac.ir

^۳ grmohtashami@um.ac.ir

۱ پیش‌گفتار

فرض کنید که هر یک از واحدهای مورد مطالعه از چندین نوع شکست مشخص شوند و برای هر واحد زمان شکست T و نوع شکست $C \in \{1, 2, \dots, k\}$ را مشاهده کنیم، شکست در این جا ممکن است به عنوان مثال تجزیه یک قطعه مکانیکی باشد که در آن چندین علت احتمالی اصلی برای شکست وجود دارد. مانند: ارتعاش، همبستگی و غیره. این یک نمونه معمولی از وضعیت ریسک‌های رقابتی در قابلیت اطمینان است و می‌توان آن را به تلاش‌های دیوید برنولی در سال (۱۷۶۰) برای ریشه‌کن کردن خطر مرگ از ابتلا به عفونت از سایر علل دیگر دانست. این در واقع یک نمونه کلاسیک از ریسک‌های رقابتی است که افراد در معرفی چندین علت مرگ قرار دارند. کاربرد مشابه در جمعیت‌شناسی و علم آماری وجود دارد و معمولاً تحت نام تجزیه و تحلیل کاهش پله‌ای چندگانه رخ می‌دهد. یک مفصل دو متغیره، یک تابع توزیع دو متغیره از دو متغیر تصادفی با توزیع یونیفرم در بازه $[0, 1]$ می‌باشد (نلسن، ۲۰۰۶). با استفاده از مفصل، می‌توان ساختار وابستگی بین دو متغیر تصادفی با توزیع‌های حاشیه‌ای دلخواه را بررسی کرد. مفصل‌های شناخته‌شده از جمله کلیتون، فرانک، جنست، گامبل و جو از خانواده مفصل‌های ارشمیدسی است و مفصل‌های گاوسی و t از خانواده بیضوی هستند. (نلسن-۲۰۰۶) در این مقاله بر مفصل فارلی-گامبل-مورگسترن (FGM) و مفصل FGM تعمیم‌یافته تمرکز می‌کنیم. مفصل مربوط به توزیع FGM توسط مورگسترن (۱۹۵۶) معرفی شده است. از کاربردهای آن، مطالعه روی مبتلایان دیابت است که توسط لوزارا و همکاران (۲۰۱۳) انجام شده و نشان دادند که در میان یک مجموعه از مفصل‌ها، مفصل FGM برای داده‌ها مناسب‌تر است. گامبل (۱۹۶۰b) توزیع FGM را با حاشیه‌های نمایی مورد مطالعه قرار داد و فارلی (۱۹۶۰) برای توابع خاص $A(\cdot)$ و $B(\cdot)$ ، توزیع تعمیم‌یافته FGM را به صورت

$$F(x, y) = F_1(x)F_2(y)[1 + \theta \{1 - A(F_1(x))\} \{1 - B(F_2(y))\}].$$

معرفی کرد. با توجه به خواص ریاضی آن، مفصل تعمیم‌یافته FGM توسط بسیاری از نویسندگان که شامل بایراموف و کاتز (۲۰۰۲)، نلسن (۲۰۰۶)، لی و ژی (۲۰۰۰)، امینی و همکاران (۲۰۱۱) و دوما و گیاردانو (۲۰۱۳، ۲۰۱۶) مورد بررسی قرار گرفته است. با توجه به اینکه راه‌های مختلفی برای تعمیم مفصل FGM وجود دارد، بر تعمیم‌های خاصی از مفصل FGM که توسط بایراموف و کاتز (۲۰۰۲) و لی و ژی (۲۰۰۰) معرفی شده، تمرکز می‌کنیم. در بخش اول، به معرفی ریسک‌های رقابتی و معرفی مدل ریسک‌های رقابتی پرداخته‌ایم. در بخش دوم، اندازه‌های مفید را برای تحلیل مدل‌های ریسک‌های رقابتی دو متغیره و همچنین توابع زیرتوزیع را تحت تعمیم‌های مفصل FGM که شامل تعمیم بایراموف و کاتز (۲۰۰۲)، لی و ژی (۲۰۰۰) است را با حاشیه‌های پر نوع ۳ به دست می‌آوریم و اندازه‌گیری قابلیت اطمینان را برای این تعمیم‌ها ارائه می‌دهیم.

۲ مدل ریسک رقابتی

در ریسک‌های رقابتی، نتیجه مشاهده شده شامل زمان شکست T یا زمان تا رخداد پیشامد مورد نظر و یا سانسورشدگی است. و علت شکست و یا متغیر سانسورشدگی می‌باشد که برابر یک است اگر پیشامد مورد نظر مشاهده شود و در غیر این صورت برابر صفر است. زمان شکست T به‌عنوان متغیر پیوسته در نظر گرفته شده و علت C می‌تواند یک عدد ثابت با مقادیر $j = 1, \dots, k$ باشد. از لحاظ احتمالی، مدل توزیع توأم (T, C) قابل شناسایی است که این را می‌توان مشخص کرد و در اصطلاح توابع زیر توزیع نامیده می‌شوند. فرض کنید (T, C) بردار تصادفی زمان شکست و علت شکست باشد، در این صورت برای هر $t > 0$ و $j = 1, 2, \dots, k$ داریم:

الف. تابع زیرتوزیع:

احتمال حاشیه‌ای شکست از یک علت یا پیشامد خاص را با تابع بروز تجمعی اندازه می‌گیریم. این تابع یکی از پرکاربردترین توابعها در تحلیل داده‌های ریسک‌های رقابتی است. این تابع برای پیشامد نوع j به صورت زیر تعریف می‌شود،

$$F_j(t) = P(T \leq t, C = j).$$

تابع توزیع کلی، احتمال رخ دادن یک پیشامد از هر نوعی یا علتی را تا قبل از زمان t اندازه می‌گیرد و برابر با مجموعه همه $F_j(t)$ برای همه پیشامدهاست.

$$F(t) = P(T \leq t) = \sum_{j=1}^k P(T \leq t, C = j) = \sum_{j=1}^k F_j(t)$$

در غیاب ریسک‌های رقابتی محدوده تابع توزیع کلی فاصله $[0, 1]$ است. در حضور ریسک‌های رقابتی، تابع توزیع کلی حداکثر تا مقدار $p(C = j)$ را می‌گیرد زیرا،

$$\lim F_j(t) = P(C = j) \leq 1 \quad j = 1, \dots, p$$

از این رو واژه توزیع برای آن مناسب نیست و به آن زیرتوزیع گویند.

ب. تابع زیرچگالی:

$$f_j(t) = F_j'(t)$$

پ. تابع زیربقا:

$$\bar{F}_j(t) = P(T > t, C = j) \quad j = 1, \dots, k$$

ت. توزیع حاشیه‌ای متغیر تصادفی C :

$$\Pi_j = P(C = j) = F_j(\infty) = \bar{F}_j(0) \quad j = 1, \dots, k$$

توابع زیرتوزیع و زیربقاء با هم مرتبط هستند زیرا برای همه $j = 1, \dots, k$

$$F_j(t) + \bar{F}_j(t) = P(T \leq t, C = j) + P(T > t, C = j) = P(C = j) = \Pi_j$$

۳ اندازه‌های ریسک رقابتی با ساختار وابستگی FGM تعمیم یافته

یکی از روش‌های مورد استفاده در مدل‌سازی داده‌های بقاء چندمتغیره با معلوم بودن توابع بقاء کناری، استفاده از توابع مفصل می‌باشد که در سال‌های اخیر گسترش یافته است. روش‌های پارامتری و ناپارامتری مختلفی برای مدل‌سازی داده‌های ریسک‌های رقابتی پیشنهاد شده است. اما روش‌های پارامتری به دلیل مشکل شناسایی پذیری کمتر مورد توجه قرار گرفته‌اند. محققین مختلفی استفاده از تابع مفصل را به منظور حل مشکل شناسایی پذیری پیشنهاد کرده‌اند. در این مدل‌ها توزیع‌های حاشیه‌ای به گونه‌ای کاملاً پارامتری انتخاب می‌شوند. زمانی که ساختار وابستگی بین متغیرهای پنهان توسط یک مفصل معروف بیان می‌شود، ژانگ و کلین نشان دادند توزیع‌های حاشیه‌ای قابل شناسایی هستند. آن‌ها حالت خاص $k = 2$ را در نظر گرفتند. پس مفصل T_1 و T_2 را به صورت زیر ارائه دادند.

$$C(u_1, u_2) = H(G_1^{-1}(u_1), G_2^{-1}(u_2)), \quad (u_1, u_2) \in [0, 1] \times [0, 1]$$

که یک تابع توزیع توأم $[0, 1] \times [0, 1]$ با حاشیه‌ای‌های یونیفرم است. ژانگ و کلین (۱۹۵۵) ارائه دادند که اگر مفصل شناخته شده باشد، پس توابع حاشیه‌ای G_1 و G_2 به وضوح توسط تابع زیرتوزیع F_1 و F_2 تعیین می‌شوند. بنابراین قادر به برآورد توزیع‌های حاشیه‌ای از مشاهدات (T, C) هستیم. ساخت توزیع‌های چند متغیره با داشتن حاشیه‌ای‌ها به عنوان یک مسئله مورد توجه آماردانان قرار گرفته است. یکی از مشهورترین توزیع‌های دو متغیره که فارلی-گامبل-مورگنسترن (FGM) نامیده شده است کلاسی از توزیع‌های دو متغیره برای هر زوج از تابع توزیع‌های تجمعی یک متغیره F_1 و F_2 (c.d.f)، به شکل

$$H(x, y) = F_1(x)F_2(y) \{1 + \theta \bar{F}_1(x)\bar{F}_2(y)\} \quad -1 \leq \theta \leq 1$$

تعریف می‌شود که در واقع توسط فارلی (۱۹۶۰)، گامبل (۱۹۶۰) و مورگنسترن (۱۹۵۶) معرفی شده است. از کاربردهای آن، مطالعه بر مبتلایان دیابت است که توسط لوزاداو همکارانش (۲۰۱۳) انجام شده و نشان دادند که در میان یک مجموعه از مفصل‌ها، مفصل FGM برای داده‌ها مناسب‌تر است. مورگنسترن توزیع FGM با حاشیه‌های کوشی را مورد مطالعه قرار داد و این خانواده توسط گامبل برای توابع حاشیه‌ای نمایشی بررسی شد و توسط فارلی تعمیم داده شد که فارلی (۱۹۶۰) برای توابع خاص $A(\cdot)$ و $B(\cdot)$ ، توزیع تعمیم یافته FGM را به صورت

$$F(x, y) = F_1(x)F_2(y)[1 + \theta \{1 - A(F_1(X))\} \{1 - B(F_2(y))\}],$$

معرفی کرد. محدودیت شناخته شده این توزیع، عدم اجازه مدل بندی داده ها با وابستگی زیاد توسط این توزیع است. برای مثال ضرایب وابستگی تاوکنندال و روی اسپیرمن این خانواده به ترتیب در بازه های $[\frac{-2}{9}, \frac{1}{9}]$ ، $[\frac{-1}{3}, \frac{1}{3}]$ قرار دارد. به همین دلیل و شکل تحلیلی ساده این خانواده، تعمیم های متنوعی از این خانواده در متون آماری ارائه شده است، که به عنوان نمونه می توان به تعمیم هایی که توسط بایراموف و کاتز (۲۰۰۲) و لی و ژی (۲۰۰۰) معرفی شده، بپردازیم. بایراموف و کاتز (۲۰۰۲) برای گسترش دامنه ضرایب وابستگی ρ و τ ، مفصل FGM را به صورت زیر تعمیم دادند

$$C(u, v) = uv \{1 + \theta(1 - u^p)^q(1 - v^p)^q\}, \quad p, q \geq 1 \quad (1)$$

مفصل تعمیم یافته FGM با پارامترهای $p, q \geq 1$ به صورت،

$$C(u, v) = uv \{1 + \theta(1 - u^p)^q(1 - v^p)^q\}.$$

که در آن،

$$-\min \left\{ 1, \frac{1}{p^q} \left(\frac{1 + pq}{q - 1} \right)^{q-1} \right\} \leq \theta \leq \frac{1}{p^q} \left(\frac{1 + pq}{q - 1} \right)^{q-1}.$$

و چگالی مفصل به صورت،

$$c(u, v) = 1 + \theta(1 - u^p)^{q-1} \{1 - (1 + pq)u^p\} (1 - v^p)^{q-1} \{1 - (1 + pq)v^p\}.$$

و همچنین مفصل تعمیم یافته FGM که توسط لی و ژی (۲۰۰۰) معرفی شده به صورت

$$C(u, v) = uv + \theta u^q v^q (1 - u)^p (1 - v)^p \quad p, q \geq 1 \quad (2)$$

که در آن

$$-\min \left\{ \frac{1}{[B^+(a, b)]^2}, \frac{1}{[B^-(a, b)]^2} \right\} \leq \theta \leq -\frac{1}{B^+(a, b)B^-(a, b)}$$

در حالت کلی برای هر دو تعمیم ذکر شده زمانی که $p = q = 1$ باشد همان ساختار مفصل FGM را داریم. توزیع بر نوع ۳، متعلق به سیستم بر است که به طور گسترده در برنامه ها مورد استفاده قرار می گیرد. از این توزیع نویسندگان مختلف برای تجزیه و تحلیل داده های طول عمر با ریسک های رقابتی مورد بحث قرار داده اند. که تابع چگالی و تابع توزیع آن به ترتیب به صورت زیر است،

$$f(x) = \alpha \delta x^{-\alpha-1} (1 + x^{-\delta})^{-\alpha-1}$$

و

$$F(x) = (1 + x^{-\delta})^{-\alpha}, \quad x > 0, \alpha > 0, \delta > 0$$

۱.۳ اندازه‌های ریسک رقابتی وابسته

در این بخش، برخی از اندازه‌های مفید را برای تحلیل مدل‌های ریسک‌های رقابتی وابسته تحت مفصل FGM تعمیم یافته به دست می‌آوریم. بر روی اندازه‌های مربوط به توابع زیرتوزیع تمرکز می‌کنیم که نقش اساسی در مدل‌های ریسک رقابتی دارد. فرض کنید X, Y زمان‌های شکست پیوسته، مفصل تعمیم یافته FGM در معادله (۱) و (۲) است. مشاهده می‌کنیم اولین زمان شکست $T = \min(X, Y)$ و شکست از علت ۱ اگر $C = ۱$ یا $X \leq Y$ یا $C = ۲$ اگر $Y < X$ باشد. توابع زیرتوزیع از علت شکست ۱ ($C = ۱$) و علت شکست ۲ ($C = ۲$) به ترتیب به صورت

$$\begin{aligned} F(1, t) &= P(C = 1, T \leq t) = \int_0^t f(1, z) dz, \\ F(2, t) &= P(C = 2, T \leq t) = \int_0^t f(2, z) dz \end{aligned} \quad (3)$$

می‌باشد که در آن،

$$\begin{aligned} f(1, t) &= \frac{-\partial \bar{F}(x, y)}{dx|_{x=y=t}}, \\ f(2, t) &= \frac{-\partial \bar{F}(x, y)}{dy|_{x=y=t}} \end{aligned} \quad (4)$$

توابع زیرچگالی نامیده می‌شوند.

قضیه ۱.۳. فرض کنید (X, Y) یک بردار تصادفی با ساختار وابستگی مفصل FGM تعمیم یافته باشد در این صورت:

(الف) اگر مدل (۱) انتخاب شود آنگاه:

$$\begin{aligned} F(2, t) &= F_2(t) - \int_0^t F_1(z) F_2(z) dz - \theta \sum_{i=0}^q \sum_{j=0}^{q-1} \binom{q}{i} \binom{q-1}{j} (-1)^{i+j} \\ &\times \left\{ \int_0^t F_1(z)^{p_i+1} F_2(z)^{p_j} F_2(z) dz - (1 + pq) \int_0^t F_1(z)^{p_i+1} F_2(z)^{p_j+p} f_2(z) dz \right\} \end{aligned}$$

(ب) اگر مدل (۲) انتخاب شود آنگاه:

$$\begin{aligned} F(2, t) &= F_2(t) - \int_0^t F_1(z) f_2(z) dz - \theta \sum_{i=0}^p \sum_{j=0}^{p-1} \binom{p}{i} \binom{p-1}{j} (-1)^{i+j} \\ &\times \left\{ q \int_0^t F_1(z)^{i+q} F_2(z)^{j+q-1} f_2(z) dz - (p+q) \int_0^t F_1(z)^{i+q} F_2(z)^{j+q} f_2(z) dz \right\}. \end{aligned}$$

توابع زیرتوزیع هستند. با روش مشابه در هر حالت $F(1, t)$ نیز به دست می‌آید.

نتیجه ۱.۳. فرض کنید شرایط قضیه ۱.۳ برقرار باشد، علاوه بر این،

$$F_1(x) = (1 + x^{-\delta})^{-\alpha},$$

$$F_2(x) = (1 + y^{-\delta})^{-\beta}$$

توابع توزیع حاشیه‌ای بر باشند که در آن $\alpha > 0$ و $\beta > 0$ و $\delta > 0$ تعریف می‌شود. در این صورت،

الف) اگر مدل (۱) انتخاب شود آنگاه:

$$F(y, t) = H_\delta(t)^\beta - \frac{\beta}{\alpha + \beta} H_\delta(t)^{\alpha + \beta} - \theta \beta \sum_{i=0}^q \sum_{j=0}^{q-1} \binom{q}{i} \binom{q-1}{j} (-1)^{i+j} \\ \times \left\{ \frac{H_\delta(t)^{\alpha(p_i+1) + \beta(p_j+1)}}{\alpha(p_i+1) + \beta(p_j+1)} - \frac{(1 + pq) H_\delta(t)^{\alpha(p_i+1) + \beta(p_j+p+1)}}{\alpha(p_i+1) + \beta(p_j+p+1)} \right\}$$

ب) اگر مدل (۲) انتخاب شود آنگاه:

$$F(y, t) = H_\delta(t)^\beta - \frac{\beta}{\alpha + \beta} H_\delta(t)^{\alpha + \beta} - \theta \beta \sum_{i=0}^p \sum_{j=0}^{p-1} \binom{p}{i} \binom{p-1}{j} (-1)^{i+j} \\ \times \left\{ \frac{q H_\delta(t)^{\alpha(i+q) + \beta(j+q)}}{\alpha(i+q) + \beta(j+q)} - \frac{(p+q) H_\delta(t)^{\alpha(i+q) + \beta(j+q+1)}}{\alpha(i+q) + \beta(j+q+1)} \right\}.$$

که در آن

$$K_{\alpha\delta} = \alpha \delta t^{-\delta-1}$$

$$K_{\beta\delta} = \beta \delta t^{-\delta-1}$$

$$H_\delta(t) = (1 + t^{-\delta})^{-1}$$

لم ۱.۳. فرض کنید (X, Y) یک بردار تصادفی با توزیع توام F و توابع توزیع حاشیه‌ای F_1 و F_2 باشد، علاوه بر این

$C = 2$ علت شکست است در این صورت،

$$R = P(Y < X) = F(y, +\infty)$$

نتیجه ۲.۳. فرض کنید شرایط نتیجه ۱.۳ برقرار باشد،

الف) اگر مدل (۱) انتخاب شود آنگاه:

$$F(y, \infty) = \frac{\alpha}{\alpha + \beta} - \theta \beta \sum_{i=0}^q \sum_{j=0}^{q-1} \binom{q}{i} \binom{q-1}{j} (-1)^{i+j} \\ \times \left\{ \frac{\beta(q - p^2 q j - pq) - \alpha p(pq i + q)}{[\alpha(p_i+1) + \beta(p_j+1)][\alpha(p_i+1) + \beta(p_j+p+1)]} \right\}$$

ب) اگر مدل (۲) انتخاب شود آنگاه:

$$F(r, \infty) = \frac{\alpha}{\alpha + \beta} - \theta \beta \sum_{i=0}^p \sum_{j=0}^{p-1} \binom{p}{i} \binom{p-1}{j} (-1)^{i+j} \\ \times \left\{ \frac{\beta(q - pq - pj) - \alpha p(i + q)}{[\alpha(i + q) + \beta(j + q)][\alpha(i + q) + \beta(j + q + 1)]} \right\}$$

مثال ۱.۳. فرض کنید شرایط نتیجه ۲.۳ برقرار باشد،

الف) اگر مدل (۱) انتخاب شود آنگاه:

۱ اگر $p = q = 1$ آنگاه:

$$R = \frac{\alpha}{\alpha + \beta} + \theta \frac{\alpha\beta(\alpha - \beta)}{(\alpha + \beta)(2\alpha + \beta)(\alpha + 2\beta)}$$

۲ اگر $p = q = 2$ آنگاه:

$$R = \frac{\alpha}{\alpha + \beta} + \theta \frac{\alpha^2 \beta^2 (\alpha - \beta) [1472(\alpha + \beta)^2 + 768\alpha\beta]}{(\alpha + \beta)(\alpha + 3\beta)(\alpha + 5\beta)(3\alpha + \beta)(3\alpha + 5\beta)(5\alpha + \beta)(5\alpha + 3\beta)}$$

۳ اگر $p = 1, q = 2$ آنگاه:

$$R = \frac{\alpha}{\alpha + \beta} + \theta \frac{564\alpha^5 \beta^2 + 1294\alpha^4 \beta^3 + 736\alpha^3 \beta^4 + 214\alpha^2 \beta^5 - 144\alpha\beta^6}{(\alpha + \beta)(\alpha + 2\beta)(\alpha + 3\beta)(2\alpha + \beta)(2\alpha + 3\beta)(3\alpha + \beta)(3\alpha + 2\beta)^2}$$

۴ اگر $p = 2, q = 1$ آنگاه:

$$R = \frac{\alpha}{\alpha + \beta} + \theta \frac{4\alpha\beta(3\alpha^2 + 2\alpha\beta - \beta^2)}{(\alpha + \beta)(\alpha + 3\beta)(3\alpha + \beta)(3\alpha + 3\beta)}$$

ب) اگر مدل (۲) انتخاب شود آنگاه:

۱ اگر $p = q = 1$ آنگاه:

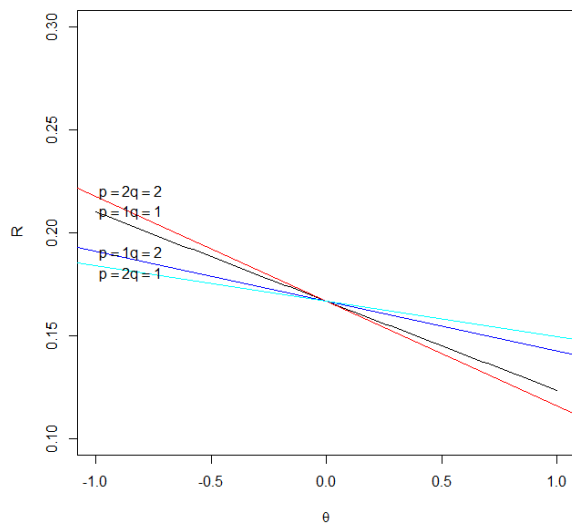
$$R = \frac{\alpha}{\alpha + \beta} + \theta \frac{\alpha\beta(\alpha - \beta)}{(\alpha + \beta)(2\alpha + \beta)(\alpha + 2\beta)}$$

۲ اگر $p = q = 2$ آنگاه:

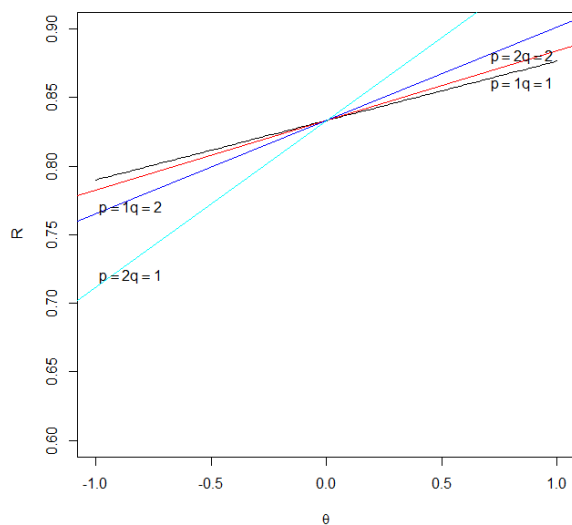
$$R = \frac{\alpha}{\alpha + \beta} + \theta \frac{\alpha^2 \beta^2 (\alpha - \beta) [26(\alpha + \beta)^2 + 3\alpha\beta]}{(\alpha + \beta)(2\alpha + 3\beta)(\alpha + 2\beta)(3\alpha + 2\beta)(3\alpha + 4\beta)(2\alpha + \beta)(4\alpha + 3\beta)}$$

۳ اگر $p = 1, q = 2$ آنگاه:

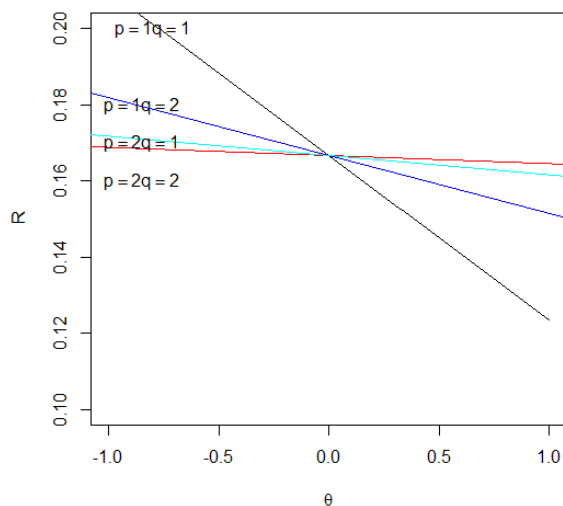
$$R = \frac{\alpha}{\alpha + \beta} + \theta \frac{\alpha\beta(\alpha - \beta)}{(\alpha + \beta)(2\alpha + 3\beta)(3\alpha + 2\beta)}$$



شکل ۱: نمودار اندازه قابلیت اطمینان برای مدل (۱) با حاشیه‌های بر نوع ۳ با پارامترهای $\alpha = 1$ و $\beta = 5$ و به‌ازای مقادیر مختلف p و q .



شکل ۲: نمودار اندازه قابلیت اطمینان برای مدل (۱) با حاشیه‌های بر نوع ۳ با پارامترهای $\alpha = 10$ و $\beta = 2$ و به‌ازای مقادیر مختلف p و q .



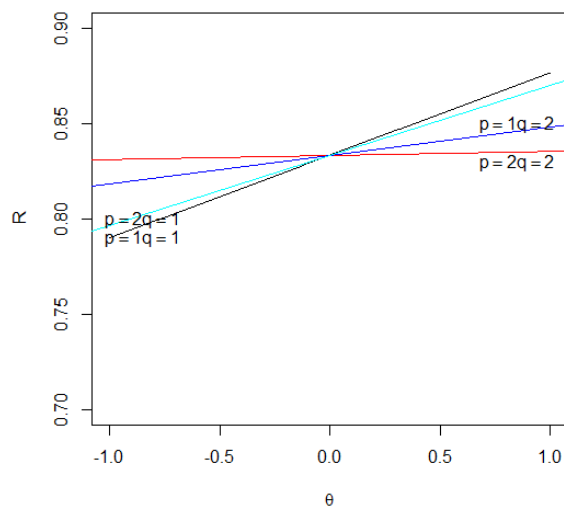
شکل ۳: نمودار اندازه قابلیت اطمینان برای مدل (۲) با حاشیه‌های بر نوع ۳ با پارامترهای $\alpha = 1$ و $\beta = 5$ و به‌ازای مقادیر مختلف p و q .

۴ اگر $p = 2, q = 1$ آنگاه:

$$R = \frac{\alpha}{\alpha + \beta} + \theta \frac{180\alpha^5\beta^2 + 432\alpha^4\beta^3 + 238\alpha^3\beta^4 + 2\alpha^2\beta^5 - 12\alpha\beta^6}{(\alpha + \beta)(\alpha + 2\beta)(\alpha + 3\beta)(2\alpha + \beta)(2\alpha + 3\beta)(3\alpha + \beta)(3\alpha + 2\beta)}$$

R تابعی خطی بر حسب θ می‌باشد، ضریب θ شیب خط و عرض از مبدا آن $\frac{\alpha}{\alpha + \beta}$ است. چون α و β همواره مثبت است پس $0 < \frac{\alpha}{\alpha + \beta} < 1$ می‌باشد. از طرفی چون مخرج ضریب θ و $\alpha\beta$ همواره مثبت است پس شیب هم علامت $(\alpha - \beta)$ است یعنی اگر $\alpha > \beta$ شیب خط صعودی و اگر $\alpha < \beta$ باشد شیب خط نزولی و در صورتی که $\alpha = \beta$ باشد شیب خط صفر است. در شکل‌های ۱ و ۳ چون $\alpha < \beta$ در نظر گرفته شده ملاحظه می‌کنیم که شیب خط نزولی و برای مقادیر مختلف p و q شدت نزولی شدن متفاوت است. در شکل‌های ۲ و ۴ چون $\alpha > \beta$ ، شیب خط صعودی و برای مقادیر مختلف p و q شدت صعودی بودن متفاوت است.

اگر در مدل R ، عرض از مبدا و شیب خط را به ترتیب با نماد R_I و D نشان دهیم آنگاه $R_I = \frac{\alpha}{\alpha + \beta}$ ، اندازه قابلیت در حالت استقلال با حاشیه‌های بر نوع ۳ و D ، به عنوان مثال در حالتی که $p = q = 1$ باشد، $D = \frac{\alpha\beta(\alpha - \beta)}{(\alpha + \beta)(2\alpha + \beta)(\alpha + 2\beta)}$ ، وابسته به پارامترهای α و β است اما چون D ضریب θ می‌باشد پس میزان وابستگی



شکل ۴: نمودار اندازه قابلیت اطمینان برای مدل (۲) با حاشیه‌های بر نوع ۳ با پارامترهای $\alpha = 10$ و $\beta = 2$ و به‌ازای مقادیر مختلف p و q .

را در معادله R نشان می‌دهد.

اگر $\theta > 0$ ، وابستگی مثبت باشد آنگاه $R > R_I$ یعنی اگر مولفه‌های سیستم با طول عمرهای X و Y به طور مثبت وابسته باشند آنگاه قابلیت سیستم در حالت استقلال کمتر از حالت وابسته است.
اگر $\theta < 0$ ، وابستگی منفی باشد آنگاه $R < R_I$ یعنی اگر مولفه‌های سیستم با طول عمرهای X و Y به طور منفی وابسته باشند آنگاه قابلیت سیستم در حالت استقلال بیشتر از حالت وابسته است.

مراجع

- [1] Shih, J. H., Emura, T. (2016). Bivariate dependence measures and bivariate competing risks models under the generalized FGM copula. *Statistical Papers*, 1-18.
- [2] Bairamov, I., Kotz, S. (2002). Dependence structure and symmetry of Huang-Kotz FGM distributions and their extensions. *Metrika*, 56(1), 55-72.
- [3] Lai, C. D., Xie, M. (2000). A new family of positive quadrant dependent bivariate distributions. *Statistics probability letters*, 46(4), 359-364.

- [4] Nelsen, R. B. (2006). An introduction to copulas, 2nd. New York: SpringerScience Business Media.
- [5] Morgenstern, D. (1956). Einfache beispiele zweidimensionaler verteilungen. *Mitteilungsblatt fur Mathematische Statistik*, 8, 234-235.
- [6] Louzada, F., Suzuki, A. K., Cancho, V. G. (2013). The FGM long-term bivariate survival copula model: modeling, Bayesian estimation, and case influence diagnostics. *Communications in Statistics-Theory and Methods*, 42(4), 673-691.
- [7] Gumbel, E. J. (1960). Bivariate exponential distributions. *Journal of the American Statistical Association*, 55(292), 698-707.
- [8] Farlie, D. J. (1960). The performance of some correlation coefficients for a general bivariate distribution. *Biometrika*, 47(3/4), 307-323.
- [9] Amini, M., Jabbari, H., Mohtashami Borzadaran, G. R. (2011). Aspects of dependence in generalized Farlie-Gumbel-Morgenstern distributions. *Communications in Statistics-Simulation and Computation*, 40(8), 1192-1205.
- [10] Domma, F., Giordano, S. (2013). A copula-based approach to account for dependence in stress-strength models. *Statistical Papers*, 54(3), 807-826.
- [11] Domma, F., Giordano, S. (2016). Concomitants of m-generalized order statistics from generalized Farlie-Gumbel-Morgenstern distribution family. *Journal of Computational and Applied Mathematics*, 294, 413-435.



پنجمین سمینار
نظریه مفصل و کاربردهای آن
۱۰ و ۱۱ بهمن ۱۳۹۷



نتایج درباره‌ی معیار نادرستی در نظریه‌ی اطلاع براساس تابع مفصل

حسینی، ت^۱ احمدی، ج^۲

گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده

تاکنون، معیارهای زیادی برای اندازه‌گیری اطلاعات نهفته در مشاهدات از توزیع‌های یک متغیره، معرفی شده است. بعضی از این معیارها مانند آنتروپی شانون، نادرستی کریچ و تعمیم‌های آنها، به حالت دو متغیره توسعه داده شده‌اند. در این مقاله، به دنبال برقراری ارتباط بین نظریه‌ی اطلاع و نظریه‌ی مفصل از طریق بکارگیری قضیه‌ی اسکالر هستیم. برای این منظور، دو معیار اطلاعاتی جدید تحت عنوان‌های نادرستی مفصلی و نادرستی گذشته‌ی تجمعی مفصلی را معرفی و خواص آنها را بررسی می‌کنیم.

کلمات کلیدی: آنتروپی شانون، اندازه‌ی نادرستی دو متغیره، اندازه‌ی نادرستی کریچ، اندازه‌ی نادرستی گذشته‌ی تجمعی دو متغیره، تابع مفصل

^۱ hoseini_toktam@mail.um.ac.ir

^۲ ahmadi-j@um.ac.ir

۱ مقدمه

ارتباط بین توزیع توأم یک بردار تصادفی با توابع توزیع حاشیه‌ای آن‌ها، یکی از موضوعات مورد توجه در مباحث احتمال و مدل‌سازی است. این موضوع در قالب نظریه‌ی مفصل بعنوان یک ابزار جدید در نظریه‌ی توزیع‌ها توسط اسکالر (۱۹۵۹) مطرح شده است. ژانگ و همکاران (۲۰۰۸) و کالبرگ (۲۰۰۸) مفهوم مفصل را بعنوان ابزاری مفید برای مدل‌سازی ساختار وابستگی بیان کرده‌اند. در واقع، نظریه‌ی مفصل روشی برای بدست آوردن توزیع‌های کناری، مشخص نمی‌کند بلکه در این نظریه، شیوه‌هایی برای ساختن تابع‌های مفصل گوناگون ارائه می‌شود که از طریق آن توابع مفصل و با داشتن توزیع‌های کناری، می‌توان به توزیع‌های توأم مختلفی رسید. بنابراین این نظریه، منجر به یک تابع مفصل یکتا نمی‌شود. راه‌های مختلفی برای ساختن مفصل‌ها و انتخاب بهترین مفصل وجود دارد. توابع مفصل بصورت توزیع‌های چندمتغیره‌ای که دارای کناری‌های یکنواخت هستند، تعریف می‌شوند. به بیان دقیق‌تر، تعریف زیر را داریم.

تعریف ۱.۱. $C(u, v)$ را تابع مفصل دومتغیره یا به اختصار، مفصل گویند که تابعی از I^2 به $I = [0, 1]$ تعریف می‌شود و در شرایط زیر باید صدق کند،

$$1- \text{ برای هر } u, v \in I \text{ هر } C(u, 0) = 0 = C(0, v)$$

$$2- \text{ به ازای هر } u, v \in I \text{ هر } C(u, 1) = u, C(1, v) = v$$

۳- اگر $u_1, u_2, v_1, v_2 \in I$ بطوری که $u_1 \leq u_2$ و $v_1 \leq v_2$ ، آنگاه

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

قضیه‌ی اسکالر (۱۹۵۹)، بیان می‌کند که برای هر تابع توزیع تجمعی مانند F ، با توابع توزیع کناری F_1 و F_2 ، تابع مفصلی مانند C ، وجود دارد طوری که

$$F(x, y) = C(F_1(x), F_2(y)), \quad x, y \in \bar{\mathbb{R}} = [-\infty, +\infty]. \quad (1)$$

در این قضیه، بیان شده است که اگر توابع کناری، پیوسته باشند مفصل C یکتا، در غیر این صورت به طور یکتا در $Ran F_1 \times Ran F_2$ تعیین می‌گردد. بالعکس اگر C یک مفصل باشد و F_1, F_2 توزیع‌های کناری باشند، آنگاه تابع توزیع توأم F ، توسط رابطه‌ی (۱) تعیین می‌گردد به طوری که F_1 و F_2 توزیع‌های کناری F باشند. با تغییر متغیر $u = F_1(x)$ و $v = F_2(y)$ در رابطه‌ی (۱) نتیجه می‌شود که

$$C(u, v) = F(F_1^{-1}(u), F_2^{-1}(v)), \quad (2)$$

که در آن $F_i^{-1}(u) = \inf\{x : F_i(x) \geq u\}$. چگالی مفصل متناظر با مفصل (۲) بصورت

$$c(u, v) = \frac{\partial^2}{\partial u \partial v} C(u, v) = \frac{\partial^2}{\partial u \partial v} F(F_1^{-1}(u), F_2^{-1}(v)), \quad u, v \in (0, 1), \quad (3)$$

بیان می‌شود. بنابراین تابع چگالی احتمال توأم بصورت

$$f(x, y) = f_1(x)f_2(y)c(F_1(x), F_2(y)), \quad x, y \in \mathbb{R}, \quad (4)$$

بدست می‌آید. برای جزئیات بیشتر درباره‌ی نظریه‌ی مفصل و کاربردهای آن به نلسن (۲۰۰۷) مراجعه شود. یکی از مفاهیم پایه در مبحث نظریه‌ی اطلاع، مفهوم آنتروپی است که اولین بار توسط شانون (۱۹۴۸) به منظور کمی ساختن اطلاعات از دست رفته در سیگنال‌های ارتباطی تلفن معرفی شد. این اندازه برای توزیع‌های پیوسته با تکیه گاه نامنفی بصورت

$$H(X_1) = - \int_{-\infty}^{+\infty} f_1(x) \log f_1(x) dx, \quad (5)$$

تعریف می‌شود که در آن f_1 تابع چگالی احتمال متغیر تصادفی نامنفی X_1 است.

بسیاری از محققین مفاهیم و شاخص‌های متعددی براساس آنتروپی شانون با کاربردهای مختلفی معرفی کردند. در این بین کریج (۱۹۶۱) و نات (۱۹۶۸) شاخصی تحت عنوان اندازه‌ی نادرستی به عنوان ابزاری برای اندازه‌گیری خطا در نتایج آزمایشگاهی معرفی نمودند که به نوعی تعمیمی از آنتروپی شانون است. فرض کنید Y_1 متغیر تصادفی نامنفی دیگری با تابع چگالی احتمال g_1 ، تابع توزیع احتمال G_1 و تابع بقای $\bar{G}_1 = 1 - G_1$ باشد. اگر f_1 توزیع واقعی متناظر با مشاهدات و g_1 توزیع مطرح شده توسط آزمایش‌گر باشد، آن‌گاه اندازه‌ی نادرستی متناظر با این دو متغیر تصادفی، بصورت

$$I(X_1, Y_1) = - \int_{-\infty}^{+\infty} f_1(x) \log g_1(x) dx, \quad (6)$$

تعریف شده است.

در نظریه‌ی قابلیت اعتماد، در مطالعه‌ی یک سیستم چند مؤلفه‌ای با هر مؤلفه‌ی دارای طول عمری وابسته به طول عمرهای بعدی، توزیع‌های عمر چند متغیره، نقش دارند. در سال‌های اخیر، موضوع اندازه‌گیری اطلاعات در توزیع‌های دو متغیره (چند متغیره) مورد توجه قرار گرفته است.

فرض کنید $\underline{X} = (X_1, X_2)$ و $\underline{Y} = (Y_1, Y_2)$ دو بردار تصادفی دومتغیره‌ی پیوسته‌ی نامنفی بترتیب با توابع چگالی احتمال توأم f و g باشند، در این صورت، معیار نادرستی در حالت دو متغیره بصورت

$$BI(\underline{X}, \underline{Y}) = - \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \log g(x, y) dy dx, \quad (7)$$

بدست می‌آید. جاش و کاندو (۲۰۱۷) معیار نادرستی گذشته‌ی تجمعی در حالت دو متغیره را بصورت

$$BCPI(\underline{X}, \underline{Y}) = - \int_{.}^{+\infty} \int_{.}^{+\infty} F(x, y) \log G(x, y) dy dx, \quad (۸)$$

بیان کردند، که در آن F و G ، بترتیب، تابع‌های توزیع توأم این دو بردار تصادفی می‌باشند. معیارهای تعریف شده در روابط (۷) و (۸) در حالتی که \underline{X} و \underline{Y} دارای توزیع‌های یکسانی هستند، بترتیب، به آنتروپی دو متغیره و آنتروپی گذشته‌ی تجمعی دو متغیره تبدیل می‌شود.

۲ معیار نادرستی بر مبنای تابع چگالی مفصل

فرض کنید $\underline{X} = (X_1, X_2)$ و $\underline{Y} = (Y_1, Y_2)$ دو بردار تصادفی پیوسته‌ی نامنفی، بترتیب با توابع چگالی احتمال توأم f و g ، تابع‌های توزیع توأم F و G و توابع بقای توأم \bar{F} و \bar{G} هستند طوری که f_i و g_i بترتیب، چگالی‌های احتمال کناری f و g ، همچنین F_i و G_i بترتیب، تابع‌های توزیع حاشیه‌ای F و G و \bar{F}_i و \bar{G}_i نیز بترتیب، توابع بقای کناری \bar{F} و \bar{G} باشند. بعلاوه فرض کنید توابع چگالی مفصل $c_{\underline{X}}$ و $c_{\underline{Y}}$ وجود داشته باشند بطوری که

$$f(x, y) = f_1(x)f_2(y)c_{\underline{X}}(F_1(x), F_2(y)), \quad x, y \in \mathbb{R}^+, \quad (۹)$$

و

$$g(x, y) = g_1(x)g_2(y)c_{\underline{Y}}(G_1(x), G_2(y)), \quad x, y \in \mathbb{R}^+. \quad (۱۰)$$

در این حالت معیار نادرستی برحسب تابع چگالی مفصل را بصورت

$$CI(\underline{X}, \underline{Y}) = - \int_{.}^1 \int_{.}^1 c_{\underline{X}}(u, v) \log c_{\underline{Y}}(G_1(F_1^{-1}(u)), G_2(F_2^{-1}(v))) dv du, \quad (۱۱)$$

تعریف و به آن نادرستی مفصلی می‌گوییم. در ادامه نتایجی برای نادرستی مفصلی و ارتباط آن با معیار نادرستی در حالت یک متغیره و دو متغیره بیان می‌کنیم.

گزاره ۱۰۲. فرض کنید X_i و Y_i ، بترتیب، متغیرهای حاشیه‌ای بردارهای تصادفی \underline{X} و \underline{Y} باشند ($i = 1, 2$)، در این صورت

$$CI(\underline{X}, \underline{Y}) = BI(\underline{X}, \underline{Y}) - I(X_1, Y_1) - I(X_2, Y_2). \quad (۱۲)$$

با توجه به رابطه‌ی (۳) می‌توان نتیجه گرفت که اگر Y_1 و Y_2 دو متغیر تصادفی مستقل از هم باشند آن‌گاه $CI(\underline{X}, \underline{Y}) = 0$. بطور خاص، اگر متغیرهای تصادفی Y_1 و Y_2 بترتیب با X_1 و X_2 هم‌توزیع باشند آن‌گاه

برابری (۳) به

$$CI(\underline{X}, \underline{Y}) = - \int \int c_{\underline{X}}(u, v) \log c_{\underline{Y}}(u, v) dv du, \quad (13)$$

تبدیل می‌شود و به جای رابطه‌ی (۴) تساوی

$$CI(\underline{X}, \underline{Y}) = BI(\underline{X}, \underline{Y}) - H(X_1) - H(X_2),$$

را داریم. همچنین، فرض کنید بردارهای تصادفی \underline{X} و \underline{Y} هم‌توزیع باشند، در این صورت، نادرستی مفصلی (۱۳) به آنتروپی مفصلی در سانگ و ژانگ (۲۰۱۸) تبدیل می‌شود. در گزاره‌های بعدی تحت فرض برقراری مدل‌های نرخ خطر متناسب و نرخ خطر متناسب معکوس بین توزیع‌های کناری، نتایجی برای نادرستی مفصلی بدست می‌آوریم.

گزاره ۲.۲. اگر برای هر $t > 0$ داشته باشیم

الف) $G_2(t) = G_1^\alpha(t)$ که α یک مقدار حقیقی نامنفی و مخالف با یک است، آنگاه

$$CI(\underline{X}, \underline{Y}) = BI(\underline{X}, \underline{Y}) - I(X_1, Y_1) - I(X_2, Y_1) + \log \alpha + (\alpha - 1)E[\log G_1(X_2)].$$

ب) $\bar{G}_2(t) = \bar{G}_1^\alpha(t)$ آنگاه

$$CI(\underline{X}, \underline{Y}) = BI(\underline{X}, \underline{Y}) - I(X_1, Y_1) - I(X_2, Y_1) + \log \alpha + (\alpha - 1)E[\log \bar{G}_1(X_2)].$$

ج) اگر علاوه بر فرض الف و ب، بترتیب داشته باشیم $F_2(t) = G_1^\beta(t)$ و $\bar{F}_2(t) = \bar{G}_1^\beta(t)$ با $\beta > 0, \beta \neq 1$ آنگاه

$$CI(\underline{X}, \underline{Y}) = BI(\underline{X}, \underline{Y}) - I(X_1, Y_1) - H(X_2) + \log \frac{\alpha}{\beta} + \frac{\beta - \alpha}{\beta}.$$

گزاره ۳.۲. فرض کنید برای متغیرهای تصادفی X_1 و Y_1

$$\forall t \geq 0, G_1(t) = F_1^\alpha(t),$$

یا

$$\forall t \geq 0, \bar{G}_1(t) = \bar{F}_1^\alpha(t),$$

و برای X_2 و Y_2

$$\forall t \geq 0, G_2(t) = F_2^\beta(t),$$

یا

$$\forall t \geq 0, \bar{G}_\tau(t) = \bar{F}_\tau^\beta(t),$$

برقرار باشد، که α و β دو مقدار حقیقی نامنفی و مخالف با یک هستند. در این صورت

$$CI(\underline{X}, \underline{Y}) = BI(\underline{X}, \underline{Y}) - H(X_\tau) - H(Y_\tau) + \log \alpha \beta - (\alpha + \beta - 2).$$

گزاره ۴.۲. فرض کنید $X_\tau = \varphi(X_1)$ و $Y_\tau = \varphi(Y_1)$.

(الف) اگر φ یک تابع نامنفی اکیدا صعودی و مشتق‌پذیر باشد طوری که $\varphi(+\infty) = +\infty$ و $\varphi(0) = 0$ آن‌گاه

$$CI(\underline{X}, \underline{Y}) = BI(\underline{X}, \underline{Y}) - \tau I(X_1, Y_1) - E[\log \varphi'(X_1)],$$

(ب) اگر φ تابعی نامنفی، اکیدا نزولی و مشتق‌پذیر باشد. همچنین داشته باشیم $\varphi(+\infty) = 0$ و $\varphi(0) = +\infty$ آن‌گاه

$$CI(\underline{X}, \underline{Y}) = BI(\underline{X}, \underline{Y}) + E[\log |\varphi'(X_1)|],$$

که در آن φ' مشتق تابع φ است.

۳ معیار نادرستی بر مبنای تابع مفصل

علاوه بر مفروضات بخش دوم، فرض کنید توابع مفصل $C_{\underline{X}}$ و $C_{\underline{Y}}$ وجود داشته باشد بطوری‌که

$$F(x, y) = C_{\underline{X}}(F_1(x), F_2(y)),$$

و

$$G(x, y) = C_{\underline{Y}}(G_1(x), G_2(y)).$$

در این حالت، نادرستی گذشته‌ی تجمعی مفصلی را بصورت

$$CCPI(\underline{X}, \underline{Y}) = - \int_0^1 \int_0^1 C_{\underline{X}}(u, v) \log C_{\underline{Y}}(G_1(F_1^{-1}(u)), G_2(F_2^{-1}(v))) dv du, \quad (14)$$

تعریف می‌کنیم. همان‌طور که در تعریف تابع مفصل بیان شد، این تابع مقادیر بازه‌ی $(0, 1)$ را اختیار می‌کند به همین دلیل، واضح است که معیار تعریف شده در (۵)، همواره نامنفی است در حالی که نادرستی مفصلی لزومی ندارد نامنفی باشد. فرض کنید متغیرهای تصادفی Y_1 و Y_2 بترتیب با X_1 و X_2 هم‌توزیع هستند. در این صورت معیار (۵) به

$$CCPI(\underline{X}, \underline{Y}) = - \int_0^1 \int_0^1 C_{\underline{X}}(u, v) \log C_{\underline{Y}}(u, v) dv du,$$

تبدیل می‌شود.

گزاره ۱.۳. اگر متغیرهای تصادفی Y_1 و Y_2 مستقل از یکدیگر باشند و برای هر $t > 0$ داشته باشیم $G_1(t) = F_1^\alpha(t)$ و $G_2(t) = F_2^\beta(t)$ که α و β مقادیر حقیقی نامنفی و مخالف با یک هستند. آن‌گاه

$$CCPI(\underline{X}, \underline{Y}) \leq \frac{\wedge}{\vee}(\alpha + \beta).$$

گزاره ۲.۳. فرض کنید متغیرهای تصادفی Y_1 و Y_2 بترتیب با X_1 و X_2 هم‌توزیع‌اند. اگر متغیرهای تصادفی Y_1 و Y_2 مستقل از یکدیگر باشند، آن‌گاه

$$CCPI(\underline{X}, \underline{Y}) \leq \frac{16}{27}.$$

در ادامه نتایجی را برای سه بردار تصادفی دومتغیره بررسی می‌کنیم. برای این منظور فرض کنید $\underline{Z} = (Z_1, Z_2)$ یک بردار تصادفی پیوسته‌ی نامنفی با تابع توزیع توأم $H(x, y) = C_{\underline{Z}}(H_1(x), H_2(y))$ باشد.

گزاره ۳.۳. فرض کنید تساوی‌های $H_1(x) = F_1^\alpha(x)$ ، $H_2(y) = F_2^\beta(y)$ و $C_{\underline{X}}(u, v) = C_{\underline{Z}}(u, v)$ برقرار باشند.

الف) اگر $\alpha, \beta > 1$ آن‌گاه

$$CCPI(\underline{Z}, \underline{Y}) \leq \alpha\beta CCPI(\underline{X}, \underline{Y}).$$

ب) اگر $0 < \alpha, \beta < 1$ آن‌گاه

$$CCPI(\underline{Z}, \underline{Y}) \geq \alpha\beta CCPI(\underline{X}, \underline{Y}).$$

تعریف ۱.۳. (ترتیب مربعی پایین). بردار تصادفی \underline{X} در ترتیب مربعی پایین کوچک‌تر یا مساوی با بردار تصادفی \underline{Y} نامیده می‌شود اگر و تنها اگر

$$\forall (x, y) \in \mathbb{R}_+^2, \quad F(x, y) \geq G(x, y),$$

و با نماد $\underline{X} \leq_{LO} \underline{Y}$ نشان داده می‌شود.

گزاره ۴.۳. اگر $\underline{Y} \leq_{LO} \underline{Z}$ ، آن‌گاه

$$CCPI(\underline{X}, \underline{Y}) \leq CCPI(\underline{X}, \underline{Z}).$$

در گزاره‌ی یاد شده، اگر بردارهای تصادفی \underline{X} و \underline{Z} هم‌توزیع باشند آن‌گاه

$$CCPI(\underline{X}, \underline{Y}) \leq CCPI(\underline{X}, \underline{X}),$$

که در آن $CCPI(\underline{X}, \underline{X})$ ، به آنتروپی گذشته‌ی تجمعی مفصلی معروف است و با نماد $CCPE(\underline{X})$ نشان داده می‌شود.

گزاره ۵.۳. اگر $\underline{Y} \leq_{LO} \underline{Z}$ و متغیرهای تصادفی Z_1 و Z_2 ، بترتیب، هم‌توزیع با Y_1 و Y_2 باشند، آنگاه

$$CCPI(\underline{Z}, \underline{X}) \leq CCPI(\underline{Y}, \underline{X}).$$

گزاره ۶.۳. فرض کنید $\underline{X} \leq_{LO} \underline{Z}$ و $\underline{Y} \leq_{LO} \underline{Z}$ ، اگر متغیرهای تصادفی Z_1 و Z_2 ، بترتیب، هم‌توزیع با X_1 و X_2 باشند، آنگاه

$$CCPI(\underline{Z}, \underline{Y}) \leq CCPI(\underline{X}, \underline{Y}) \leq CCPI(\underline{X}, \underline{Z}).$$

در گزاره‌ی زیر برقراری نامساوی مثلث را برای نادرستی گذشته‌ی تجمعی مفصلی، بررسی می‌کنیم.

گزاره ۷.۳. فرض کنید $\underline{Y} \leq_{LO} \underline{Z}$ در این صورت داریم

(الف)

$$CCPI(\underline{X}, \underline{Y}) \leq CCPI(\underline{X}, \underline{Z}) + CCPI(\underline{Z}, \underline{Y}).$$

(ب) اگر Z_1 و Z_2 ، بترتیب هم‌توزیع با Y_1 و Y_2 باشند آنگاه

$$CCPI(\underline{Z}, \underline{X}) \leq CCPI(\underline{Z}, \underline{Y}) + CCPI(\underline{Y}, \underline{X}).$$

گزاره ۸.۳. اگر $\underline{X} \leq_{LO} \underline{Z} \leq_{LO} \underline{Y}$ و متغیرهای تصادفی Z_1 و Z_2 ، بترتیب، هم‌توزیع با X_1 و X_2 باشند آنگاه

$$CCPI(\underline{X}, \underline{Y}) \geq \max\{CCPI(\underline{X}, \underline{Z}), CCPI(\underline{Z}, \underline{Y})\}.$$

گزاره ۹.۳. اگر $\underline{W} = (W_1, W_2)$ یک بردار تصادفی پیوسته‌ی نامنفی با تابع توزیع توأم

$$D(x, y) = C_{\underline{W}}(D_1(x), D_2(y)),$$

باشد، طوری که $Z_i = \varphi_i(X_i)$ و $W_i = \varphi_i(Y_i)$ و φ_i توابعی اکیدا صعودی و نامنفی باشند، آنگاه

$$CCPI(\underline{Z}, \underline{W}) = CCPI(\underline{X}, \underline{Y}).$$

گزاره ۱۰.۳. اگر \underline{X} و \underline{Y} بصورت $G(x, y) = F^\alpha(x, y)$ با $\alpha > 0$ و $\alpha \neq 1$ در مدل نرخ خطر متناسب معکوس صدق کنند، آنگاه

$$CCPI(\underline{X}, \underline{Y}) = \alpha CCPE(\underline{X}).$$

از گزاره‌ی بالا می‌توان نتیجه گرفت که در صورت برقراری شرایط، اگر $\alpha > 1$ آنگاه

$$CCPI(\underline{X}, \underline{Y}) \geq CCPE(\underline{X}),$$

و برای $0 < \alpha < 1$ جهت نامساوی برعکس می‌شود.

۴ مثال

در این بخش، به منظور بررسی نتایج بدست آمده، دو مثال ارائه می‌شود.

مثال ۱.۴. فرض کنید Z_1 و Z_2 دارای حداکثر وابستگی باشند، یعنی $C_Z(u, v) = \min\{u, v\}$. همچنین داشته باشیم $Z_i \leq_{st} X_i, Z_i \leq_{st} Y_i$ برای $i = 1, 2$ (برای جزئیات درباره‌ی خواص و کاربرد این تابع مفصل به نلسن و همکاران (۲۰۰۱) مراجعه نمایید). در این صورت ترتیب‌های $\underline{Z} \leq_{LO} \underline{X}$ و $\underline{Z} \leq_{LO} \underline{Y}$ را داریم. بنا به گزاره‌ی ۴.۳ نتیجه می‌شود که

$$CCPI(\underline{X}, \underline{Z}) \leq CCPI(\underline{X}, \underline{Y}) \quad \text{و} \quad CCPI(\underline{Y}, \underline{Z}) \leq CCPI(\underline{Y}, \underline{X}).$$

و نامساوی مثلث بصورت

$$CCPI(\underline{X}, \underline{Z}) \leq CCPI(\underline{X}, \underline{Y}) + CCPI(\underline{Y}, \underline{Z}),$$

و

$$CCPI(\underline{Y}, \underline{Z}) \leq CCPI(\underline{Y}, \underline{X}) + CCPI(\underline{X}, \underline{Z}),$$

برقرار است. حال اگر متغیرهای تصادفی Z_1 و Z_2 ، بترتیب، هم‌توزیع با Y_1 و Y_2 باشند، براساس گزاره‌ی ۵.۳ داریم

$$CCPI(\underline{Y}, \underline{X}) \leq CCPI(\underline{Z}, \underline{X}).$$

در این مورد نامساوی مثلث بصورت

$$CCPI(\underline{Y}, \underline{X}) \leq CCPI(\underline{Y}, \underline{Z}) + CCPI(\underline{Z}, \underline{X}),$$

برقرار می‌شود. اگر متغیرهای تصادفی Z_1 و Z_2 ، بترتیب، هم‌توزیع با X_1 و X_2 باشند، براساس گزاره‌ی ۵.۳ داریم

$$CCPI(\underline{X}, \underline{Y}) \leq CCPI(\underline{Z}, \underline{Y}),$$

و در این مورد نامساوی مثلث بصورت

$$CCPI(\underline{X}, \underline{Y}) \leq CCPI(\underline{X}, \underline{Z}) + CCPI(\underline{Z}, \underline{Y}),$$

برقرار است.

مثال ۲.۴. اگر Z_1 و Z_2 دارای حداقل وابستگی باشند یعنی $C_Z(u, v) = \max\{u + v - 1, 0\}$. همچنین داشته باشیم $Z_i \geq_{st} X_i, Z_i \geq_{st} Y_i$ برای $i = 1, 2$ (برای جزئیات درباره‌ی خواص و کاربرد این تابع مفصل به نلسن و همکاران (۲۰۰۱) مراجعه نمایید). در این صورت ترتیب‌های $\underline{Z} \geq_{LO} \underline{X}$ و $\underline{Z} \geq_{LO} \underline{Y}$ را داریم. با توجه به گزاره‌ی ۴.۳ داریم

$$CCPI(\underline{X}, \underline{Y}) \leq CCPI(\underline{X}, \underline{Z}) \quad \text{و} \quad CCPI(\underline{Y}, \underline{X}) \leq CCPI(\underline{Y}, \underline{Z}).$$

و نامساوی مثلث بصورت

$$CCPI(\underline{X}, \underline{Y}) \leq CCPI(\underline{X}, \underline{Z}) + CCPI(\underline{Z}, \underline{Y}),$$

و

$$CCPI(\underline{Y}, \underline{X}) \leq CCPI(\underline{Y}, \underline{Z}) + CCPI(\underline{Z}, \underline{X}),$$

برقرار است. حال اگر متغیرهای تصادفی Z_1 و Z_2 ، بترتیب، هم‌توزیع با Y_1 و Y_2 باشند، براساس گزاره‌ی ۵.۳ داریم

$$CCPI(\underline{Z}, \underline{X}) \leq CCPI(\underline{Y}, \underline{X}),$$

و در این مورد نامساوی مثلث بصورت

$$CCPI(\underline{Z}, \underline{X}) \leq CCPI(\underline{Z}, \underline{Y}) + CCPI(\underline{Y}, \underline{X}),$$

برقرار می‌شود و اگر متغیرهای تصادفی Z_1 و Z_2 ، بترتیب، هم‌توزیع با X_1 و X_2 باشند، براساس گزاره‌ی ۵.۳ داریم

$$CCPI(\underline{Z}, \underline{Y}) \leq CCPI(\underline{X}, \underline{Y}).$$

و در این مورد نامساوی مثلث بصورت

$$CCPI(\underline{Z}, \underline{Y}) \leq CCPI(\underline{Z}, \underline{X}) + CCPI(\underline{X}, \underline{Y}),$$

برقرار است.

۵ جمع‌بندی و نتیجه‌گیری

در این مقاله با استفاده از تابع چگالی مفصل و تابع مفصل به دنبال برقراری ارتباط بین دو نظریه‌ی اطلاع بویژه معیار نادرستی و نظریه‌ی مفصل بودیم. در بخش دوم با استفاده از تابع چگالی مفصل، معیار جدیدی تحت عنوان نادرستی مفصلی را معرفی کردیم و برای آن، نتایجی را بیان کردیم. در بخش سوم، معیار جدید نادرستی گذشته‌ی تجمعی مفصلی را با استفاده از تابع مفصل تعریف کردیم و نتایجی را برای آن ذکر کردیم. بخش چهارم نیز به ارائه‌ی دو مثال اختصاص داده شد. با توجه به مطالب ارائه شده در این مقاله مشاهده می‌کنیم که استفاده از نظریه‌ی مفصل و مفاهیم آن می‌تواند نقش مفیدی در گسترش نظریه‌ی اطلاع، بویژه معیار نادرستی ایفا کند. علاوه‌براین، در نظریه‌ی مفصل می‌توان از نادرستی در حالت دو متغیره (چند متغیره) برای تشخیص خطا در تعیین مدل دو متغیره (چند متغیره) با استفاده از قضیه‌ی اسکالر نیز استفاده نمود.

مراجع

- [1] Ghosh, A., Kundu, C. (2017). Bivariate extension of (dynamic) cumulative residual and past inaccuracy measures. *Statistical Papers*, 1-28.
- [2] Kallenberg, W. C. (2008). Modelling dependence. *Insurance: Mathematics and Economics*, 42(1), 127-146.
- [3] Kerridge, D. F. (1961). Inaccuracy and inference. *Journal of the Royal Statistical Society. Series B (Methodological)*, 184-194.
- [4] Nath, P. (1968). Inaccuracy and coding theory. *Metrika*, 13(1), 123-135.
- [5] Nelsen, R. B. (2007). *An Introduction to Copulas*. Springer Science Business Media.
- [6] Nelsen, R. B., Quesada-Molina, J. J., Rodríguez-Lallena, J. A., Ubeda-Flores, M. (2001). Bounds on bivariate distribution functions with given margins and measures of association. *Communications in Statistics-Theory and Methods*, 30(6), 1055-1062.
- [7] Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27(3), 379-423.
- [8] Singh, V. P., Zhang, L. (2018). Copula-entropy theory for multivariate stochastic modeling in water engineering. *Geoscience Letters*, 5(1), 6.

- [9] Sklar, M. (1959). Fonctions de repartition an dimensions et leurs marges. Publ. inst. statist. univ. Paris, 8, 229-231.
- [10] Zhang, D., Wells, M. T., Peng, L. (2008). Nonparametric estimation of the dependence function for a multivariate extreme value distribution. Journal of Multivariate Analysis, 99(4), 577-588.



پنجمین سمینار
نظریه مفصل و کاربردهای آن
۱۰ و ۱۱ بهمن ۱۳۹۷



ارزش در معرض ریسک شرطی برای مفصل‌های با وابستگی دمی

سالاری‌نژاد، ز^۱ میرحسینی، س.م.^۲
گروه آمار، دانشکده علوم ریاضی، دانشگاه یزد

چکیده

در این مقاله به بررسی دو معیار ارزش در معرض ریسک و ارزش در معرض ریسک شرطی از دیدگاه مفصل و ارائه معایب استفاده از ارزش در معرض ریسک شرطی استاندارد پرداخته می‌شود و سپس ارزش در معرض ریسک شرطی اصلاح‌شده برای مفصل‌هایی با وابستگی دمی غیربديهی مطالعه می‌شود.

کلمات کلیدی: مفصل، وابستگی دمی، ارزش در معرض ریسک، ارزش در معرض ریسک شرطی، چندک شرطی

۱ مقدمه

از دیدگاه نظری، هر فعالیت اقتصادی توأم با درجه ریسک است. ریسک را نمی‌توان کاملاً حذف کرد بنابراین نگرش علمی به مسئله ریسک چیزی جز مدیریت و اندازه‌گیری آن نیست. از جمله معیارهای اندازه‌گیری ریسک

^۱z.salarinegad@gmail.com

^۲mmirhoseini@yazd.ac.ir

می‌توان دو معیار ارزش در معرض ریسک (VaR) ^۱ و ارزش در معرض ریسک شرطی (CoVaR) ^۲ نام برد. اولین بار VaR توسط مورگان در سال ۱۹۹۴ مطرح شد [۷]. آرتزرنر و همکاران ^۳ در سال ۱۹۹۹ برخی از ویژگی‌هایی که VaR باید داشته باشد را پیشنهاد نمودند [۲]. نویسندگان بسیاری از جمله پیرسون ^۴ [۱۳] در سال ۲۰۰۲، چرنوژوکوف و دو ^۵ [۵] در سال ۲۰۰۸، برناردینو و همکاران ^۶ [۳] در سال ۲۰۱۵، ارزش در معرض ریسک را در پژوهش‌های خود مورد بررسی قرار داده‌اند. CoVaR یک معیار ریسک سیستماتیک در سال ۲۰۰۸ توسط آدریان و برانر میر مطرح شد [۱]. در این میان افرادی مانند گیراردی و ارگون ^۷ در سال ۲۰۱۳، مانیک و اسچانینگ ^۸ در سال ۲۰۱۴، برناردی و همکاران ^۹ در سال ۲۰۱۵، هر دل و همکاران ^{۱۰} در سال ۲۰۱۶ و او و پاتون ^{۱۱} در سال ۲۰۱۶ به مطالعه موضوعاتی در ارتباط با CoVaR و محاسبات CoVaR و کاربردهای آن در علوم اقتصادی و مالی پرداختند [۱]. با توجه به این که وابستگی متغیرها نقش اساسی در محاسبه CoVaR ایفا می‌کنند و با عنایت به اینکه محاسبه CoVaR به دم توزیع وابسته است بنابراین استفاده از مفصل‌هایی که وابستگی دمی دارند، گزینه‌های مناسب‌تری برای محاسبه CoVaR هستند. در این مقاله به بررسی و مطالعه CoVaR برای مفصل‌هایی که وابستگی دمی دارند، پرداخته می‌شود.

۲ معیارهای ریسک

به منظور حفظ یک مؤسسه مالی باید ریسک این مؤسسه مالی کاهش دهیم. برای محاسبه ریسک یک مؤسسه مالی از معیار VaR استفاده می‌شود.

تعریف ۱.۰.۲ [۹] (VaR) به ازای هر $\alpha \in (0, 1)$ ، ارزش در معرض ریسک در سطح α برای متغیر تصادفی X به صورت زیر تعریف می‌شود

$$VaR_{\alpha}(X) = -Q_{\alpha}^{+}(X) = Q_{1-\alpha}^{-}(-X),$$

که $Q_{\alpha}^{+}(X)$ چندک بالایی α ام متغیر تصادفی X و $Q_{1-\alpha}^{-}(-X)$ چندک پایینی $1 - \alpha$ ام متغیر تصادفی

^۱ Value at Risk

^۲ Conditional Value at Risk

^۳ Artzner et al

^۴ Pearson

^۵ Chernozhukov and Du

^۶ Bernardino et al

^۷ Girardi and Ergun

^۸ Mainik and schaanning

^۹ Bernardi et al

^{۱۰} Hardle et al

^{۱۱} Oh and Patton

X - است.

در برخی موارد شرایط بسیار خطرناک است یعنی ریسک یک مؤسسه مالی بر روی ریسک مؤسسه مالی دیگر تأثیرگذار است و به صورت یک دومینو باعث نابودی مؤسسه مالی دیگر می‌شود. در چنین حالتی فرض کنید که مؤسسات مالی مثلاً X متحمل زیان قابل توجهی شده است. با توجه به زیان این مؤسسه، مؤسسه مالی دیگر مثل Y چه میزان کمک برای اینکه متحمل زیان نشود، نیاز دارد. برای بدست آوردن این میزان کمک از معیار $CoVaR$ استفاده می‌شود.

تعریف ۲.۲. $[CoVaR]$ فرض کنید X و Y دو متغیر تصادفی باشند. به ازای هر $\alpha, \beta \in (0, 1)$

$$CoVaR(Y|X) := VaR_{\beta}(Y|X \in E),$$

که در آن E یک زیر مجموعه بورل از خط حقیقی و نشان دهنده یک پیشامد ناسازگار در مورد X است.

اغلب اوقات E شامل یک نقطه و یا نیم‌خط محدود به یک آستانه است. آدریان و برونرمایر از ساختار E شامل یک نقطه استفاده کردند. برای تمایز رویکرد آن‌ها، $CoVaR$ معرفی شده توسط آن‌ها را $CoVaR$ استاندارد می‌نامیم؛ و به طور خاص؛ VaR شرطی استاندارد در سطح (α, β) به صورت VaR در سطح β از Y تحت شرایطی تعریف می‌شود که $X = -VaR_{\alpha}(X)$. در ادامه به تعریف $CoVaR^=$ پرداخته می‌شود.

$$CoVaR_{\alpha, \beta}^=(Y|X) = VaR_{\beta}(Y|X = -VaR_{\alpha}(X)) \quad (1)$$

استفاده از $CoVaR^=$ در عمل با اشکالاتی همراه است [۱۱]:

(۱) محاسبه $CoVaR^=$ دشوار و در برخی موارد ناممکن است.

(۲) $CoVaR^=$ تابع صعودی از پارامتر وابستگی نیست.

(۳) $CoVaR^=$ نمی‌تواند ریسک سیستماتیک را که در زمان بیشتر مشخص می‌شود، شناسایی کند.

به همین دلیل مانیک و اسچانینگ [۱۱] در سال ۲۰۱۴، $CoVaR$ استاندارد را اصلاح و $CoVaR$ اصلاح شده را براساس VaR شرطی اصلاح شده به صورت زیر تعریف کردند.

تعریف ۳.۲. ($CoVaR$ اصلاح شده) [۹] - $CoVaR$ اصلاح شده در سطح (α, β) به عنوان VaR در سطح

β برای متغیر تصادفی Y به شرط $X \leq -VaR_{\alpha}(X)$ است تعریف می‌شود به عبارتی دیگر

$$CoVaR_{\alpha, \beta}^{\leq}(Y|X) = VaR_{\beta}(Y|X \leq -VaR_{\alpha}(X)). \quad (2)$$

می‌توان عبارت ۲ را از منظر چندک به صورت زیر بیان کرد:

$$CoVaR_{\alpha, \beta}^{\leq}(Y|X) = -Q_{\beta}^+(Y|X \leq Q_{\alpha}^+(X)). \quad (3)$$

از تعریف CoVaR مشخص است که باید وابستگی بین X و Y را مدل‌سازی کرد و این را می‌توان با استفاده از مفصل انجام داد.

۳ مفصل

ابتدا به بیان تعریف مفصل دو متغیره پرداخته می‌شود.

تعریف ۱.۰۳. (مفصل) فرض کنید که $I = [0, 1]$ یک بازه باشد. تابع دو متغیره $C : I^2 \rightarrow I$ مفصل نامیده می‌شود اگر شرایط زیر برقرار باشد:

(۱)

$$\forall u \in [0, 1], \quad C(u, 0) = C(0, u) = 0;$$

(۲)

$$\forall u, v \in [0, 1], \quad C(u, 1) = u, C(1, v) = v;$$

(۳)

$$\forall u_1, u_2, v_1, v_2 \in [0, 1], \quad u_1 \leq v_1, u_2 \leq v_2$$

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

مفصل $\hat{C}(u, v)$ را مفصل بقای $C(u, v)$ نامند که به صورت $\hat{C}(u, v) = u + v - 1 + C(1 - u, 1 - v)$ تعریف می‌شود.

نکته ۱.۰۳. [۹] زمانی که C یک مفصل مربوط به سود مؤسسات مالی X و Y باشد، آنگاه مفصل بقاء متناظر یعنی \hat{C} ، مفصل مربوط به ضرر مؤسسات مالی یعنی $L_1 = -X$ و $L_2 = -Y$ است.

همچنان که ملاحظه می‌شود محاسبه CoVaR به دم توزیع وابسته است. بنابراین به ارائه تعریف بسط دمی و وابستگی دمی پرداخته می‌شود.

تعریف ۲.۰۳. (بسط دمی) [۹] مفصل C داری بسط دمی در رأس مربع واحد $(0, 0)$ است اگر به ازای همه x, y های نامنفی

$$\lim_{t \rightarrow 0^+} \frac{C(tx, ty)}{t}$$

موجود باشد.

تابع

$$L : [0, \infty]^2 \rightarrow [0, \infty), \quad L(x, y) = \lim_{t \rightarrow 0^+} \frac{C(tx, ty)}{t}$$

را تابع وابستگی دمی یا به اصلاح، بسط دمی نامیده می‌شود.

می‌توان تابع وابستگی دمی را بر اساس مفصل بقاء به صورت زیر نوشت:

$$\hat{L}(x, y) = \lim_{t \rightarrow 0} \frac{\hat{C}(tx, ty)}{t} = \lim_{t \rightarrow 0} \frac{tx + ty - 1 + C(1 - tx, 1 - ty)}{t}$$

توجه داشته باشید که $L(1, 1)$ و $\hat{L}(1, 1)$ برابر با ضرایب وابستگی دمی پایین و بالایی هستند.

۴ CoVaR اصلاح شده مفصل مبنا

[۴] فرض کنید $C(u, v)$ یک مفصل از متغیرهای تصادفی X و Y و دارای توابع توزیع پیوسته F_X و F_Y باشد، آنگاه

$$\mathbb{P}(Y \leq y | X \leq Q_{\alpha}^{+}(X)) = \frac{\mathbb{P}(Y \leq y \wedge X \leq Q_{\alpha}^{+}(X))}{\alpha} = \frac{C(\alpha, F_Y(y))}{\alpha}.$$

بنابراین

$$CoVaR_{\alpha, \beta}^{\leq}(Y|X) = VaR_{w_{*}}(Y),$$

که در آن بزرگترین جواب معاله $w_{*} = w_{*}(\alpha, \beta, C)$

$$C(\alpha, w_{*}) = \alpha\beta, \quad (۴)$$

است. توجه داشته باشید که

$$w_{*} = -CoVaR_{\alpha, \beta}^{\leq}(F_Y(Y)|F_X(X)).$$

قضیه ۱۰۴. [۹] فرض کنید $C_i(u, v), i = 1, 2$ توسط یک مفصل از متغیر تصادفی X_i و Y_i و دارای توابع توزیع پیوسته F_{X_i} و F_{Y_i} و $\alpha, \beta \in (0, 1]$ چند آستانه ثابت تشکیل شده است. اگر

$$\forall (u, v) \in [0, 1]^2 \quad C_1(u, v) \leq C_2(u, v)$$

آنگاه

$$w_{*}(\alpha, \beta, C_1) \geq w_{*}(\alpha, \beta, C_2). \quad (۵)$$

علاوه بر این، اگر

$$\forall t \in (-\infty, +\infty) \quad F_{Y_1}(t) \geq F_{Y_2}(t)$$

آنگاه

$$CoVaR_{\alpha, \beta}^{\leq}(Y_1|X_1) \leq CoVaR_{\alpha, \beta}^{\leq}(Y_2|X_2). \quad (۶)$$

[۹]

نتیجه ۱.۰۴. [۹] فرض کنید C هر مفصل دو بعدی باشد. آنگاه

$$\alpha\beta \leq w_*(\alpha, \beta, C) \leq 1 - \alpha(1 - \beta).$$

گوییم مفصل C دارای وابستگی ربعی مثبت^{۱۲} (PQD) است هرگاه مفصل C بر مفصل استقلال مسلط باشد به عبارت دیگر

$$\forall (u, v) \in [0, 1]^2 \quad C(u, v) \geq uv = \Pi(u, v),$$

هرگاه مفصل C دارای خاصیت PQD باشد، به دلیل اینکه $w_*(\alpha, \beta, C) = \beta$ می توان کران بیان شده در نتیجه ۱.۴ را ساده تر کرد.

نتیجه ۲.۰۴. [۹] فرض کنید C یک مفصل PQD باشد آنگاه

$$\alpha\beta \leq w_*(\alpha, \beta, C) \leq \beta.$$

۵ محاسبه CoVaR اصلاح شده برای مفصل های با وابستگی دمی

در ادامه مثال هایی برای محاسبه ارزش در معرض ریسک شرطی اصلاح شده برای مفصل هایی با وابستگی دمی غیربدیهی، بیان می شود [۴].

مثال ۱.۰۵. برای $\theta > 0$ مفصل C_θ ماردیا-تاکاهاشی-کلیتون^{۱۳} را به صورت زیر در نظر می گیریم.

$$C_\theta^{MTR}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} \quad (۷)$$

^{۱۲}Positively Quadrant Dependence

^{۱۳}Mardia-Takahasi-Claytoncopula

ابتدا ارزش در معرض ریسک شرطی استاندارد، $v_*(\alpha, \beta, C_\theta^{MTR})$ را به صورت زیر به دست می آوریم:

$$\begin{aligned} \frac{D_u C_\theta^{MTC}}{\partial u} |_{u=\alpha} &= (\alpha^{-\theta} + v^{-\theta} - 1)^{-\frac{1+\theta}{\theta}} \alpha^{-\theta-1} = \beta \\ \rightarrow (\alpha^{-\theta} + v^{-\theta} - 1) \alpha^{-1-\theta} &= \beta^{\frac{\theta}{1+\theta}} \rightarrow \alpha^{-\theta} + v^{-\theta} - 1 = \frac{\beta^{\frac{\theta}{1+\theta}}}{\alpha^{-1-\theta}} \\ \rightarrow v^{-\theta} &= \frac{\beta^{\frac{\theta}{1+\theta}}}{\alpha^{-1-\theta}} - \alpha^{-\theta} + 1 \rightarrow v^{-\theta} = \alpha(\beta^{\frac{\theta}{1+\theta}} - 1 + \alpha^\theta) \\ \rightarrow v_* &= \alpha(\beta^{\frac{\theta}{1+\theta}} - 1 + \alpha^\theta)^{-\frac{1}{\theta}} \end{aligned}$$

در ادامه ارزش در معرض ریسک شرطی اصلاح شده، $w_*(\alpha, \beta, C_\theta^{MTR})$ را به صورت زیر به دست می آوریم.

$$\begin{aligned} C_\theta^{MTC}(u, v) |_{u=\alpha} &= \alpha\beta \rightarrow (\alpha^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} = \alpha\beta \rightarrow (\alpha^{-\theta} + v^{-\theta} - 1) = (\alpha\beta)^{-\theta} \\ \rightarrow v^{-\theta} &= (\alpha\beta)^{-\theta} - \alpha^{-\theta} + 1 \rightarrow w_* = ((\alpha\beta)^{-\theta} - \alpha^{-\theta} + 1)^{-\frac{1}{\theta}} = \left(\frac{1}{(\alpha\beta)^{-\theta}} - \frac{1}{\alpha^{-\theta}} + 1\right)^{-\frac{1}{\theta}} \\ \rightarrow \frac{(1 - \beta^{-\theta} + (\alpha\beta)^{-\theta})^{-\frac{1}{\theta}}}{((\alpha\beta)^{-\theta})^{-\frac{1}{\theta}}} &= \alpha\beta(1 - \beta^\theta + (\alpha\beta)^\theta)^{-\frac{1}{\theta}} \rightarrow w_* = \alpha\beta(1 - \beta^\theta + (\alpha\beta)^\theta)^{-\frac{1}{\theta}} \end{aligned}$$

در جدول زیر مقادیر CoVaR استاندارد (v_*) و CoVaR اصلاح شده (w_*) برای مفصل های مخروطی، مقادیر فرین و LTI آورده شده است [۹].

جدول ۱: محاسبه مقادیر CoVaR استاندارد و CoVaR اصلاح شده برای چند مفصل با وابستگی دمی

نام مفصل	تابع مفصل	v_*	w_*
مفصل مخروطی	$\max(ul(\frac{v}{u}), u + v - 1)$	$\alpha I^{[-1]}(\beta)$	$\alpha I^{-1}(\beta)$
مفصل مقادیر فرین	$C_l(u, v) = uv \exp(-\ln(u)l(\frac{\ln(v)}{\ln(u)}))$	$\alpha I^{-1}(\beta) + O(\alpha^\gamma)$	$\alpha I^{-1}(\beta) + O(\alpha^\gamma)$
مفصل LTI	$f(\frac{g(w)}{\alpha}) = \beta$	$f(\alpha I_f^{[-1]}(\beta))$	$f(\alpha g(\beta))$

۶ جمع بندی

محاسبه CoVaR استاندارد در برخی موارد با اشکالاتی چون محاسبه دشوار، صعودی نبودن نسبت به پارامتر وابستگی، همراه است. به همین دلیل به محاسبه CoVaR اصلاح شده برای مفصل ها با وابستگی دمی، پرداخته شد.

مراجع

- [1] Adrian, T. and M.K. Brunnermeier (2008). CoVaR. *Am. Econ. Rev.* **106**(7), 1705-1741.
- [2] Artzner, F. Delbaen, J. M. Eber, and D. Heath, (1999). Coherent Measures of Risk. *Mathematical Finance.* **9**, 203-28
- [3] Di Bernardino, E., Fernández-Ponce, J. M., Palacios-Rodríguez, F., Rodríguez-Griñolo, M. R., (2015). Extremal Quantiles and Value-at-Risk, *The New Palgrave Dictionary of Economics. Second Edition.*(1), 271-292
- [4] Bernardi, M., F. Durante, and P. Jaworski (2017). CoVaR of families of copulas. *Stat. Probabil. Lett.***120**, 8–17.
- [5] Chernozhukov, V., and S. Du (2008). Extremal Quantiles and Value-at-Risk, *The New Palgrave Dictionary of Economics*, Second Edition(1), 271–292
- [6] Embrechts, P. (2009). Copulas: a personal view. *J.Risk Insur.* **76**(3), 639–650.
- [7] Hull, John C (2002). *Risk Management and Financial Institutions*, New Jersey **12**(1), 259–260.
- [8] Jaworski, P., F. Durante, W. Härdle, and T. Rychlik, editors (2010) *Copula Theory and its Applications*. Springer, Heidelberg.
- [9] Jaworski, P. (2017). *On Conditional Value-at-Risk (CoVaR) for tail-dependent copulas* Institute of Mathematics, University of Warsaw, Poland, **5**, 1-19.
- [10] Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall CRC, Boca Raton FL.
- [11] Mainik and Schaanning (2014). *On dependence consistency of CoVaR and some other systemic risk measures* Stat. Risk Model. **31**(1), 49–77.
- [12] Nelsen, R.B. (2006). *An Introduction to Copulas*. Second edition. Springer, New York.
- [13] Pearson, Neil (2002). *Risk Budgeting: Portfolio Problem Solving with Value-at-Risk*. John Wiley & Sons. **ISBN 978-0-471-40556-6**
- [14] Sklar, A. (1959). *Fonctions de répartition a n dimensions et leurs marges*. Publ. Inst.Statist. Univ. Paris **8**, 229–231.



مدل‌بندی وابستگی در داده‌های طولی نامتعادل با استفاده از مفصل‌های جفتی

سفیدی، س ۱ گنجعلی، م ۲ باغفلکی، ت ۳

۱،۲ گروه آمار، دانشگاه شهید بهشتی

۳ گروه آمار، دانشگاه تربیت مدرس

چکیده

یکی از روش‌های مدل‌بندی وابستگی در داده‌های چند متغیره، استفاده از مفصل D- و این است. در این مقاله به بررسی مدل‌بندی وابستگی در داده‌های طولی نامتعادل می‌پردازیم. یکی دیگر از روش‌های مدل‌بندی داده‌های طولی استفاده از مدل آمیخته خطی است. در این مقاله نشان می‌دهیم که مدل مفصل D- و این بسیار انعطاف‌پذیرتر از مدل آمیخته خطی برای داده‌های طولی همراه با همبستگی همگن میان آزمودنی‌ها است. این رویکرد می‌تواند داده‌های گم‌شده از نوع انصراف را بدون نیاز به از بین بردن آنها مدل‌بندی کند. گم‌شدگی در این رویکرد از نوع تصادفی در نظر گرفته شده است. مقایسه مدل‌ها با معیار اطلاع بیزی صورت می‌گیرد. کاربرد داده‌های مطالعه ایدز مورد بررسی قرار می‌گیرد.

کلمات کلیدی: داده‌های طولی نامتعادل، مدل مفصل D- و این، گم‌شدگی از نوع انصراف، داده‌های مطالعه ایدز.

^۱ s_sefidi@sbu.ac.ir

^۲

^۳

۱ مقدمه

اندازه‌گیری‌های تکراری در مطالعات طولی در حوزه‌های بسیاری مانند نجوم، اقتصاد، پزشکی و روانشناسی رواج بسیاری پیدا کرده است. مدل‌های آمیخته خطی ابزاری برای بررسی این‌گونه مطالعات هستند. معرفی بیشتر این مدل‌ها را می‌توان در [۴] و [۱۳] یافت. از طرفی مفصل‌ها به دلیل امکان ترسیم توزیع‌های حاشیه‌ای و ساختار وابستگی از محبوبیت بسیاری برخوردار هستند [۱۱]. در نتیجه مفصل‌ها برای مدل‌بندی داده‌های طولی نیز مورد استفاده قرار گرفته‌اند. این رویکرد ابتدا توسط [۷] استفاده شد. [۶] از مفصل‌گوسی برای مدل‌های چندمتغیره استفاده کردند. [۱۰] به این نتیجه رسیدند که مفصل‌های بیضوی به مراتب مناسبتر از مفصل‌های ارشمیدوسی در مدل‌بندی وابستگی سریالی در زمینه داده‌های طولی هستند.

مفصل‌های D- و این کلاس خاصی از مفصل‌های واین هستند که برای مدل‌بندی وابستگی سریالی مناسب هستند (برای مطالعه بیشتر رجوع کنید به [۲] و [۱]). مدل مفصل D- واین، تابع چگالی را در یک نقطه از زمان به دنباله‌ای از چگالی‌های مفصل دو متغیره و چگالی‌های حاشیه‌ای تجزیه می‌کند که هر یک از این مفصل‌های دو متغیره می‌توانند به دلخواه انتخاب شوند. در نتیجه مدل پیشنهادی مدلی انعطاف‌پذیر خواهد بود. [۱۲] با رویکرد بیزی از این مدل برای داده‌های طولی استفاده نمودند. داده‌های طولی چندمتغیره در [۹] و [۸] مورد بررسی قرار گرفتند. در این مقاله مدل مفصل D- واین را با حاشیه‌های دلخواه برای داده‌های طولی نامتعادل به کار می‌بریم و آن را با مدل آمیخته خطی مقایسه می‌کنیم. معیار مقایسه معیار اطلاع بیزی است.

۲ D- واین برای داده‌های کامل

ابتدا به معرفی ساختار مفصل D- واین برای داده‌های کامل می‌پردازیم. واین‌ها مدل‌های گرافیکی هستند که می‌توانند چگالی مفصل چند متغیره را به حاصل ضرب مفصل‌های دو متغیره موسوم به مفصل‌های جفتی تبدیل کنند. D- واین‌ها زیر کلاسی از واین‌ها هستند که امروزه به دلیل انعطاف‌پذیری و تفسیرپذیری بالا از محبوبیت بسیاری برخوردارند.

فرض کنید بردار تصادفی $U_{1:d} = (U_1, \dots, U_d)$ با توزیع‌های حاشیه‌ای یکنواخت باشد. چگالی D- واین [۳] به صورت زیر تعریف می‌شود:

$$c_{1:d}(u_1, \dots, u_d) = \prod_{l=1}^{d-1} \prod_{k=1}^{d-l} c_{k, k+l; (k+1):(k+l-1)}(C_{k|(k+1):(k+l-1)}(u_k | u_{k+1}, \dots, u_{k+l-1}), \quad (1)$$

$$C_{k+l|(k+1):(k+l-1)}(u_{k+l} | u_{k+1}, \dots, u_{k+l-1}); u_{k+l} | u_{k+1}, \dots, u_{k+l-1}).$$

نویسندگان بسیاری فرض ساده‌سازی را اعمال کردند و ثابت کردند که مفصل جفتی

$$C_{k,k+l;(k+1):(k+l+1)}(\cdot, \cdot; u_{k+1}, \dots, u_{k+l-1})$$

وابسته به مقادیر شرطی کننده $u_{k+1}, \dots, u_{k+l-1}$ نیست. در ادامه فرض می‌کنیم که این مدل پارامتری را می‌توان با خانواده مفصل‌های جفتی

$$C = \{c_{k,k+l;(k+1):(k+l+1)} | k = 1, \dots, d-l, l = 1, \dots, d-1\} \quad (۲)$$

و مجموعه پارامترهای وابسته

$$\theta = \{\theta_{k,k+l;(k+1):(k+l+1)} | k = 1, \dots, d-l, l = 1, \dots, d-1\} \quad (۳)$$

شناسایی نمود.

۳ -D- واین برای داده‌های ناکامل

داده طولی $\mathcal{Y} = \{y^1, \dots, y^n\}$ را در نظر بگیرید که در آن $n \in \mathbb{N}$ تعداد آزمودنی‌ها \mathbb{R}^{d_i} ها $y^i = (y^i_1, \dots, y^i_{d_i})$ مشاهدات فرد i ام با $d_i \in \{1, \dots, d\}$ اندازه‌گیری که $d = \max_{i \in \{1, \dots, n\}} \{d_i\}$ است. فرض کنید برای $j = 1, \dots, d$ تعداد آزمودنی‌هایی است که متغیر پاسخ آن‌ها در j زمان مشاهده شده است و $n_j = 0$ است اگر طول هیچ یک از y^i ها برابر با j نباشد. حال مجموعه داده \mathcal{Y} را به زیر مجموعه‌هایی شامل گروه‌هایی از افراد با تعداد اندازه‌گیری مشابه تقسیم می‌کنیم. برای $j = 1, \dots, d$ ، مشاهدات گروه j را به صورت $\mathcal{Y}^j = \{y^i | i \in I_j\}$ بیان می‌کنیم که $I_j = \{i, y^i \in \mathbb{R}^j\}$ است. در رویکرد پیشنهادی در این مقاله ابتدا باید توزیع حاشیه‌ای تک متغیره F_j^i برای $y_j^i \in \mathbb{R}$ به دست آوریم. مقادیر $F_j^i(y_j^i) \in [0, 1]$ ، $j = 1, \dots, d_i$ ، $i = 1, \dots, n$ در تعیین ساختار وابستگی به کار می‌روند. داده‌های مفصل $\mathcal{U} = \{u^1, \dots, u^n\}$ شامل مشاهدات $u^i = (u^i_1, \dots, u^i_{d_i})' \in [0, 1]^{d_i}$ ، $i = 1, \dots, n$ هستند. مانند تعریف فوق قرار می‌دهیم $\mathcal{U}^j = \{u^i | i \in I_j\}$ که شامل همه مشاهدات به طول j برای $j = 1, \dots, d$ است. حال به معرفی مدل D- واین برای داده‌های ناکامل می‌پردازیم. Killiches: ۲۰۱۷. به دلیل اینکه داده‌ها از اندازه‌گیری‌های تکراری به دست می‌آیند، یک ترتیب زمانی متوالی بین مشاهدات هر آزمودنی وجود دارد. بنابراین استفاده از مفصل D- واین از مرتبه‌های $d, 3, 2, \dots$ پیشنهاد می‌شود. با در نظر گرفتن فرض ساده‌سازی مدل‌های D- واین را برای چگالی‌های مفصل گروه‌های $d, 3, 2, \dots$ j تعریف می‌کنیم.

چگالی مفصل $c_{1:j}^j$ را به کمک $j(j-1)/2$ مفصل‌های جفتی

$$\mathcal{C}^j = \{c_{k,k+l;(k+1):(k+l+1)}^j | k = 1, \dots, j-l, l = 1, \dots, j-1\} \quad (۴)$$

و پارامترهای متناظر

$$\theta^j = \{\theta_{k,k+l;(k+1):(k+l+1)}^j | k = 1, \dots, j-l, l = 1, \dots, j-1\} \quad (۵)$$

برای $j = 2, \dots, d$ معرفی می‌کنیم. برای برآورد \mathcal{C}^j و θ^j از تابع درستنمایی استفاده می‌کنیم که زیرمجموعه‌هایی از \mathcal{U} را در شامل می‌شود که مشاهداتی به طول j دارند. تابع درستنمایی و لگاریتم آن به صورت زیر تعریف می‌شوند:

$$\mathcal{L}_j(\mathcal{C}^j, \theta^j | \mathcal{U}^j) = \prod_{i \in I_j} c_{1:j}^j(u_1^i, \dots, u_j^i | \mathcal{C}^j, \theta^j) \quad (۶)$$

$$\log \mathcal{L}_j(\mathcal{C}^j, \theta^j | \mathcal{U}^j) = \sum_{i \in I_j} \log c_{1:j}^j(u_1^i, \dots, u_j^i | \mathcal{C}^j, \theta^j), \quad (۷)$$

در نتیجه تابع درستنمایی مدل کلی به صورت زیر است

$$\log \mathcal{L}(\mathcal{C}^2, \dots, \mathcal{C}^d, \theta^2, \dots, \theta^d | \mathcal{U}) = \sum_{j=2}^d \log \mathcal{L}_j(\mathcal{C}^j, \theta^j | \mathcal{U}^j), \quad (۸)$$

۱.۳ مثال

برای روشن‌تر شدن بحث فوق به مثالی از داده‌های طولی با حداکثر $d = 4$ مشاهده برای هر آزمودنی توجه کنید. داده‌های طولی ۴ بعدی $\mathcal{U} = \{u^1, \dots, u^n\}$ را در نظر بگیرید که $n = n_2 + n_3 + n_4$ است. حال \mathcal{U} را به گروه‌های $\mathcal{U}^j = \{u^i | i \in I_j\}$ برای $j = 2, 3, 4$ افراز می‌کنیم که $I_2 = \{i | u^i \in \mathbb{R}^2\}$ ، $I_3 = \{i | u^i \in \mathbb{R}^3\}$ و $I_4 = \{i | u^i \in \mathbb{R}^4\}$ است. بنابراین تابع لگاریتم درستنمایی به مجموعه‌ای با ۶ مفصل جفتی به صورت $\theta = (\theta_{1,2}, \theta_{2,3}, \theta_{3,4}, \theta_{1,3;2}, \theta_{2,4;3}, \theta_{1,4;2,3})$ و پارامترهای متناظر $\mathcal{C} = (c_{1,2}, c_{2,3}, c_{3,4}, c_{1,3;2}, c_{2,4;3}, c_{1,4;2,3})$ وابسته است. شکل ۱ نمایش شماتیکی از مدل کامل $c_{1,4}$ نشان می‌دهد. تابع درستنمایی برای این مدل

شکل ۱: D- واین ۴ بعدی

به صورت زیر است:

$$\begin{aligned} \log \mathcal{L}(\mathcal{C}, \theta | \mathcal{U}) &= \sum_{i \in I_r} \log c_{1:2}(u_1^i, u_2^i | c_{1,2}, \theta_{1,2}) \\ &+ \sum_{i \in I_r} \log c_{1:3}(u_1^i, u_2^i, u_3^i | c_{1,2}, c_{2,3}, c_{1,3;2}, \theta_{1,2}, \theta_{2,3}, \theta_{1,3;2}) \\ &+ \sum_{i \in I_r} \log c_{1:4}(u_1^i, u_2^i, u_3^i, u_4^i | c_{1,2}, c_{2,3}, c_{3,4}, c_{1,3;2}, c_{2,4;3}, c_{1,4;2,3}, \theta_{1,2}, \theta_{2,3}, \theta_{3,4}, \theta_{1,3;2}, \theta_{2,4;3}, \theta_{1,4;2,3}) \end{aligned} \quad (9)$$

همان طور که در مدل ۹ می بینیم، مفصل های جفتی و پارامترهای متناظر گروه j به طور مستقل از گروه های دیگر برآورد می شوند. برای افزایش دقت برآوردها این مدل را به فرم زیر تغییر می دهیم:

$$\begin{aligned} \log \mathcal{L}(\mathcal{C}, \theta | \mathcal{U}) &= \sum_{i \in I_r \cup I_r \cup I_r} \log c_{1:2}(u_1^i, u_2^i | c_{1,2}, \theta_{1,2}) \\ &+ \sum_{i \in I_r \cup I_r} [\log c_{2,3}(u_2^i, u_3^i; \theta_{2,3}) + \log_{1,3;2}(C_{1|2}(u_1^i | u_2^i; \theta_{1,2}), C_{3|2}(u_3^i | u_2^i; \theta_{2,3}); \theta_{1,3;2})] \\ &+ \sum_{i \in I_r} [\log c_{3,4}(u_3^i, u_4^i; \theta_{3,4}) + \log c_{2,4;3}(C_{2|3}(u_2^i | u_3^i; \theta_{2,3}), C_{4|3}(u_4^i | u_3^i; \theta_{3,4}); \theta_{2,4;3})] \\ &+ \log_{1,4;2,3}(C_{1|3;2}(C_{1|2}(u_1^i | u_2^i; \theta_{1,2}) | C_{3|2}(u_3^i | u_2^i; \theta_{2,3}); \theta_{1,3;2}), \\ &(C_{4|2;3}(C_{2|3}(u_2^i | u_3^i; \theta_{2,3}) | C_{4|3}(u_4^i | u_3^i; \theta_{3,4}); \theta_{2,4;3}; \theta_{1,4;2,3})) \end{aligned} \quad (10)$$

در این مدل حداکثر استفاده از تمام گروه های ۲ و ۳ و ۴ برای برآورد پارامترها می شود.

۴ مثال کاربردی

در این بخش می خواهیم مدل ارایه شده در بخش قبل را برای مدل بندی مجموعه داده طولی MACS^۱ در [۱۴] به کار گیریم. این مجموعه داده وضعیت بیماری ۲۸۳ فرد مبتلا به ایدز را در طول سال های ۱۹۸۴ تا ۱۹۹۱ مورد بررسی قرار می دهد. متغیر پاسخ این مجموعه داده درصد سلول $CD4$ بیماران بعد از ابتلا به عفونت HIV است. مصرف سیگار با مقادیر ۱ و ۰، سن بیماران و درصد سلول $CD4$ قبل از ابتلا به بیماری سه متغیر کمکی هستند که برای این مطالعه مورد بررسی قرار می گیرند. همه افراد برای انجام ملاقات کلینیکی نیم سالانه برای انجام اندازه گیری درصد سلول های $CD4$ و وضعیت بالینی برنامه ریزی شده بودند. اما بسیاری از افراد بارها ملاقات های برنامه ریزی شده خود را از دست دادند و نتوانستند برای انجام معاینات پزشکی حاضر شوند. به دلیل اینکه تعداد مشاهدات در مراجعات پایانی کم است ما روی ۵ مشاهده اول از هر فرد تمرکز می کنیم. برای دست یافتن به داده مفصل ابتدا با استفاده از مدل آمیخته خطی توزیع حاشیه ای تک متغیره مربوط به هر مشاهده را به دست می آوریم. برای اینکار مدل عرض از مبدا تصادفی را با ساختار همبستگی مناسب به مدل برازش می دهیم. از این مدل برای مقایسه با مدل پیشنهادی استفاده می کنیم.

^۱Multi-center AIDS Cohort Study

جدول ۱: مقادیر AIC ، BIC و $Loglik$

AIC	BIC	$loglik$	
-۵۷۵/۴	-۵۴۶/۲۴	۲۹۵/۷	مدل D- واین کلی
-۵۷۳/۰۹	-۵۵۱/۲۱	۲۹۲/۵۴	مدل D- واین گوسی
۷۷۹/۴	۸۱۵/۸	-۳۷۹/۷	مدل آمیخته خطی

پس از به دست آوردن داده‌های مفصل مدل D- واین ۱۰ را با $d = 5$ برای این مجموعه داده به کار می‌بریم. مدل D- واین ۵ بعدی برای داده کامل به اختصار به صورت زیر نمایش داده می‌شود:

$$f(x_1, \dots, x_5) = \left[\prod_{k=1}^5 f_k(x_k) \right] \cdot c_{1,2} \cdot c_{2,3} \cdot c_{3,4} \cdot c_{4,5} \cdot c_{1,3|2} \cdot c_{2,4|3} \cdot c_{3,5|4} \cdot c_{1,4|2,3} \cdot c_{2,5|3,4} \cdot c_{1,5|2,3,4} \quad (11)$$

برای نمایش انعطاف‌پذیری مدل پیشنهادی، از مدل D- واین گوسی که تمام مفصل‌های جفتی آن از مفصل گوسی استفاده شده است، برای مقایسه با مدل پیشنهادی استفاده می‌کنیم. مدل پیشنهادی با مشاهدات تبدیل یافته در پکیج *VineCopula* اجرا می‌شود. آزمون استقلال برای اجتناب از برآورد پارامتر اضافی در این پکیج انجام می‌شود. معیار انتخاب مفصل جفتی معیار اطلاع بیزی BIC است. در مدل D- واین پیشنهادی برآورد پارامترها در درخت اول برابر با $c_{1,2} = Gaussian(\hat{\tau}_{1,2} = 0.52)$ و $c_{2,3} = survivalGumbel(\hat{\tau}_{2,3} = 0.54)$ ، $c_{3,4} = t-copula(\hat{\tau}_{3,4} = 0.57)$ ، $c_{4,5} = Gaussian(\hat{\tau}_{4,5} = 0.53)$ است و در درخت دوم برابر با $c_{1,3|2} = \hat{\tau}_{1,3|2} = 0.10$ ، $c_{2,4|3} = \hat{\tau}_{2,4|3} = 0.13$ ، $c_{3,5|4} = \hat{\tau}_{3,5|4} = 0.15$ و $c_{1,4|2,3} = \hat{\tau}_{1,4|2,3} = 0.10$ ، $c_{2,5|3,4} = \hat{\tau}_{2,5|3,4} = 0.13$ ، $c_{1,5|2,3,4} = \hat{\tau}_{1,5|2,3,4} = 0.10$ است. در درخت‌های ۳ و ۴ کندال صفر است. در مدل D- واین گوسی برآورد پارامترها در درخت اول برابر با $\hat{\tau}_{1,2} = 0.52$ ، $\hat{\tau}_{2,3} = 0.53$ ، $\hat{\tau}_{3,4} = 0.56$ و $\hat{\tau}_{4,5} = 0.53$ است و در درخت دوم برابر با $\hat{\tau}_{1,3|2} = 0.10$ ، $\hat{\tau}_{2,4|3} = 0.13$ ، $\hat{\tau}_{3,5|4} = 0.15$ و $\hat{\tau}_{1,4|2,3} = 0.10$ است. در هر دو حالت ساختار وابستگی مارکوف مرتبه دوم است زیرا در درخت‌های ۳ و ۴ استقلال وجود دارد. جدول ۱ مقادیر AIC ، BIC و $Loglik$ سه مدل را نشان می‌دهد. عملکرد مدل D- واین کلی به طور یکنواخت بهتر از مدل D- واین محدود شده گوسی و مدل آمیخته خطی است. همچنین به دلیل انعطاف‌پذیری، هر دو مدل بر پایه D- واین عملکردی بهتر از مدل آمیخته خطی دارند.

۵ نتیجه گیری

در این مقاله، ابتدا به معرفی داده‌های طولی ناکامل پرداختیم. سپس مدل مفصل D و این را برای داده کامل و ناکامل توضیح دادیم. تابع درستنمایی مدل D و این را برای مثالی از داده ناکامل با حداکثر ۴ مشاهده با جزئیات بیان کردیم. در نهایت این مدل را برای مجموعه داده طولی با حداکثر ۵ زمان مورد بررسی قرار دادیم. مقایسه این مدل با مدل آمیخته خطی نشان داد که مقدار لگاریتم درستنمایی و AIC و BIC برازش بسیار بهتری را نشان دادند. با استفاده از این مدل میزان وابستگی زمان‌ها به هم با معیار τ کندهال مشخص شدند.

مراجع

- [1] Aas, K., Czado, C., Frigessi, A., and Bakken, H. (2009), *Pair-Copula Constructions of Multiple Dependence*, Insurance: Mathematics and Economics, **44 (2)**, 182-198.
- [2] Bedford, T., and Cooke, R. (2002), *Vines- A New Graphical Model for Dependent Random Variables*, The Annals of Statistics, **30**, 1031-1068.
- [3] Czado, C. (2010). *Pair-copula constructions of multivariate copulas*. In Jaworski, P., Durante, F., Hardle, W. K., and Rychlik, T., editors, *Copula Theory and Its Applications*, pages 93-109. Springer Berlin Heidelberg, Berlin, Heidelberg.
- [4] Diggle, P. J., Heagerty, P., Liang, K. Y. and Zeger, S. L. (2002). *Analysis of longitudinal data (2nd ed)*. Oxford University Press, Oxford, U.K..
- [5] Killiches, M., and Czado, C. (2018). *A D-vine copula based model for repeated measurements extending linear mixed models with homogeneous correlation*. arXiv preprint arXiv: 1705.06261. (to appear in Biometrics).
- [6] Lindsey, J. and Lindsey, P. (2006). *Multivariate distributions with correlation matrices for nonlinear repeated measurements*. Computational Statistics Data Analysis, **50(3)**:720-732.
- [7] Meester, S. G. and MacKay, J. (1994). *A parametric model for cluster correlated categorical data*. Biometrics, **50(4)**:954-963.

- [8] Nai Ruscone, M. and Osmetti, S. A. (2017). *Modelling the dependence in multivariate longitudinal data by pair copula decomposition*. In Ferraro, M. B. e. a., editor, *Soft Methods for Data Science*, pages 373-380. Springer International Publishing, Cham.
- [9] Smith M. (2015). *Copula modelling of dependence in multivariate time series*. *International Journal of Forecasting*, **31**: 815-833.
- [10] Sun, J., Frees, E. W., and Rosenberg, M. A. (2008). *Heavy-tailed longitudinal data modeling using copulas*. *Insurance: Mathematics and Economics*, **42(2)**: 817-830.
- [11] Sklar, A. (1959), *Fonctions de repartition à n dimensions et leurs marges*, *Publications de l'Institut de Statistique de l'Universit de Paris*, **8**, 229-231.
- [12] Smith, M., Min A., Almeida C., Czado C. (2010), *Modeling longitudinal data using a pair-copula decomposition of serial dependence*, *Journal of the American Statistical Association*, **105**, 1467-1479.
- [13] Verbeke, G. and Molenberghs, G. (2009). *Linear mixed models for longitudinal data*. Springer Science Business Media.
- [14] Wu, H., and Zhang, J. T. (2004). *Nonparametric regression methods for longitudinal data analysis*. *Wiley Series in Probability and Statistics*, New York.



پنجمین سمینار
نظریه مفصل و کاربردهای آن

۱۰ و ۱۱ بهمن ۱۳۹۷



تعیین اندازه ریسک در سهام شرکت‌های بیمه بوسیله تابع مفصل

سیف اللهی، س^۱ بیورانی، ح^۲

گروه آمار، دانشکده علوم ریاضی، دانشگاه تبریز

چکیده

از مهمترین اندازه ریسک در مباحث مالی، ریسک دمی شرطی می‌باشد. در این مقاله ریسک دمی شرطی با استفاده از تابع مفصل را ارائه می‌دهیم که در آن ریسک دمی شرطی متغیر با در نظر گرفتن وابستگی آن با متغیر دیگر محاسبه می‌شود. ابتدا ریسک دمی شرطی مفصلی معرفی می‌شود و سپس اندازه ریسک برای مفصل کلایتون و گامیل را مشخص و با در نظر گرفتن مقادیر مختلف برای پارامتر وابستگی مقادیر آن محاسبه و ارائه شده است. در پایان، با در نظر گرفتن داده‌های واقعی مربوط به آخرین قیمت معامله شده روزانه سهام شرکت‌های بیمه آسیا، پارسیان، البرز و ملت در طی دو سال ۲۰۱۷ و ۲۰۱۸ اندازه ریسک معرفی شده حساب و نتایج بدست آمده ارائه شده است.

کلمات کلیدی: امید دمی شرطی، تابع مفصل، اندازه وابستگی، وابستگی ربعی مثبت.

^۱ s.seifollahi.k@gmail.com

^۲ bevrani@gmail.com

۱ پیش‌گفتار

مفهوم ریسک نقش کلیدی در بازارهای مالی ایفا می‌کند از این جهت شناسایی انواع ریسک، اندازه‌گیری و مدیریت آن از اهمیت بالایی برخوردار است. ارزش در معرض خطر (VaR)، امید دمی شرطی (CTE) و اندازه ریسک پنهان (DRM) از جمله معیارهای مهم اندازه‌گیری ریسک بوده که در سال‌های اخیر وارد مباحث مالی شده است.

در تحلیل ریسک CTE به صورت امید شرطی ریسک در صورتی که ریسک از مقدار در معرض خطر آن بیشتر باشد تعریف می‌شود. به عبارت دیگر برای هر مقدار حقیقی $\alpha \in (0, 1)$ ،

$$CTE_X(\alpha) := E(X|X > VaR_X(\alpha))$$

که در آن $VaR_X(\alpha) = \inf\{x; F(x) \geq \alpha\}$. اگر مجبور به کنترل افزایش دو ریسک X_1 و X_2 بطور همزمان باشیم در این صورت نمی‌توان از CTE استفاده کرد لذا نیاز به فرمول جدیدی است، بنابراین با مقدار

$$E(X_1|X_1 > VaR_X(\alpha), X_2 > VaR_X(t)); \quad \forall \alpha, t \in (0, 1)$$

سروکار خواهیم داشت. اگر زوج متغیرهای (X_1, X_2) مستقل از هم باشند، فرمول بالا همان مقدار برای متغیر ریسک خواهد شد و اگر دو متغیر وابسته باشند نیاز به تعیین تابع توزیع توأم می‌باشد. با فرض در دست داشتن توزیع‌های حاشیه‌ای، می‌توان با استفاده از قضیه اسکالر [۳] و استفاده از توابع مفصل توزیع توأم را تعیین کرد.

هدف اصلی مقاله در نظر گرفتن اطلاعات وابستگی ربعی مثبت (PQD) بین ریسک‌هاست تا بتوان اندازه ریسک مالی را حساب کرد. برای این منظور، اندازه ریسک به صورت زیر تعریف می‌شود:

$$CCTE_{X_1}(t) = E(X_1|X_1 > VaR_X(\alpha), X_2 > VaR_X(t)); \quad \forall \alpha, t \in (0, 1) \quad (1)$$

این ریسک را امید دمی شرطی مفصلی تعریف می‌شود. این اندازه ریسک در تمام ویژگی‌های اندازه ریسک منسجم صدق می‌کند.

۲ امید دمی شرطی مفصلی

رایجترین اندازه ریسک در مباحث مالی، VaR و CTE هستند که در این بین VaR برای زمانی که توزیع ریسک نرمال نباشد دارای ایراداتی است. CTE برای متغیر تصادفی پیوسته X به صورت زیر تعریف می‌شود:

$$CTE_X(\alpha) = \frac{1}{1-\alpha} \int_{\alpha}^1 VaR_X(u).du; \quad \forall \alpha \in (0, 1) \quad (2)$$

حال فرض کنید که دو ریسک X_1 و X_2 با توزیع‌های حاشیه‌ای F_{X_i} و توزیع توأم H بهم وابسته باشند. X_1 را ریسک هدف و X_2 را ریسک مرتبط می‌نامیم. در این صورت دیگر فرمول ۳.۳ مفید نخواهد بود، حال با در نظر گرفتن وابستگی بین ریسک هدف و ریسک مرتبط، امید دمی شرطی مفصلی ارائه می‌شود. که در رابطه ۱ داده شده است [۲].

مطابق با قضیه اسکالار [۳]، یک مفصل یکتا $C: [0, 1]^d \rightarrow [0, 1]$ وجود دارد بطوریکه

$$H(x_1, x_2) = C(F_{X_1}(x_1), F_{X_2}(x_2))$$

فرض کنید توابع بقا به صورت زیر تعریف شده باشد:

$$\bar{F}_{X_i}(x_i) = p(X_i > x_i) \quad i: 1, 2$$

$$\bar{H}(x_1, x_2) = p(X_1 > x_1, X_2 > x_2)$$

و تابعی که \bar{F} و \bar{H} را به هم وصل می‌کند را با \bar{C} (مفصل بقا) نشان دهیم، در این صورت \bar{C} یک تابع مفصل است و

$$\bar{C}(u, v) = u + v - 1 + C(u, v)$$

که مفصل X_1 و X_2 می‌باشد. اگر فرض شود که C مطلقاً پیوسته با تابع چگالی c است، داریم

$$\bar{C}(1 - \alpha, 1 - t) = \int_{\alpha}^1 J_t(u) du; \quad \forall \alpha \in (0, 1)$$

که در آن

$$J_t(u) = \int_t^1 c(u, v) dv; \quad \forall \alpha \in (0, 1)$$

همچنین

$$\bar{C}(1 - \alpha, 1 - t) = 1 - \alpha - t + C(\alpha, t).$$

گزاره ۱.۲. فرض کنید (X_1, X_2) یک متغیر تصادفی دو متغیره با توزیع توأمی که با تابع مفصل باشد. برای یک مقدار ثابت α و تمام $t \in (0, 1)$ ، امید دمی شرطی مفصلی به صورت زیر است:

$$CCTE_{X_1}(t) = \frac{\int_{\alpha}^1 J_t(u) F_{X_1}^{-1}(u) du}{\int_{\alpha}^1 J_t(u) du} \quad (3)$$

که در آن $F_{X_1}^{-1}$ ، معکوس تابع توزیع F_{X_1} می باشد [۲].

۳ مثال عددی

۱.۳ CCTE با استفاده از تابع مفصل کلایتون

خانواده کلایتون اولین بار توسط کلایتون [۱] مطرح شده است. فرم مفصل کلایتون برای $\theta > 0$ به صورت زیر تعریف شده است:

$$C_{\theta}^c(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta} \quad (۴)$$

برای مفصل کلایتون داریم:

$$\lambda_l = 2^{-1/\theta}, \quad \lambda_u = 0, \quad \tau = \theta/(\theta + 2)$$

که در آن τ همان کندال تائو می باشد.

فرض کنید توابع بقای حاشیه‌ای‌ها به صورت زیر تعریف شود:

$$\bar{F}_{X_i}(x) = \begin{cases} (1+x)^{-\gamma} & \text{if } x \geq 0 \\ 1 & \text{if } o.w \end{cases} \quad (۵)$$

در این صورت برای هر سطح ثابت α خواهیم داشت:

$$CTE_X(\alpha) = \frac{\gamma}{\gamma-1} (1-\alpha)^{-1/\gamma} - 1, \quad VaR_X(\alpha) = (1-\alpha)^{-1/\gamma} - 1 \quad (۶)$$

$$CCTE_X(t) = \frac{1}{\bar{C}_{\theta}^c(1-\alpha, 1-t)} \left((1-\alpha)CTE_X(\alpha) - \int_{\alpha}^1 \frac{u^{-\theta-1}((1-\alpha)^{-1/\gamma} - 1)}{(u^{-\theta} + t^{-\theta} - 1)^{1+1/\theta}} du \right) \quad (۷)$$

با استفاده از روابط بین پارامتر وابستگی و کندال تائو، می توان مقدار پارامتر وابسته را برای وابستگی ضعیف، متوسط و قوی انتخاب تا مقدار CCTE را محاسبه و مقایسه کرد.

جدول ۲ اندازه ریسک CCTE را به ازای مقادیر مختلف α و t نمایش می دهد. مقایسه مقادیر بدست آمده نشان می دهد که مقادیر ریسک مربوط به متغیر X_2 بیشتر از دو متغیر دیگر می باشد، بنابراین این متغیر بطور واضح خطرناکتر از X_1 و X_2 است. مفصل کلایتون برای کاربردهایی مناسب است که دو پیشامد مقادیر کم را با هم تجربه کنند، چون وابستگی در دم پایین قوی و در دم بالا ضعیف است.

جدول ۱: مقدار وابستگی دمی، کندال تائو و پارامتر مفصل کلایتون

τ	θ_i	λ_i
۰,۲	۰,۵	۰,۲۵۰
۰,۵	۲	۰,۷۰۷
۰,۸۵۷	۱۲	۰,۹۴۳

۲.۳ CCTE با استفاده از تابع مفصل گامبل

این خانواده توسط گامبل [۴] معرفی شده است، اما چون در [۵] مورد بحث قرار گرفته است، به عنوان خانواده گامبل- هوگل هم معروف است. مفصل گامبل به صورت زیر می‌باشد:

$$C_{\theta}^G(u, v) = \exp\{-((- \ln u)^{\theta} + (- \ln v)^{\theta} - 1)^{-1/\theta}\} \quad \theta \in [1, \infty) \quad (۸)$$

برای این مفصل داریم

$$\lambda_l = ۰, \quad \lambda_u = ۲ - ۲^{-1/\theta}, \quad \tau = (\theta - 1)/\theta$$

این نتایج نشان می‌دهد که اگر مشاهدات نشان دهنده وابستگی قوی برای مقادیر بزرگتر و وابستگی ضعیف برای مقادیر کوچکتر باشد، مفصل گامبل انتخاب مناسبی خواهد بود.

با فرض اینکه توابع بقا حاشیه‌ای به فرم ۵ باشد، در این صورت اندازه ریسک امید دمی شرطی به صورت زیر خواهد بود:

$$CCTE_X(t) = \frac{1}{\bar{C}_{\theta}^G(1 - \alpha, 1 - t)} ((1 - \alpha)CTE_X(\alpha) - \int_{\alpha}^1 \frac{((1 - u)^{-1/\theta} - 1)C_{\theta}^G(u, v)(- \ln u)^{\theta-1}}{u((- \ln u)^{\theta} + (- \ln v)^{\theta} - 1)^{-1/\theta}} du) \quad (۹)$$

که در آن

$$\bar{C}_{\theta}^G(1 - \alpha, 1 - t) = 1 - \alpha - t + C_{\theta}^G(\alpha, t) \quad (۱۰)$$

با استفاده از روابط بین پارامتر وابستگی و کندال تائو، می‌توان مقدار پارامتر وابسته را برای وابستگی ضعیف، متوسط و قوی انتخاب تا مقدار CCTE را محاسبه و مقایسه کرد.

جدول ۳ اندازه ریسک $CCTE$ را به ازای مقادیر مختلف α و t نمایش می‌دهد. مقایسه مقادیر بدست آمده نشان می‌دهد متغیر X_3 بطور واضح خطرناکتر از X_1 و X_2 است.

۴ کاربرد

برای ارائه کاربرد واقعی از اندازه ریسک تعریف شده، از بین سهام شرکت‌های بیمه، چهار سهام یعنی بیمه آسیا، پارسیان، البرز و ملت را انتخاب و داده‌های مربوط به آخرین قیمت معامله شده روزانه آنها را برای دو سال ۲۰۱۷ و ۲۰۱۸ مورد تحلیل قرار می‌دهیم. برآورد ضریب همبستگی این چهار متغیر در جدول ۵ نمایش داده شده است.

جدول ۶ نتایج آزمون نیکویی برازش دو متغیره با استفاده از پکیج gofCopula در نرم افزار R برای سه تابع مفصل گاوس، کلایتون و گامبل برای چهار متغیر انتخاب شده را نشان می‌دهد و بر اساس بیشترین p مقدار بدست آمده، توابع مفصل را برای متغیرها انتخاب کردیم.

اکثر تابع مفصل انتخاب شده برای متغیرها، مفصل گاوس می‌باشد که به صورت زیر است:

$$c_\rho(u, v) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{\frac{\rho xy - \rho^2(x^2 + y^2)}{2(1-\rho^2)}\right\}$$

که در آن $x = F^{-1}(u)$ و $y = F^{-1}(v)$ و ρ پارامتر وابستگی مفصل گاوس می‌باشد. بعد از انتخاب مفصل مناسب برای هر دو متغیر، پارامتر وابستگی آن با استفاده از رابطه بین پارامتر وابستگی و کندال تاو تعیین و در جدول ۷ نشان داده‌ایم.

برازش توزیع برای متغیرها نشان داد که متغیرها دارای توزیع پارتو می‌باشند که برآورد پارامترهای آن را به روش حدکثر درستمایی تعیین کردیم. اندازه ریسک معرفی شده $CCTE$ برای مقادیر $\alpha = 0.90$ و $t = 0.90$ با استفاده از مفاصل انتخاب شده و به کمک توزیع پارتو محاسبه و خلاصه نتایج را در جدول ۸ نمایش دادیم. با توجه به مقادیر بدست آمده مشخص شد که ریسک مربوط به زوج (بیمه البرز و بیمه ملت) دارای کمترین ریسک می‌باشد که در آن بیمه البرز ریسک هدف و بیمه ملت ریسک مرتبط می‌باشد.

مراجع

- [1] Clayton, D.G. (1978), *A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence*, Biometrika 65, 141–151.
- [2] Brahim, b. Benatia, F. and Yahia, (2018) D. *Copula conditional tail expectation for multivariate financial risks*, Arab J Math Sci 24 (1), 82–100.
- [3] Sklar, A. (1959), *Fonctions de repartition a n dimensions et leurs marges*, Publ. Inst. Statist. Univ. Paris 8. 22–231.
- [4] Gumbel, E.J. (1960), *Bivariate exponential distributions*, J. Amer. Statist. Assoc. 55, 698–707.
- [5] [25] Hougaard, P.A (1986), *class of multivariate failure time distributions*, Biometrika 73, 671–678.

جدول ۲: اندازه ریسک با $\gamma = 1/5$ با در نظر گرفتن مفصل کلایتون

۰,۹۹۰۰	۰,۹۶۷۵	۰,۹۴۵۰	۰,۹۲۲۵	۰,۹۰۰۰	
۲۰,۵۴۴۳۵	۸,۸۱۹۲۳۷	۵,۹۱۴۴۵۹	۴,۵۰۱۳۰۳	۳,۶۴۱۵۸۹	VaR
۶۳,۶۳۳۰۴	۲۸,۴۵۷۷۱۲	۱۹,۷۴۳۳۷۷	۱۵,۵۰۳۹۰۸	۱۲,۹۲۴۷۶۷	CTE
$\theta = 0.5$					
۰,۹۹۰۰	۰,۹۶۷۵	۰,۹۴۵۰	۰,۹۲۲۵	۰,۹۰۰۰	t
۶۳,۷۰۶۰۳	۲۸,۵۶۶۹۳	۱۹,۸۷۴۹۱	۱۵,۶۵۲۹۸	۱۳,۰۸۸۸۰	۰,۹۰۰۰
۶۳,۷۰۷۸۷	۲۸,۵۶۹۶۹	۱۹,۸۷۸۲۲	۱۵,۶۵۶۶۵	۱۳,۰۹۲۹۳	۰,۹۲۲۵
۷۰,۹۷۳.	۲۸,۵۷۲۴۳	۱۹,۸۸۱۵۲	۱۵,۶۶۰۳۹	۱۳,۰۹۷۰۱	۰,۹۴۵۰
۶۳,۷۱۱۵۶	۲۸,۵۷۵۲۰	۱۹,۸۸۴۶۵	۱۵,۶۶۴۱۶	۱۳,۱۰۱۰۸	۰,۹۶۷۵
۶۳,۷۱۳۱۶	۲۸,۵۷۷۹۳	۱۹,۸۸۸۱۷	۱۵,۶۶۷۶۷	۱۳,۱۰۵۱۷	۰,۹۹۰۰
$\theta = 2$					
۰,۹۹۰۰	۰,۹۶۷۵	۰,۹۴۵۰	۰,۹۲۲۵	۰,۹۰۰۰	t
۶۳,۸۸۲۶۷	۲۸,۸۳۳۷۷	۲۰,۱۹۹۲۴	۱۶,۰۲۳۸۳	۱۳,۵۰۰۶۷	۰,۹۰۰۰
۶۳,۸۹۸۷۴	۲۸,۸۵۷۷۳	۲۰,۲۲۷۹۵	۱۶,۰۵۶۲۲	۱۳,۵۳۶۱۳	۰,۹۲۲۵
۶۳,۹۱۵۳۶	۲۸,۸۸۲۴۶	۲۰,۲۵۷۵۸	۱۶,۰۸۹۵۸	۱۳,۵۷۲۶۰	۰,۹۴۵۰
۶۳,۹۳۲۵۲	۲۸,۹۰۷۹۸	۲۰,۲۸۸۰۹	۱۶,۱۲۳۹۷	۱۳,۶۱۰۱۵	۰,۹۶۷۵
۶۳,۹۵۰۲۸	۲۸,۹۳۴۲۹	۲۰,۳۱۹۸۱	۱۶,۱۵۹۴۱	۱۳,۶۴۸۷۷	۰,۹۹۰۰
$\theta = 12$					
۰,۹۹۰۰	۰,۹۶۷۵	۰,۹۴۵۰	۰,۹۲۲۵	۰,۹۰۰۰	t
۶۴,۱۶۹۰۱	۲۹,۳۳۱۳۱	۲۰,۸۸۸۳۶	۱۶,۹۱۳۴۵	۱۴,۶۰۵۱۵	۰,۹۰۰۰
۶۴,۳۶۳۵۵	۲۹,۶۳۷۷۵	۲۱,۲۷۴۱۴	۱۷,۳۶۶۷۹	۱۵,۱۱۸۰۲	۰,۹۲۲۵
۶۴,۶۱۹۲۰	۳۰,۰۳۳۲۸	۲۱,۷۶۲۷۶	۱۷,۹۳۰۱۵	۱۵,۷۴۳۶۹	۰,۹۴۵۰
۶۴,۹۵۱۸۰	۳۰,۵۳۶۹۱	۲۲,۳۷۱۹۸	۱۸,۶۱۸۷۸	۱۶,۴۹۴۸۲	۰,۹۶۷۵
۶۵,۳۸۰۲۴	۳۱,۱۶۹۴۵	۲۳,۱۱۹۹۵	۱۹,۴۴۷۶۲	۱۷,۳۸۳۷۳	۰,۹۹۰۰

جدول ۳: اندازه ریسک با $\gamma = 1/5$ با در نظر گرفتن مفصل گامبل

۰,۹۹۰۰	۰,۹۶۷۵	۰,۹۴۵۰	۰,۹۲۲۵	۰,۹۰۰۰	
۲۰,۵۴۴۳۵	۸,۸۱۹۲۳۷	۵,۹۱۴۴۵۹	۴,۵۰۱۳۰۳	۳,۶۴۱۵۸۹	VaR
۶۳,۶۳۳۰۴	۲۸,۴۵۷۷۱۲	۱۹,۷۴۳۳۷۷	۱۵,۵۰۳۹۰۸	۱۲,۹۲۴۷۶۷	CTE
$\theta = 1.1$					
۰,۹۹۰۰	۰,۹۶۷۵	۰,۹۴۵۰	۰,۹۲۲۵	۰,۹۰۰۰	t
۷۱,۹۹۲۷۵	۳۲,۵۵۶۹۷	۲۲,۶۹۹۰۶	۱۷,۸۷۹۳۳	۱۴,۹۳۷۰۲	۰,۹۰۰۰
۷۴,۱۳۳۹۴	۳۳,۶۶۷۲۹	۲۳,۵۰۷۶۴	۱۸,۵۲۸۹۰	۱۵,۴۸۵۰۰	۰,۹۲۲۵
۷۷,۶۴۵۲۴	۳۵,۵۳۴۹۶	۲۴,۸۷۳۷۹	۱۹,۶۲۶۷۱	۱۶,۴۱۰۲۶	۰,۹۴۵۰
۸۴,۷۲۶۵۲	۳۹,۴۵۴۷۵	۲۷,۷۶۰۶۶	۲۱,۹۴۸۷۵	۱۸,۳۶۵۹۲	۰,۹۶۷۵
۱۱۱,۱۸۶۸۱	۵۵,۲۷۵۶۸	۳۹,۵۸۸۰۸	۳۱,۵۰۷۳۴	۲۶,۴۲۹۳۰	۰,۹۹۰۰
$\theta = 2$					
۰,۹۹۰۰	۰,۹۶۷۵	۰,۹۴۵۰	۰,۹۲۲۵	۰,۹۰۰۰	t
۶۳,۳۱۱۵۲	۳۰,۸۴۹۰۹	۲۲,۸۴۲۱۲	۱۹,۲۰۹۲۰	۱۷,۱۵۸۱۸	۰,۹۰۰۰
۶۵,۶۰۶۳۵	۳۱,۷۶۶۷۴	۲۴,۰۵۹۷۲	۲۰,۶۵۳۶۰	۱۸,۷۶۹۳۰	۰,۹۲۲۵
۶۶,۵۸۳۴۲	۳۳,۵۴۳۷۴	۲۶,۳۸۳۷۰	۲۳,۳۳۸۵۷	۲۱,۶۹۱۱۵	۰,۹۴۵۰
۶۹,۰۷۴۷۴	۳۸,۱۲۸۴۰	۳۲,۰۷۰۷۸	۲۹,۶۰۷۵۸	۲۸,۲۹۳۵۲	۰,۹۶۷۵
۸۵,۳۸۵۳۴	۶۳,۳۱۱۵۲	۵۹,۵۹۹۳۱	۵۸,۰۸۹۹۸	۵۷,۲۷۲۵۹	۰,۹۹۰۰
$\theta = 10$					
۰,۹۹۰۰	۰,۹۶۷۵	۰,۹۴۵۰	۰,۹۲۲۵	۰,۹۰۰۰	t
۶۳,۶۳۳۰۴	۲۸,۴۵۷۷۶	۱۹,۷۴۷۹۶	۱۵,۵۹۱۸۵	۱۳,۵۷۳۰۱	۰,۹۰۰۰
۶۳,۶۳۳۰۴	۲۸,۴۵۸۲۹	۱۹,۷۹۴۸۵	۱۶,۲۸۲۹۹	۱۵,۵۴۴۶۴	۰,۹۲۲۵
۶۳,۶۳۳۰۴	۲۸,۴۷۲۰۹	۲۰,۷۳۵۹۴	۱۹,۷۲۶۴۵	۱۹,۶۷۲۱۶	۰,۹۴۵۰
۶۳,۶۳۳۱۳	۲۹,۸۸۵۹۶	۲۸,۳۶۳۱۶	۲۸,۳۴۶۵۲	۲۸,۳۴۵۸۶	۰,۹۶۷۵
۶۶,۸۰۷۲۵	۶۳,۳۸۱۳۸	۶۳,۳۸۱۲۷	۶۳,۳۸۱۲۶	۶۳,۳۸۱۲۶	۰,۹۹۰۰

جدول ۴: مقدار وابستگی دمی، کندال تائو و پارامتر مفصل گامبل

τ	θ_i	λ_u
۰,۰۱۳	۱,۰۱	۰,۰۰۹
۰,۵۸۵	۲	۰,۵۰۰
۰,۹۲۸	۱۰	۰,۹۰۰

جدول ۵: برآورد کندال تائو برای سهام چهار شرکت بیمه

متغیرها	آسیا	پارسیان	البرز	ملت
آسیا	۱			
پارسیان	۰,۳۹۱۸	۱		
البرز	۰,۴۸۳۵۹	۰,۶۷۷۲	۱	
ملت	۰,۳۴۶۴	۰,۰۷۲۵۱	۰,۲۴۲۵	۱

جدول ۶: مقدار p-value آزمون نیکویی برازش براساس بوت استرپ با بعد دو به روش Sn

مفصل	پارسیان	البرز	ملت
گاوس	۰,۶۱۶	۰,۲۱۲	۰,۱۸۱
کلایتون	۰,۲۶۰	۰,۷۹۷	۰,۱۷۵
گامبل	۰,۲۳۲	۰,۸۱۳	۰,۱۷۱
گاوس	-	۰,۱۶۵	۰,۰۷۷
کلایتون	-	۰,۰۰۸	۰,۱۲۱
گامبل	-	۰,۰۰۲	۰,۱۱۶
گاوس	-	-	۰,۲۷۰
کلایتون	-	-	۰,۲۵۷
گامبل	-	-	۰,۲۴۶

جدول ۷: برآورد پارامترهای وابستگی مفصل انتخاب شده با استفاده از برآوردهای کندال تائو

متغیرها	آسیا	پارسیان	البرز	ملت
آسیا	-	۰,۵۷۳۶۲	۱,۸۷۲۸۹	۰,۵۱۷۷۰
پارسیان	۰,۵۷۳۶۲	-	۰,۸۷۴۱۶	۰,۱۵۶۳۸
البرز	۱,۸۷۲۸۹	۰,۸۷۴۱۶	-	۰,۳۷۱۸۰
ملت	۰,۵۱۷۷۰	۰,۱۵۶۳۸	۰,۳۷۱۸۰	-

جدول ۸: اندازه ریسک *CCTE* با $\alpha = ۰,۹۰$ و $t = ۰,۹۰$

متغیر	آسیا	پارسیان	البرز	ملت
آسیا	-	۵۲۰۴,۹۲۱	۵۲۱۸,۰۰۹	۵۲۹۹,۳۶۹
پارسیان	۳۶۷۵,۶۷۸	-	۳۶۰۱,۸۸۹	۳۶۳۱,۴۳۳
البرز	۲۵۱۹,۷۵۹	۲۶۰۷,۱۹۳	-	۲۴۹۷,۱۹۴
ملت	۲۷۱۳,۰۹۸	۲۶۹۳,۳۰۵	۲۷۱۳,۷۶۱	-



پنجمین سمینار
نظریه مفصل و کاربردهای آن
۱۰ و ۱۱ بهمن ۱۳۹۷



انتخاب مفصل با استفاده از معیارهای نظریه اطلاع

شیبانی،^۱ دهقان، م.ح.^۲ احمدزاده، ح.^۳

گروه آمار، دانشکده ریاضی، دانشگاه سیستان و بلوچستان

چکیده

امروزه ساختار همبستگی بین متغیرها و شناخت آنها از مسائل مهم و مورد نیاز جامعه علمی می‌باشد، این مهم را می‌توان با استفاده از توابع مفصل حل کرد. از آنجا که مفصل‌های متنوعی وجود دارد که بر مجموعه‌ای از داده‌ها برازش یابند، انتخاب بهترین مفصل یکی از مسائل روز می‌باشد. در این راستا برای انتخاب بهترین مدل مفصل از روش‌های CF ، AIC و KL استفاده می‌شود و معیارهای فوق‌الذکر بر روی داده‌های واقعی اعمال و نتایج مورد بررسی قرار می‌گیرد.

کلمات کلیدی: آمیخته مقیاسی بتا پرایم، توابع فوق هندسی، وابستگی دمی، مرتبه دمی

^۱ sheybanielahe@gmail.com

^۲ mhdehghan@math.usb.ac.ir

^۳ hahmadzadeh@math.usb.ac.ir

۱ پیش‌گفتار

گفته می‌شود احتمالاً تاریخ مفصل^۱ در سال ۱۹۵۱ توسط فرشه^۲ معرفی شده است. او مسأله زیر را برای حالت دو بعدی مطالعه کرد:

با توابع توزیع F_1 و F_2 از متغیرهای تصادفی X_1 و X_2 در فضای احتمالی یکسان (Ω, F, p) ، در مورد مجموعه $\Gamma(F_1, F_2)$ از تابع چگالی دو متغیره‌ای که حاشیه‌های آن F_1 و F_2 است، چه می‌توان گفت؟ لازم به ذکر است که مجموعه $\Gamma(F_1, F_2)$ ، که در حال حاضر کلاس فرشه F_1 و F_2 نامیده می‌شود، غیر تهی بوده است. اگر X_1 و X_2 مستقل نباشند، آنگاه تابع توزیع

$F(x_1, x_2) = F_1(x_1)F_2(x_2)$ همیشه به $\Gamma(F_1, F_2)$ متعلق است اما دیگر عناصر $\Gamma(F_1, F_2)$ به وضوح نیست. مطالعات مقدماتی درباره این مسأله را فرون^۳، فرشه^۴ ۱۹۵۶ و گمبل^۴ ۱۹۵۸ انجام دادند. در سال ۱۹۵۹، اسکالر^۵ [۳] با ارائه مفهوم و نام مفصل، اثبات قضیه‌ای که پس از آن به نام او شناخته می‌شود، بهترین نتیجه را در این زمینه بدست آورد. او، برت^۶ و شوایزر^۷ در کار خود بر فضای متریک آماری پیشرفت کرده بودند، تا حدی که منگر^۸ ارتباط دادن نتایج خود به فرشه را ارزشمند دانست. فرشه علاقه‌مند بود و از او خواست تا بیانیه‌ای را برای حساب‌های ارائه شده بنویسد. این امر منجر به تبادل نامه بین اسکالر و فرشه شد که

این کار باعث تبادل نامه‌هایی در رابطه با آنچه او برای من چندین پاکت از پرینت‌های مربوطه به کار او و هم‌دانشگاهیان‌ش روی توزیع حاشیه‌ای ارسال کرده است با فرشه شد. این مقالات برای بسیاری از کارهای بعدی مهم بودند. با این حال، در آن زمان، مهمترین مقاله برای اسکالر مقاله‌ی فرون ۱۹۵۶ بود. فرون در مطالعه توزیع‌های سه‌بعدی توابع کمکی را که بر روی مکعب واحد تعریف شده بودند، را معرفی کرد که این توزیع‌ها را با حاشیه‌های یک‌بعدی مرتبط می‌داند. اسکالر مشاهده کرد که توابع مشابه را می‌توان در واحد n -مکعبی^۹ برای تمام $n \geq 2$ تعریف کرد و به طور مشابه توزیع n -بعدی را به حاشیه‌های یک‌بعدی پیوند می‌دهد. او پس از آنکه ویژگی‌های اساسی این توابع را بررسی کرد، در مورد آنها به فرشه نامه‌ای نوشت. فرشه از اسکالر خواست تا در مورد آنها یادداشتی به زبان فرانسه بنویسد. در حین نگارش این مقاله، اسکالر به

^۱ Copula

^۲ Frechet

^۳ Feron

^۴ Gumbel

^۵ Sklar

^۶ Bert

^۷ Schweizer

^۸ Menger

^۹ n-cube

این نتیجه رسید که برای این توابع یک نام نیاز دارد. از آنجایی که دانستن کلمه‌ی "مفصل" به عنوان یک عبارت دستوری، برای یک کلمه یا عبارت که یک موضوع و گزاره را پیوند می‌دهد، لذا او احساس کرد که این یک نام مناسب برای تابعی خواهد بود که یک توزیع چند بعدی را به حاشیه‌های یک بعدی پیوند می‌دهد و از آن استفاده می‌شود. فرشه یادداشت اسکالر را دریافت کرد، یک گزاره ریاضی را اصلاح کرد و چند تصحیح جزئی بر متن فرانسه او اعمال کرد و یادداشت‌های او را در موسسه آماری دانشگاه پاریس منتشر کرد. جالب است که در این مقاله، نویسنده "ابی اسکالر"^{۱۰} به عنوان "م. اسکالر"^{۱۱} نامیده شد. (آیا باید به عنوان آقا ۱۲ انتخاب شود؟). اثبات قضیه اسکالر در ۱۹۵۹ ارائه نشد، اما طرحی از آن در ۱۹۷۳ ارائه شد. به طوری که برای چند سال افراد حرفه‌ای در این زمینه باید آن را با اتکا به یادداشت‌های نوشته شده توسط خود اسکالر بازسازی می‌کردند. همچنین باید اشاره کرد که برخی اثبات‌های "غیرمستقیم" قضیه اسکالر (بدون اشاره به مفصل) بعداً توسط مور^{۱۳}، اسپردیل^{۱۴} و دلایو^{۱۵} کشف شدند. در حدود ۱۵ سال تمام نتایج مربوط به مفصل‌ها در چارچوب نظریه فضای متریک احتمال توسط شوایزر و اسکالر^{۱۹۸۳} به دست آمده است. رویدادی که به نفع جامعه آماری روی مفصل‌ها در اواسط دهه هفتاد به طور کاملاً تصادفی رخ داد. آن بود که، یک مقاله با عنوان معیارهای وابستگی توسط رنی^{۱۶} ۱۹۵۹ را خواند و متوجه این موضوع شد که او به راحتی می‌تواند با استفاده از مفصل چنین اقداماتی را انجام دهد. اولین بلوک‌های ساختاری توسط شوایزر و ولف^{۱۷} ۱۹۸۱ و پایان‌نامه دکتری ولف در دانشگاه ماساچوست^{۱۸} امهرست^{۱۹} ۱۹۹۷ اعلان شد. این نتایج در جامعه آماری در مقاله شوایزر و ولف ۱۹۸۱ ارائه و با مقاله‌ی ولف ۱۹۸۰ مقایسه شده است. با این حال برای چندین سال دیگر، فصل ۶ کتاب بنیادی توسط شوایزر و اسکالر^{۱۹۸۳}، به نظریه‌ی فضاهای متریک احتمال اختصاص داده شده و در سال ۱۹۸۳ منتشر شده، اما منبع اصلی اطلاعات اولیه در مورد مفصل‌ها بود. بعد از انتشار این مقالات و کتاب، دانشجویان و همکاران سریعتر مشغول شدند. علاوه بر این از آنجا که علاقه به سوالات وابستگی آماری افزایش یافته است، سایرین از جهات مختلف به این موضوع روی آوردند. در سال ۱۹۸۶، مقاله پرطرفدار "the joy of copulas" توسط جنست^{۲۰} ۱۹۸۶ و مک‌کی^{۲۱}، توجه بیشتری را به خود جلب کرد. در ۱۹۹۰، دلایو اولین کنفرانس اختصاص داده شده به

^{۱۰} Abe Sklar^{۱۱} M. Sklar^{۱۲} Monsieur^{۱۳} Moore^{۱۴} Spruill^{۱۵} Deheuvels^{۱۶} A. Renyi^{۱۷} Wolff^{۱۸} Massachusetts^{۱۹} Amherst^{۲۰} Genest^{۲۱} MacKay

مفصل‌ها را سازماندهی کرد که "توزیع احتمال با حاشیه‌های داده شده" نامیده شد. این کنفرانس اولین بار در مجموعه‌ای از کنفرانس‌هایی برگزار شد که تا حد زیادی به این فرضیه کمک کرد، زیرا هریک از آنها شانس ارئه نتایج یک شخص را پیشنهاد کرده بودند، این کنفرانس‌ها در شهر سیاتل^{۲۲} توسط روشندرف^{۲۳} و شوایزر و تیلور^{۲۴} ۱۹۹۶، در شهر پراگ^{۲۵} توسط بنس^{۲۶} و استفن^{۲۷} ۱۹۹۷، در بارسلون^{۲۸} توسط کوادرس^{۲۹} و فورتیان^{۳۰} ۲۰۰۲، در کبک^{۳۱} توسط جنست^{۳۲} ۲۰۰۴، و در تارتو^{۳۲} در سال ۲۰۰۹ توسط کلو^{۳۳} برگزار شدند و برنامه‌ی بعدی قرار بود در سال ۲۰۱۰ در ساووپائولو^{۳۴} شرکت کند. در پایان دهه نود، ایده مفصل‌ها به طور فزاینده‌ای محبوب شد. دو کتاب درباره‌ی مفصل رونمایی شدند و برای دهه بعدی به مراجع استاندارد تبدیل شد. در سال ۱۹۹۷ جو^{۳۵} کتاب خود را در مدل‌های چندمتغیره نوشت و قسمت بزرگ آن را به بحث درباره مفصل‌ها منتشر کرد (با برخی نتایج جدید توسط نلسن^{۳۶} ۲۰۰۶ به چاپ رسید). اما دلیل اصلی این افزایش علاقه را باید در کشف ایده‌های مفصلی توسط محققان در چندین زمینه کاربردی مانند امور مالی یافت. همانطور که تا کنون دیده‌ایم، مفهوم مفصل هم طبیعی است و هم برای بررسی توابع چگالی چند متغیره آسان است. چرا ما شاهد رشد باور نکردنی در مقالات منتشر شده از پایان دهه نود هستیم؟ در دهه هشتاد و دهه نود، یک توسعه از روش مدیریت ربا با ریسک کمی در امور مالی و بیمه را تجربه کردیم، که بسیاری از آنها توسط دستورالعمل‌های تنظیمی جدید یا توسعه محصولات جدید رانده شدند. پیدایش مفصل‌ها در امور مالی توسط جنست^{۳۲} ۲۰۰۹ از ارزش تحقیقاتی در مورد مفصل‌ها به خصوص کاربرد مفصل‌ها سرچشمه گرفته است. در همان زمان، زمینه‌های مختلف نظیر هیدرولوژی جنست^{۳۲} ۲۰۰۷، اهمیت این مفهوم را برای ساخت مدل‌های چندمتغیره را با انعطاف‌پذیری بیشتری کشف کردند. امروزه، ارائه یک حساب کامل از همه‌ی کاربردهای مفصل به حوزه‌های بسیاری که در آن از آنها استفاده می‌شود، تقریباً غیرممکن است. همانطور که شوایزر گفت: عصر توزیع‌های مستقل و هم توزیع پایان یافته است، وقتی که وابستگی بطور جدی وجود دارد، مفصل‌ها به‌طور طبیعی وارد عمل می‌شوند. این امر برای جامعه آماری در مقیاس بزرگ برای تشخیص

^{۲۲}Seattle^{۲۳}Ruschendorf^{۲۴}Taylor^{۲۵}Prague^{۲۶}Benes^{۲۷}Stepan^{۲۸}Barcelona^{۲۹}Cuadras^{۳۰}Fortiana^{۳۱}Quebec^{۳۲}Tartu^{۳۳}Kollo^{۳۴}Sao Paulo^{۳۵}Joe^{۳۶}Nelsen

این واقعیت باقی می‌ماند و زمانی که هر عبارت آماری شامل یک بخش یا فصل در مفصل‌ها باشد موضوع به سن و سال خواهد رسید. اخیراً انتقادات زیادی در مورد مفصل‌ها و کاربردهایشان مطرح شده است و چند نفر درباره‌ی "جنون مفصل" شروع به صحبت کردند. از دیدگاه ما، این انتقادات واکنشی به چنین انتشار گسترده‌ی کاربردهای مفصلی کاملاً طبیعی بود. باید گفت که چندین نفر به اشتباه مفصل‌ها را به‌عنوان راه‌حل "همه‌ی مشکلات وابستگی اتفاقی" تعبیر کرده‌اند، قطعاً اینطور نیست! مفصل‌ها ابزاری ضروری برای درک مشکلات متعدد در مورد وابستگی تصادفی هستند، اما آنها راه‌حل مشکلات همه مدل‌های تصادفی نیستند. علی‌رغم این گستره وسیع مورد توجه در مورد مفصل‌ها، ما هنوز بر این باوریم که این مفهوم هنوز در ابتدای راه قرار دارد و چندین تحقیق دیگر ممکن است و باید به منظور تاکید بر این‌که آیا مفاهیم یا مفاهیم مبنی بر مفصل، واقعا می‌تواند در چندین کاربرد به‌عنوان یک مفهوم قوی ریاضی در نظر گرفته شود.

اگرچه امروزه مدل‌سازی پیشامدهای تصادفی توأم اغلب ضروری و به ندرت صریح است. مدل‌های مفصل یک روش متداول مدل‌سازی متغیرهای تصادفی توأم می‌باشند اگرچه آنها بلافاصله پس از کشف توسط اسکالر ۱۹۵۹ به طور گسترده مورد توجه قرار نگرفتند. این سبک تا اواخر قرن بیستم و در قرن بیست و یکم به طور گسترده در بسیاری از زمینه‌ها به تصویب رسید. فریز^{۳۷} و والدز^{۳۸} مفهوم مدل‌سازی روابط چند متغیره با استفاده از مفصل‌ها به میدان آماری را معرفی کردند و این برنامه با استفاده از مجموعه داده‌های هزینه تلفات بیمه نشان داده شد. مدل‌های مفصلی به طور شهودی در امور مالی برای همبستگی مدل پیش‌فرض استفاده می‌شوند. دیوید لی^{۳۹} در سال ۲۰۰۰ مقاله‌ای را در زمینه مدل‌سازی همبستگی‌های پیش‌فرض اعتباری با استفاده از توابع مفصل منتشر کرد. در سال ۲۰۰۸ استفاده از مفصل‌ها برای مدل همبستگی پیش‌فرض در موسسات مالی، آژانس‌های اعتباری و در میان قانون‌گذاران دولتی وجود داشت. مقاله لی شاید بیش از هر مقاله‌ی دیگری، دقت قابل توجهی در امور مالی، از تاریخ انتشار آن تا زمان بحران مالی جهانی ۲۰۰۹-۲۰۰۸، و حتی به جریان اصلی رسانه‌ها در طی و بعد از بحران مالی داشته است. (برای مثال سالمون^{۴۰} ۲۰۰۹، جونز^{۴۱} ۲۰۰۹، چن^{۴۲} و فن^{۴۳} [۷] ۲۰۰۵). با انتشار مقاله‌ای به نام "آزمون نسبت شبه درست‌نمایی برای انتخاب مدل مفصل چند متغیره به صورت شبه پارامتری"، جزئیات یک روش انتخاب مدل مفصل که پس از آن به نام روش CF شناخته شد، را بیان نمودند.

^{۳۷}Frees

^{۳۸}Valdez

^{۳۹}David Li

^{۴۰}Salmon

^{۴۱}Jones

^{۴۲}Chen

^{۴۳}Chen

۲ مفاهیم اولیه تابع مفصل

۱.۲ تابع مفصل

یک متغیر تصادفی X با تابع توزیع تجمعی اش مشخص می شود

$$F(x) = p(X \leq x) \quad x \in \mathbb{R} \quad (۱)$$

برای مجموعه‌ای از متغیرهای تصادفی X_1, \dots, X_d توابع توزیع حاشیه‌ای F_1, \dots, F_d رفتار منحصر بفرد متغیرها را توصیف می‌کند اما هیچ اطلاعاتی راجع به رفتار توأم‌شان نمی‌دهد. مفصل‌ها توابع توزیع توأمی می‌باشند که از تعدادی توزیع‌های حاشیه‌ای یکنواخت استاندارد ساخته شده‌اند. مفصل توابع توزیع با وابستگی خاص بین متغیرهای تصادفی را بهم جفت می‌کند.

تعریف ۱.۲ (مفصل). مفصل d بعدی $[0, 1]^d \rightarrow [0, 1]$ یک تابع توزیع تجمعی توأم با توزیع‌های حاشیه‌ای یکنواخت استاندارد می‌باشد. لازم به ذکر است توزیع‌هایی که یکنواخت استاندارد نمی‌باشند را می‌توان با استفاده از تبدیل انتگرال احتمال به توزیع‌های یکنواخت استاندارد تبدیل کرد. اگر U یک متغیر تصادفی یکنواخت استاندارد و F یک تابع توزیع تجمعی دلخواه با معکوس تعمیم‌یافته‌ی F^{-1} باشد، آنگاه برای هر x متعلق به \mathbb{R}

$$p(F^{-1}(U) \leq x) = p(U \leq F(x)) = F(x) \quad (۲)$$

یعنی $F^{-1}(U)$ یک متغیر تصادفی با تابع توزیع تجمعی F می‌باشد. در موارد چندمتغیره، فرض کنید C یک مفصل U_1, \dots, U_d دارای تابع توزیع تجمعی C باشند و F_1, \dots, F_d توابع توزیع تجمعی تک متغیره دلخواه باشند. در این صورت به ازای هر x_1, \dots, x_d متعلق به \mathbb{R}

$$\begin{aligned} C(F_1(x_1), \dots, F_d(x_d)) &= p(U_1 \leq F_1(x_1), \dots, U_d \leq F_d(x_d)) \\ &= p(F_1^{-1}(U_1) \leq x_1, \dots, F_d^{-1}(U_d) \leq x_d) \end{aligned}$$

و همچنین $C(F_1(x_1), \dots, F_d(x_d))$ یک تابع توزیع تجمعی چندمتغیره با حاشیه‌های F_1, \dots, F_d تعریف می‌شود. طبق تعریف، توابع مفصل تابع توزیع تجمعی می‌باشند و با نماد C نمایش داده می‌شوند، در صورت وجود مشتقات جزئی تابع مفصل C با مشتق گرفتن نسبت به هر مولفه می‌توان تابع چگالی را به دست آورد.

قضیه ۱.۲ (اسکلار). اگر F یک تابع توزیع تجمعی d بعدی با توزیع‌های حاشیه‌ای F_1, \dots, F_d باشد آنگاه مفصل C برای هر $x_i \in \mathbb{R}$ ($i = 1, \dots, d$) به قسمی وجود داشته باشد که

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \quad (۳)$$

برقرار است. از طرفی، اگر هر F_i ($i = 1, \dots, d$) به ازای هر $x_i \in \mathbb{R}$ پیوسته باشد آنگاه مفصل C یکتاست. عکس قضیه: اگر C یک مفصل و F_1, \dots, F_d تابع توزیع تجمعی دلخواه باشند، آنگاه F طبق ۳ تابع توزیع تجمعی چند متغیره با توزیع‌های حاشیه‌ای F_1, \dots, F_d می‌باشد.

۳ روش‌های انتخاب مدل مفصل

۱. معیار اطلاعات آکائیکه (AIC) ^{۴۴}

۲. واگرایی کولبک-لیب-لر (KL-d) ^{۴۵}

۳. روش (CF) ^{۴۶}

۱.۳ معیار اطلاعات آکائیکه (AIC)

معیار اطلاعات آکائیکه به طور گسترده در انتخاب مدل‌های آماری با مبانی نظریه اطلاعات مورد استفاده قرار می‌گیرد. محاسبه‌ی مقادیر AIC با استفاده از

$$AIC = -2 \log(\text{ماکسیمم درستنمایی}) + 2k \quad (۴)$$

امکان‌پذیر می‌باشد. معمولاً در عمل مدل‌ها با استفاده از روش ماکسیمم درستنمایی یا ماکسیمم شبه‌درستنمایی برآزش داده می‌شوند بنابراین مقدار AIC به سهولت برای تمام مدل‌های مورد بررسی در دسترس می‌باشد. دلتای Δ_i ، AIC بین مقدار AIC مدل i ام و مقدار AIC مدل با کمترین مقدار AIC در مجموعه مدل‌های در نظر گرفته شده را محاسبه می‌کند.

$$\Delta_i = AIC_i - AIC_{min} \quad (۵)$$

با توجه به جدول زیر Δ_i های محاسبه شده برای انتخاب مدل استفاده می‌شوند.

^{۴۴} Akaike's Information Criterion

^{۴۵} Kullback - Leibler's divergence

^{۴۶} Chen and Fan method

جدول ۱: سطح حمایت i زمین مدل تجربی

سطح حمایت i زمین مدل تجربی	Δ_i
قابل توجه	۰ - ۲
بسیار کم	۴ - ۷
اساساً هیچ کدام	$10 <$

با توجه به آزمون فرض

داده‌ها از مفصل مورد نظر پیروی می‌کنند. : H_0

داده‌ها از مفصل مورد نظر پیروی نمی‌کنند. : H_1

و اینکه ناحیه‌ی پذیرش H_0 مبنی بر وجود Δ_i های بین ۰ - ۲ می‌باشد، آزمون مذکور انجام می‌شود.

تذکر ۱.۳. با توجه به اینکه در اینجا توابع مفصل مورد بررسی می‌باشند برای محاسبه‌ی تابع لگاریتم درستنمایی به صورت زیر عمل می‌کنیم و بقیه‌ی مراحل همانطور که بیان شد اجرا می‌شود. تابع لگاریتم درستنمایی پارامتر مفصلی θ :

$$\begin{aligned} L(\theta) &= \sum_{t=1}^n \log(f_1^i(X_{1t}) \times \cdots \times f_d^i(X_{dt}) \times c(F_1^i(X_{1t}), \dots, F_d^i(X_{dt}); \theta)) \\ &= \sum_{t=1}^n \sum_{j=1}^d \log f_j^i(X_{jt}) + \sum_{t=1}^n \log c(F_1^i(X_{1t}), \dots, F_d^i(X_{dt}); \theta) \\ &= \sum_{t=1}^n \sum_{j=1}^d \log f_j^i(X_{jt}) + \sum_{t=1}^n \log c(U_{1t}, \dots, U_{dt}; \theta) \end{aligned}$$

لازم به ذکر است که در محاسبات مقادیر AIC می‌توان جمله‌ی اول $L(\theta)$ را حذف نمود زیرا در فرمول AIC عبارت ماکسیمم درستنمایی قید شده است و در رابطه‌ی فوق برای ماکسیمم شدن تابع درستنمایی، جمله‌ی اول مستقل از θ می‌باشد پس می‌توان آن را حذف نمود و تنها محاسبه‌ی جمله‌ی دوم کفایت می‌کند.

۲.۳ واگرایی کولبک-لیبلر (KL)

واگرایی کولبک-لیبلر در حالت تک‌متغیره، بین دو توزیع داده شده با چگالی‌های احتمال پیوسته $p(x)$ (چگالی واقعی) و $q(x)$ (چگالی مورد بررسی) به صورت

$$D(p||q) = \int_x p(x) \log \frac{p(x)}{q(x)} dx = \int_x p(x) \log p(x) dx - \int_x p(x) \log q(x) dx \quad (۶)$$

می باشد.

برای هر توزیع یا مدل واقعی اولین عبارت $D(p||q)$ تنها به $\log p(x)$ بستگی دارد و برای توزیع های مورد بررسی مختلف $q(x)$ ثابت و دومین عبارت از $D(p||q)$ ، $E_p[\log q(x)]$ یا مقدار منفی امید ریاضی لگاریتم درست نمایی $q(x)$ می باشد.

تذکره ۲.۳. توزیع مورد بررسی ای که با بیشینه کردن این مقدار امید ریاضی، باعث به حداقل رساندن واگرایی کولبک-لیبلر به توزیع واقعی می شود، مطلوب است. همچنین واگرایی کولبک-لیبلر برای توابع دو متغیره پیوسته به صورت زیر می باشد

$$D(p||q) = \int \int p(x, y) \log \frac{p(x, y)}{q(x, y)} dx dy. \quad (7)$$

که در اینجا با توجه به اینکه توابع مفصل مدنظر می باشد برای محاسبه ی واگرایی کولبک-لیبلر به صورت زیر عمل می کنیم و بقیه ی مراحل همانطور که بیان شد اجرا می شود.

فاصله ی کولبک-لیبلر بین دو تابع چگالی مفصل $c^f, c^g : [0, 1]^2 \rightarrow [0, \infty)$ به صورت زیر تعریف می شود

$$KL(c^f, c^g) = \int_{u \in [0, 1]^2} \log \left(\frac{c^f(u)}{c^g(u)} \right) c^f(u) du. \quad (8)$$

این نکته لازم به ذکر است که بررسی روی ۷ مفصل (فرانک، جو، گمبل، کلیتون، گالامبوس، نرمال و تی) انجام می شود. برای تعیین بهترین مفصل با مینیمم گرفتن مقدار معیار کولبک-لیبلر مفصل مورد بررسی نسبت به دیگر مفصل ها امکان پذیر می باشد.

۳.۳ روش (CF)

روش انتخاب مدل مفصل دیگری که توسط چن و فن توصیف شده است، شامل آزمون فرضی رسمی می باشد که انتخاب مدل را از میان هر تعداد مدل مفصل مورد بررسی پیشنهاد داده شده، بیان می کند. بدین صورت که هر دو مفصل مورد بررسی به طور متوالی به عنوان یک مفصل معیار در مقابل مفصل های رقابتی دیگر آزمون می شوند.

تذکره ۳.۳. هدف روش CF تعیین الگویی مبنی بر اینکه مفصل $C_2(F_1^*(x_1), \dots, F_d^*(x_d); \theta_2)$ به طور قابل توجهی به مدل مفصل واقعی $C^*(F_1^*(x_1), \dots, F_d^*(x_d); \theta^*)$ نزدیکتر از مدل مفصل معیار $C_1(F_1^*(x_1), \dots, F_d^*(x_d); \theta_1)$ در واگرایی کولبک-لیبلر است، می باشد. که در آن C^* تابع مفصل پارامتری واقعی اما نامعلوم، θ^* پارامتر واقعی اما نامعلوم، F_j^* j امین توزیع حاشیه ای پیوسته واقعی اما نامعلوم، f_j^* نیز چگالی j امین توزیع حاشیه ای پیوسته واقعی اما نامعلوم می باشند.

نکته ۱.۳. مراحل اجرای روش CF به صورت زیر می‌باشد:

۱. بیان آزمون فرض ۲. تعیین آماره‌ی نسبت شبه درستنمایی (PLR) ۳. محاسبه‌ی برآوردگر واریانس برای ساخت آماره آزمون

۱. آزمون فرض. با فرض c_1 به عنوان مفصل معیار داریم

$$H_0 : E \left[\log \left\{ \frac{c_2(U_1, \dots, U_d; \theta_2^*)}{c_1(U_1, \dots, U_d; \theta_1^*)} \right\} \right] \leq 0,$$

$$H_1 : E \left[\log \left\{ \frac{c_2(U_1, \dots, U_d; \theta_2^*)}{c_1(U_1, \dots, U_d; \theta_1^*)} \right\} \right] > 0.$$

تذکر ۴.۳. با توجه به آزمون فرض، واضح است که تحت فرض صفر امیدریاضی لگاریتم نسبت درستنمایی واقعی از مدل‌های مفصل صفر یا منفی است.

۲. آماره‌ی نسبت شبه درستنمایی^{۴۷}. (PLR) چن و فن آزمون فرضی براساس آماره‌ی نسبت شبه درستنمایی معرفی کردند زیرا مفصل واقعی و توزیع‌های حاشیه‌ای واقعی مجهول می‌باشند. آماره‌ی PLR به صورت زیر تعریف می‌شود

$$LR_n(\tilde{F}_1, \dots, \tilde{F}_d; \hat{\theta}_2, \hat{\theta}_1) = \frac{1}{n} \sum_{t=1}^n \log \left[\frac{c_2\{\tilde{F}_1(X_{1t}), \dots, \tilde{F}_d(X_{dt}); \hat{\theta}_2\}}{c_1\{\tilde{F}_1(X_{1t}), \dots, \tilde{F}_d(X_{dt}); \hat{\theta}_1\}} \right] \quad (9)$$

توزیع‌های حاشیه‌ای \tilde{F}_j به صورت ناپارامتری برآورد شدند بنابراین نتایج برآورد نسبت درستنمایی یک نسبت درستنمایی واقعی نیست در نتیجه یک نسبت شبه درستنمایی می‌باشد. آماره‌ی PLR امیدریاضی لگاریتم خارج قسمت درستنمایی دو مفصل را برآورد می‌کند. چن و فن نشان دادند که

$$LR_n(\tilde{F}_1, \dots, \tilde{F}_d; \hat{\theta}_2, \hat{\theta}_1) \xrightarrow{a.s} E \left[\log \left\{ \frac{c_2(U_1, \dots, U_d; \theta_2^*)}{c_1(U_1, \dots, U_d; \theta_1^*)} \right\} \right] \quad (10)$$

میل می‌کند. در نتیجه آماره‌ی PLR بطور تقریبی دارای توزیع نرمال با میانگین صفر و واریانس σ^2 می‌باشد.

$$\sqrt{n} \left[LR_n(\tilde{F}_1, \dots, \tilde{F}_d; \hat{\theta}_2, \hat{\theta}_1) - E \left[\log \left\{ \frac{c_2(U_1, \dots, U_d; \theta_2^*)}{c_1(U_1, \dots, U_d; \theta_1^*)} \right\} \right] \right] \rightarrow N(0, \sigma^2) \quad (11)$$

با این حال توزیع حدی مشکلاتی را ایجاد می‌کند.

نه تنها هیچ‌راهی برای فهمیدن مقدار $E \left[\log \left\{ \frac{c_2(U_1, \dots, U_d; \theta_2^*)}{c_1(U_1, \dots, U_d; \theta_1^*)} \right\} \right]$ نیست بلکه چون θ_1^* و θ_2^* به توزیع‌های حاشیه‌ای نامعلوم اما واقعی بستگی دارد، توزیع حدی آماره‌ی PLR نامعلوم می‌باشد. به راحتی می‌توان پارامتر مزاحم را حذف کرد و این کار با اجرای پیکربندی حداقل مطلوب (LFC)^{۴۸} انجام می‌شود.

^{۴۷}Pseudo-Likelihood Ratio

^{۴۸}Least Favorable Configuration

بطور کلی LFC مشخص‌کننده پارامترهایی است که احتمال هر معیار مطلوب را به حداقل می‌رساند. تحت LFC پارامتر مزاحم

$$E \left[\log \left\{ \frac{c_2(U_1, \dots, U_d; \theta_2^*)}{c_1(U_1, \dots, U_d; \theta_1^*)} \right\} \right] = 0 \quad (12)$$

می‌شود. با فرض مقدار ثابت صفر برای امیدریاضی لگاریتم درست‌نمایی، توزیع تقریبی آماره به صورت

$$\sqrt{n}[LR_n(\tilde{F}_1, \dots, \tilde{F}_d; \hat{\theta}_2, \hat{\theta}_1)] \rightarrow N(0, \sigma^2) \quad (13)$$

ساده‌تر می‌شود. با به کار بردن LFC ، فرض صفر به

$$H_0: E \left[\log \left\{ \frac{c_2(U_1, \dots, U_d; \theta_2^*)}{c_1(U_1, \dots, U_d; \theta_1^*)} \right\} \right] = 0 \quad (14)$$

تبدیل می‌شود درحالی‌که فرض مقابل تغییر نمی‌کند.

۳. محاسبه‌ی براوردگر واریانس برای آماره آزمون. اصلاح فرض صفر یک توزیع حدی منحصر به فرد از آماره‌ی PLR را تضمین می‌کند و همه اینها نیازمند یک براوردگر واریانس آماره‌ی PLR ، σ^2 است. بدین منظور براوردگری سازگار از واریانس $\hat{\sigma}^2$ را محاسبه و از $\hat{\sigma}$ برای نرمال‌سازی آماره‌ی PLR و ساختار آماره آزمون استفاده می‌شود.

$$T_n^N = \frac{\sqrt{n}}{\hat{\sigma}} LR_n(\tilde{F}_1, \dots, \tilde{F}_d; \hat{\theta}_2, \hat{\theta}_1) \quad (15)$$

آماره آزمون T_n^N دارای توزیع حدی گوسی تحت LFC می‌باشد.

$$T^N \rightarrow N(0, 1) \quad (16)$$

فرض صفر را زمانی که $T_n^N > Z_\alpha$ باشد در سطح معنی‌داری α رد می‌کند. توجه داشته باشید که ناحیه‌ی رد فقط به چندک بالای توزیع گوسی استاندارد، Z_α بستگی دارد. در نتیجه به راحتی می‌توان p مقادیر مربوط به برازش هر مفصل را محاسبه کرد و مفصلی که بهترین برازش را به داده‌ها خواهد داشت را انتخاب می‌کنیم. بیان این نکته حائز اهمیت می‌باشد که تمامی مراحل در هر ۳ روش توسط شبیه‌سازی با نرم‌افزار R ورژن ۱,۵,۲ انجام شده است.

۴ دستاوردهای تحقیق

تذکر ۱۰۴. نتایج مربوط به برازش داده‌های شبیه‌سازی شده توسط روش CF به مفصل‌های نامبرده در جدول زیر قابل ملاحظه می‌باشد. با توجه به مقادیر جدول مشهود است که درصد برازش مفصل‌های مورد پذیرش

جدول ۲: درصد برآزش داده‌های شبیه‌سازی شده به مفصل‌ها

مفصل	فرانک	گمبل	گالامبوس	کلیتون	جو	نرمال	تی
فرانک	۱۰۰	۰	۰	۰	۰	۲	۱۰۰
گمبل	۱۰۰	۰	۱۰۰	۰	۰	۰	۱
گالامبوس	۱۰۰	۰	۱۰۰	۰	۰	۱۰۰	۱
کلیتون	۱۰۰	۰	۰	۱۰۰	۰	۰	۰
جو	۰	۱۰۰	۰	۰	۰	۰	۱
نرمال	۱۰۰	۰	۰	۰	۰	۱۰۰	۱۰۰
تی	۱۰۰	۱۰۰	۰	۰	۰	۱۴	۱

توسط روش CF درصدهای بالایی را به خود اختصاص داده‌اند. می‌توان اینطور نتیجه‌گیری نمود که روش CF ، روش مناسبی برای انتخاب مدل مفصلی می‌باشد.

تذکر ۲۰۴. نتایج مربوط به برآزش داده‌های واقعی به مفصل‌های نامبرده توسط روش CF در جدول زیر قابل ملاحظه می‌باشد. با توجه به نتایج حاصل از روش CF که در آن ۱۰۰ درصد داده‌ها به مفصل‌های گمبل و

جدول ۳: درصد برآزش داده‌های واقعی توسط روش CF به مفصل‌ها

مفصل	فرانک	گمبل	گالامبوس	کلیتون	جو	نرمال	تی
CF	۰	۱۰۰	۱۰۰	۰	۰	۰	۰

گالامبوس برآزش خورده است می‌توان نتیجه گرفت که داده‌های واقعی از مفصل گمبل و گالامبوس پیروی می‌کنند.

۵ نتیجه‌گیری

در این مقاله، در مورد روش‌های انتخاب مدل مفصل آماری بحث شده است. عملکرد روش CF نسبت به روش‌های AIC و KL بررسی شده است. در بخش آخر مقاله مطالعات شبیه‌سازی برای ارزیابی عملکرد روش CF و بحث در مورد نتایج آن صورت گرفته است. در نتیجه با توجه به نتایج به دست آمده از روش CF ، KL و AIC می‌توان اینطور برداشت نمود که روش CF نسبت به دو روش دیگر برتر می‌باشد اما هزینه‌های محاسباتی قابل توجه بیشتری برای استفاده از روش CF متجاوز از روش‌های AIC و KL وجود دارد. لازم به

ذکر است که شبیه‌سازی توسط روش‌های KL و AIC در دست انجام می‌باشد و به علت ضیق وقت تا قبل از اتمام مهلت داوری ارسال خواهد شد.

مراجع

- [1] Chen, X. and Fan, Y. "Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection." *The Canadian Journal of Statistics*, 33(3):389–414 (2005).
- [2] Killiches, M. Kraus, D. Czado, C. "Using model distances to investigate the simplifying assumption, model selection and truncation levels for vine copula." *Zentrum Mathematik, Technische Universitat Munchen*, 4(2018).
- [3] Roger, J. "statistical approaches to copula model selection." *McGill University, Montral, Quebec*, 24-37(2015).
- [4] Sklar, A. "Fonctions de repartition a n dimensions et leurs marges." *Publications de l'Institut de statistique de l'Universite de Paris*, 8:229–231 (1959).



پنجمین سمینار
نظریه مفصل و کاربردهای آن
۱۰ و ۱۱ بهمن ۱۳۹۷



ساخت مفصل دومتغیره با وابستگی دمی و کاربرد آن در مدل بندی داده‌های هزینه و درآمد

عسگری، پ^۱ میرحسینی، س.م^۲

گروه آمار، دانشکده علوم ریاضی، دانشگاه یزد

چکیده

داده‌های با وابستگی دمی بالا و پایین در علوم بیمه، اقتصاد، مالی و محیط زیست بسیار رایج است. توابع مفصل می‌توانند در به دست آوردن مدل‌های آماری با وابستگی دمی بالا و پایین مفید باشند. توابع مفصل با وابستگی دمی بالا و پایین به طور هم‌زمان بسیار نادر است. در این مقاله به ارائه یک مفصل دومتغیره با وابستگی دمی بالا و پایین پرداخته و به بررسی و مطالعه ویژگی‌های آن می‌پردازیم.

کلمات کلیدی: آمیخته مقیاسی بتا پرایم، توابع فوق هندسی، وابستگی دمی، مرتبه دمی

۱ پیش‌گفتار

مدل‌بندی وابستگی با مفصل می‌تواند در مدل‌بندی داده‌هایی که دارای وابستگی هستند، مورد استفاده قرار گیرد. در زمینه‌های خاصی نظیر علوم بیمه‌ای، مالی کمی، اقتصاد و محیط زیست مدل‌بندی آماری برای

^۱ parisa.asgari1993@gmail.com

^۲ mmirhoseini@yazd.ac.ir

دم‌های توأم اغلب نقش مهمی ایفا می‌کند. یک مثال معمولی برای اهمیت مدل‌بندی وابستگی دمی، استفاده از مفصل‌های گوسی برای مدل‌بندی وابستگی قراردادهای مالی است [۵].

مقالات زیادی نیز برای مطالعه الگوهای وابستگی دمی وجود دارد. به عنوان مثال، [۷] وابستگی دمی توزیع‌های بیضوی را بررسی می‌کند. [۱] بحث جامعی در رفتار دمی مفصل‌های ارشمیدسی دارد و [۴] به بررسی الگوهای دمی برای مدل‌های آمیخته مقیاسی می‌پردازد.

در این مقاله، یک مدل مفصل دو متغیره دو پارامتری جدید ارائه می‌دهیم که دارای وابستگی کامل دمی در هر دو دم بالا و پایین است و وابسته به مقدار پارامترها، دم بالا و پایین می‌تواند متقارن انعکاسی یا نامتقارن انعکاسی باشد.

علاوه بر این اکثر محاسبات بر اساس تابع فوق هندسی گاوس F_2 و تابع فوق هندسی اپل F_1 است [۶].

۲ ساخت توزیع دو متغیره در حالت کلی

فرض کنید $H_1 \sim F_1(h_1)$ ، $H_2 \sim F_2(h_2)$ ، $R \sim G(r)$ ، $H_i \stackrel{ind}{\sim} R$ و همگی متغیرهای تصادفی نامنفی باشند. همچنین فرض کنید $X_1 = RH_1$ و $X_2 = RH_2$ ، تابع توزیع توأم (X_1, X_2) به صورت زیر خواهد بود:

$$\begin{aligned} H(x_1, x_2) &= P(X_1 \leq x_1, X_2 \leq x_2) \\ &= P(RH_1 \leq x_1, RH_2 \leq x_2) \\ &= \int_0^{\infty} P(RH_1 \leq x_1, RH_2 \leq x_2 \mid R = r) dG(r) \\ &= \int_0^{\infty} P(H_1 \leq x_1/r, H_2 \leq x_2/r) dG(r) \\ &= \int_0^{\infty} P(H_1 \leq x_1/r) P(H_2 \leq x_2/r) dG(r) \\ &= E_G[F_1(x_1/R) F_2(x_2/R)]. \end{aligned} \quad (1)$$

و توابع توزیع حاشیه‌ای X_1 و X_2 برابرند با:

$$H_1(x_1) = E_G[F_1(x_1/R)]$$

$$H_2(x_2) = E_G[F_2(x_2/R)].$$

در ادامه برخی از ویژگی‌های وابستگی این توزیع بیان می‌شود.

قضیه ۱.۰۲. اگر $R_1 \stackrel{s.t.}{\leq} R_2$ ، یعنی $G_1(r) > G_2(r)$ ، آن‌گاه

$$H_{G_1}(x_1, x_2) \geq H_{G_2}(x_1, x_2).$$

برهان. از آنجا که

$$H(x_1, x_2) = E_G[F_1(x_1/R)F_2(x_2/R)],$$

با توجه به فرض قضیه داریم

$$G_1(r) > G_2(r),$$

یا به طور معادل

$$\bar{G}_1(r) \leq \bar{G}_2(r).$$

بنابراین

$$\int_{\cdot}^{\infty} F_1(x_1/r)F_2(x_2/r)dG_1(r) > \int_{\cdot}^{\infty} F_1(x_1/r)F_2(x_2/r)dG_2(r),$$

در نتیجه

$$H_{G_1}(x_1, x_2) \geq H_{G_2}(x_1, x_2).$$

□

لم ۱.۰۲. فرض کنید X یک متغیر تصادفی و ψ_1 و ψ_2 توابعی از X باشند

(الف) اگر ψ_1 و ψ_2 هر دو صعودی یا هر دو نزولی باشند، آن‌گاه $Cov(\psi_1(X), \psi_2(X)) \geq 0$.

(ب) اگر ψ_1 صعودی (نزولی) و ψ_2 نزولی (صعودی) باشد، آن‌گاه $Cov(\psi_1(X), \psi_2(X)) \leq 0$.

□

برهان. مرجع [۳] لم ۲، ۱ صفحه ۲۳ را ببینید.

قضیه ۲.۰۲. اگر (X_1, X_2) از تابع توزیع توأم $H(x_1, x_2)$ پیروی کند، آن‌گاه (X_1, X_2) PQD است.

برهان. از آنجا که

$$H(x_1, x_2) = E_G[F_1(x_1/R)F_2(x_2/R)],$$

$$H_1(x_1) = E_G[F_1(x_1/R)] = E_G(\psi_1(R)),$$

$$H_2(x_2) = E_G[F_2(x_2/R)] = E_G(\psi_2(R)),$$

که در آن $\psi_1(R) = F_i(x_i/R), i = 1, 2$ و توابع $\psi_1(\cdot)$ و $\psi_2(\cdot)$ هر دو نسبت به r نزولی هستند، داریم

$$Cov(\psi_1(R), \psi_2(R)) \geq 0,$$

در نتیجه

$$Cov(F_1(x_1/R), F_2(x_2/R)) \geq 0.$$

یا به طور معادل

$$E_G[F_1(x_1/R)F_2(x_2/R)] \geq E_G[F_1(x_1/R)]E_G[F_2(x_2/R)],$$

در نتیجه

$$H_G(x_1, x_2) \geq H_1(x_1)H_2(x_2).$$

□

تذکره ۱۰۲. برای بردار (X_1, X_2) با تابع توزیع (۱)، مفصل متناظر به صورت زیر است:

$$C_H(u, v) = H(H_1^{-1}(u), H_2^{-1}(v)).$$

۳ حالت خاص

برای حالت خاص، فرض کنید بردار تصادفی دومتغیره (X_1, X_2) در مفروضات زیر صدق کند:

(۱) $X_i = RH_i, i = 1, 2$ که متغیرهای R و H_i ها از هم مستقلند.

(۲) R دارای توزیع $BetaPrime(\alpha, \beta)$ با تابع چگالی زیر باشد

$$f_R(r) = [Beta(\alpha, \beta)]^{-1} r^{\alpha-1} (1+r)^{-\alpha-\beta} \quad r, \alpha, \beta > 0,$$

که $\alpha, \beta > 0$ پارامترهای شکل توزیع و $Beta(\alpha, \beta)$ تابع بتا است.

(۳) $f_{H_i}(h) = (1+h)^{-2}, h \geq 0$ ؛ یعنی؛ $H_i \sim Pareto(1), i = 1, 2$

کلاس توزیع هر بردار تصادفی که در مفروضات (۱) تا (۳) صدق کند را با \mathcal{M} نشان می‌دهیم و گفته می‌شود

$$(X_1, X_2) \in \mathcal{M}(\alpha, \beta)$$

با یک بردار تصادفی (X_1, X_2) که در فرض (۱) تا (۳) صدق کند یک مفصل دومتغیره به صورت $C(u, v) =$

ساخته می‌شود. در ادامه کمیت‌های اساسی مورد نیاز در شبیه‌سازی، محاسبه و با

مفصل پیشنهادی مدل برازش داده می‌شود.

قضیه ۱.۳. تابع چگالی و تابع توزیع متغیر تصادفی $X_i, i = 1, 2$ که در فرض (۱) تا (۳) صدق کند، به صورت زیر هستند:

$$F_{X_i}(x) = 1 - \frac{\alpha}{\alpha + \beta} {}_2F_1(1, \beta, \alpha + \beta + 1, 1 - x), \quad x > 0. \quad (2)$$

$$f_{X_i}(x) = 1 - \frac{\alpha\beta}{(\alpha + \beta)(\alpha + \beta + 1)} {}_2F_1(2, \beta + 1, \alpha + \beta + 2, 1 - x), \quad x > 0. \quad (3)$$

که ${}_2F_1$ تابع فوق هندسی گاوس است.

قضیه ۲.۳. تابع توزیع توأم و تابع چگالی توأم بردار (X_1, X_2) که در فرض (۱) تا (۳) صدق کند، برای $x_1, x_2 > 0$ به صورت زیر است:

(۴)

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1) + F_{X_2}(x_2) - 1 + \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} F_1(\beta, 1, 1, \alpha + \beta + 2, 1 - x_1, 1 - x_2),$$

$$f_{X_1, X_2}(x_1, x_2) = \frac{\alpha\beta(\alpha + 1)(\beta + 1)}{(\alpha + \beta) \cdots (\alpha + \beta + 3)} F_1(\beta + 2, 2, 2, \alpha + \beta + 4, 1 - x_1, 1 - x_2), \quad (5)$$

که F_1 تابع فوق هندسی اپل است.

تذکر ۱.۳. تابع بقا توأم بردار (X_1, X_2) به صورت زیر است:

$$\bar{F}_{X_1, X_2}(x_1, x_2) = \frac{\alpha(\alpha + 1)}{(\alpha + \beta)(\alpha + \beta + 1)} F_1(\beta, 1, 1, \alpha + \beta + 2, 1 - x_1, 1 - x_2). \quad (6)$$

تذکر ۲.۳. با توجه به فرمول‌های قبل برای (X_1, X_2) ، تابع توزیع و تابع چگالی مفصل تولید شده به فرم زیر است. شایان ذکر است که دو متغیر X_1 و X_2 هم توزیع هستند

$$C(u, v) = P[X_1 \leq F_{X_1}^{-1}(u), X_2 \leq F_{X_2}^{-1}(v)], \quad (7)$$

$$c(u, v) = f_{X_1, X_2}(F_{X_1}^{-1}(u), F_{X_2}^{-1}(v)) / [f_{X_1}(F_{X_1}^{-1}(u)) f_{X_2}(F_{X_2}^{-1}(v))]. \quad (8)$$

در این بخش، مرتبه دمی، پارامترهای وابستگی دمی و معیارهای وابستگی کلی مانند ρ اسپیرمن و τ کندال را برای مدل مفصل پیشنهادی به دست می‌آوریم.

قضیه ۳.۳. مفصل تولیدی توسط بردار $(X_1, X_2) \in \mathcal{M}(\alpha, \beta)$ دارای مرتبه دمی بالا $k_U = \max\{\min(2, \beta), 1\}$ و مرتبه دمی پایین $k_L = \max\{\min(2, \alpha), 1\}$ است. اگر $0 < \alpha \leq 1$ ، آنگاه پارامتر وابستگی دمی پایین $\lambda_L = 1 - \alpha$ است. اگر $0 < \beta \leq 1$ ، آنگاه پارامتر وابستگی دمی بالا $\lambda_U = 1 - \beta$ است. اگر $\alpha = \beta$ باشد، مفصل دارای ویژگی تقارن شعاعی است.

تذکر ۳.۳. با توجه به قضیه (۳.۳)، وابستگی دمی بالا و پایین توسط پارامترهای α و β کنترل می‌شود.

قضیه ۴.۳. ρ -اسپیرمن بردار $(X_1, X_2) \in \mathcal{M}(\alpha, \beta)$ برابر است با

$$\rho = 12 \int_0^1 \int_0^1 W(x_1, x_2) dx_1 dx_2 - 3 \quad (9)$$

که در آن

$$W(x_1, x_2) = F_1(\beta, 1, 1, \alpha + \beta + 2, 1 - x_1, 1 - x_2) {}_2F_1(2, \beta + 1, \alpha + \beta + 2, 1 - x_1) \quad (10)$$

$$\times {}_2F_1(2, \beta + 1, \alpha + \beta + 2, 1 - x_2). \quad (11)$$

قضیه ۵.۳. τ - کندال بردار $(X_1, X_2) \in \mathcal{M}(\alpha, \beta)$ برابر است با

$$\tau = \frac{4}{[\text{Beta}(\alpha, \beta)]^2} \int_0^1 \int_0^1 \left(\frac{\xi \ln[\xi] + 1 - \xi}{1 - 2\xi + \xi^2} \right)^2 [(1-x)(1-y)]^{\alpha-1} [xy]^{\beta-1} dx dy - 1 \quad (12)$$

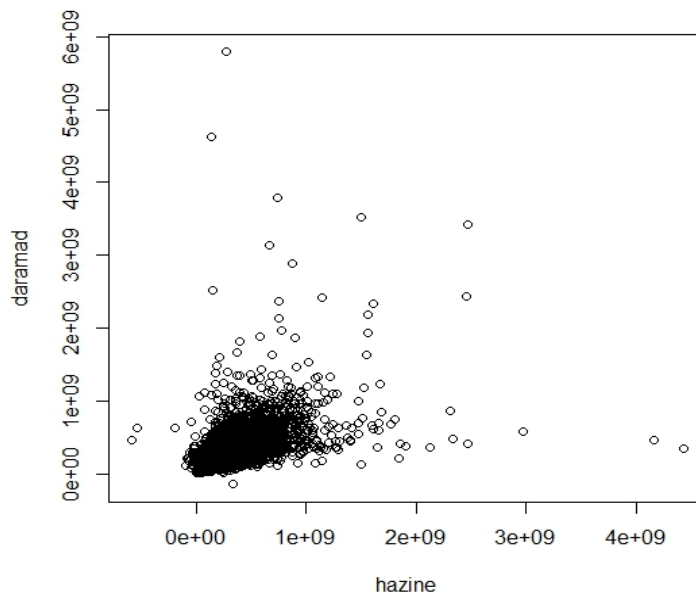
که در آن

$$\xi := [y(1-x)]/[x(1-y)]. \quad (13)$$

۴ تحلیل داده واقعی

۱.۴ داده‌های هزینه و درآمد خانوار

برای بررسی کارایی مدل، داده‌های هزینه و درآمد سال ۱۳۹۴ تعداد ۱۹۳۸۱ خانوار ایرانی که از مرکز آمار ایران اتخاذ شده، در نظر گرفته شده است. نمودار پراکنش داده‌ها و رتبه‌های یکنواخت شده در شکل ۱ و ۲ نمایش داده شده است. همانطور که دیده می‌شود داده‌ها دارای وابستگی دمی بالا و پایین هستند.

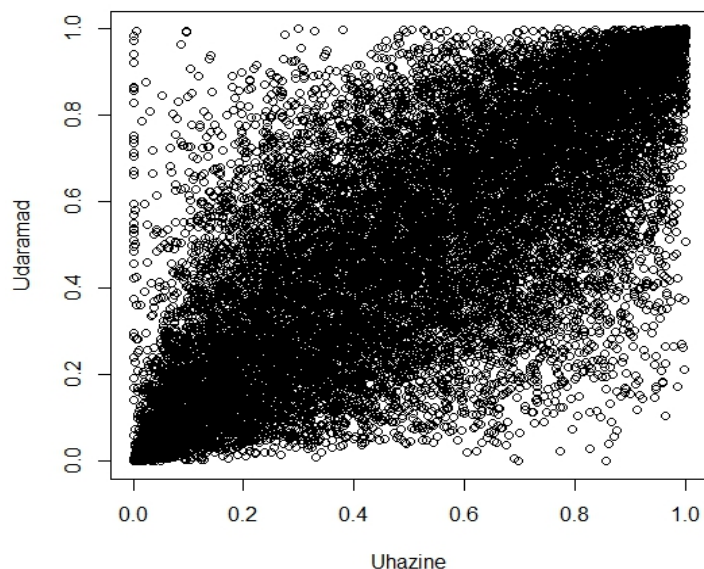


شکل ۱: داده‌های هزینه و درآمد خانوار

نتایج مربوط به برازش مفصل‌های مختلف در جدول ۱ آمده است. همانطور که ملاحظه می‌شود بر اساس معیار AIC مدل GGEE مناسب‌تر است.

جدول ۱: نتایج برازش مفصل‌های مختلف به داده‌های هزینه و درآمد خانوار

مفصل	برآورد ماکسیمم درست‌نمایی	انحراف معیار	لگاریتم درست‌نمایی	AIC
جو	$\hat{\alpha} = ۲٫۶۷$	$se(\hat{\alpha}) = ۰٫۰۲$	۷۵۱۲٫۰۷	-۱۵۰۲۲٫۱۴
کلایتون بازتاب شده	$\hat{\alpha} = ۱٫۸۶$	$se(\hat{\alpha}) = ۰٫۰۲$	۷۵۸۶٫۹	-۱۵۱۷۱٫۸
هاسلر-ریس	$\hat{\alpha} = ۱٫۶۲$	$se(\hat{\alpha}) = ۰٫۰۱$	۶۲۵۳	-۱۲۵۰۴
BB۶	$\hat{\alpha} = ۱, \hat{\beta} = ۲٫۱۷$	$se(\hat{\alpha}) = ۰٫۰۴, se(\hat{\beta}) = ۰٫۰۵$	۸۴۵۴٫۴۸	-۱۶۹۰۴٫۹۶
گمبل	$\hat{\alpha} = ۲٫۱۷$	$se(\hat{\alpha}) = ۰٫۰۱$	۸۴۵۴٫۸	-۱۶۹۰۷٫۶
t-فرین مقدار	$\hat{\alpha} = ۰٫۸۴, \hat{\beta} = ۲٫۹۲$	$se(\hat{\alpha}) = ۰٫۰۱, se(\hat{\beta}) = ۰٫۱۳$	۶۹۷۹	-۱۳۹۴۸
گالامبوس	$\hat{\alpha} = ۱٫۲۷$	$se(\hat{\alpha}) = ۰٫۰۱$	۶۷۶۲	-۱۳۵۲۲
GGEE	$\hat{\alpha} = ۰٫۱, \hat{\beta} = ۰٫۶۲$	$se(\hat{\alpha}) = ۰٫۰۶, se(\hat{\beta}) = ۰٫۰۳$	۱۱۹۷۱٫۸۳	-۲۳۹۳۹٫۶۶



شکل ۲: رتبه‌های یکنواخت شده هزینه و درآمد خانوار

مراجع

- [1] A. Charpentier and J. Segers, *Tails of multivariate Archimedean copulas*, Journal of Multivariate Analysis, 100(7):1521–1537, 2009.
- [2] L. Hua, *On a bivariate copula with both upper and lower full-range tail dependence*, Insurance: Mathematics and Economics, 73,94-104, 2017.
- [3] H. Joe, *Dependence Modeling with Copulas*, Chapman and Hall, 2014.
- [4] H. Li and Y. Sun, *Tail dependence for heavy-tailed scale mixtures of multivariate distributions*, Journal of Applied Probability, 46(4):925–937, 2009.
- [5] D. Li, *a copula function approach*, Journal of Fixed Income, 9(4):43–54, 2000.
- [6] M. J. Schlosser, *Multiple hypergeometric series: Appell series and beyond*, In Computer Algebra in Quantum Field Theory, pages 305–324. Springer, 2013.
- [7] R. Schmidt. *Tail dependence for elliptically contoured distributions*, Mathematical Methods of Operations Research, 55(2):301–327, 2002.



پنجمین سمینار
نظریه مفصل و کاربردهای آن
۱۰ و ۱۱ بهمن ۱۳۹۷



مفصل‌هایی برای ساخت حرکت براونی دو بعدی

بیک‌مرادی، ف^۱ دولتی، ع^۲ زمانی، م.ص^۲

گروه آمار، دانشکده علوم ریاضی، دانشگاه یزد

چکیده

یکی از کاربردهای حرکت براونی، به عنوان نوسانات تصادفی متغیرهای اقتصادی و مالی است که در قالب معادلات دیفرانسیل تصادفی مدل‌بندی می‌شوند. نوسانات متغیرهای یک دستگاه از معادلات دیفرانسیل تصادفی نیز در قالب یک بردار چند بعدی از حرکت‌های براونی بیان می‌شود. در مطالعه حرکت براونی چند بعدی فرض می‌شود که مؤلفه‌ها از هم مستقل هستند. این فرض در عمل قابل استفاده نیست و لازم است وابستگی مؤلفه‌ها در نظر گرفته شود. در این مقاله به ساخت حرکت براونی دو بعدی با استفاده از توابع مفصل پرداخته می‌شود.

کلمات کلیدی: حرکت براونی، مفصل، وابستگی، گاوسی.

^۱ beykmorady@gmail.com

^۲ adolati@yazd.ac.ir

^۲ zamani@yazd.ac.ir

۱ مقدمه

فرایند تصادفی $\{B_t : t \geq 0\}$ یک حرکت براونی استاندارد است اگر

(الف) با احتمال یک $B_0 = 0$ ،

(ب) برای هر $s < t$ $B_t - B_s \sim N(0, t - s)$ ،

(ج) برای هر $0 \leq t_1 \leq t_2 \leq \dots \leq t_n$ ، $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ از یکدیگر مستقل باشند،

(د) با احتمال یک تمام مسیرهای B_t پیوسته باشند. [۳].

برای $t > 0$ فرض کنید B_t^1 و B_t^2 دو حرکت براونی استاندارد روی یک فضای احتمال (Ω, \mathcal{F}, P) باشند. زوج $\mathbf{B}_t = (B_t^1, B_t^2)$ یک بردار تصادفی با مقادیری در \mathbb{R}^2 است که به آن حرکت براونی دو بعدی گفته می‌شود [۴].

به سادگی می‌توان دید که حرکت براونی دو بعدی $\{\mathbf{B}_t := (B_t^1, B_t^2) : t \geq 0\}$ مشابه حرکت براونی یک بعدی ویژگی مارکوفی را دارد و یک مارتینگل نیز هست. در حرکت براونی دو بعدی مؤلفه‌ها مستقل فرض می‌شوند که در عمل فرض مناسبی برای مدل‌سازی نیست. در بیشتر مسائل کاربردی با حالت‌هایی مواجه می‌شویم که با دو حرکت براونی توصیف می‌شوند که مستقل نیستند.

یکی از مسائل مهم در بازار انرژی، قیمت‌گذاری و کاهش ریسک معاملات است. فرض کنید X_t قیمت برق در زمان $t > 0$ ، Y_t قیمت سوخت مصرفی نیروگاه در زمان $t > 0$ و H هزینه تولید ۱ واحد برق با مصرف ۱ واحد سوخت باشد که به آن ضریب گرمایی گفته می‌شود. درآمد نیروگاه را با $\max\{X_t - HY_t - K, 0\}$ می‌توان مدل‌بندی کرد که در آن K نشان‌دهنده هزینه ثابت است. برای مطالعه رفتار احتمالی درآمد این نیروگاه لازم است که توزیع توأم X_t و Y_t مدل‌بندی شود. از آنجا که X_t و Y_t مستقل نیستند، بایستی مدلی برای وابستگی بین آنها در نظر گرفته شود. برای اینکه این نیروگاه سودده باشد بایستی $P(X_t - HY_t \geq x)$ برای هر $x \geq 0$ ، حداقل $\frac{1}{4}$ باشد. فرض کنید X_t و Y_t با دو حرکت براونی B_t^1 و B_t^2 به صورت $X_t = \sigma_1 B_t^1$ و $Y_t = \sigma_2 B_t^2$ مدل‌بندی شده باشد و ضریب همبستگی B_t^1 و B_t^2 برابر مقدار ثابت ρ باشد. اگر وابستگی را گاوسی (نرمال) دو متغیره فرض کنیم یعنی $(B_t^1, B_t^2) \sim N_2(0, 0, t, t, \rho)$ ، در این صورت با توجه به تقارن آن، برای هر $x \in \mathbb{R}$ داریم،

$$P(X_t - HY_t \geq x) = P(X_t - HY_t \leq -x).$$

در نتیجه برای $x \geq 0$ خواهیم داشت،

$$P(X_t - HY_t \geq x) \leq \frac{1}{4}.$$

بنابراین با در نظر گرفتن وابستگی گاوسی دو متغیره نمی‌توان سود این نیروگاه را مدل‌سازی نمود، هر چند توزیع حاشیه‌ای هر دو متغیر X_t و Y_t با توزیع گاوسی قابل بیان است. یک راه برای حل این مشکل استفاده از ساختار وابستگی دیگری به جز گاوسی دو متغیره است که اجازه دهد کمیت سود $X_t - HY_t$ نامتقارن باشد. توابع مفصل ابزار مفیدی در مدل‌سازی وابستگی بین متغیرهای تصادفی هستند. در این مقاله با استفاده از مفصل به عنوان ساختار وابستگی به مطالعه حرکت براونی دو بعدی پرداخته می‌شود. البته هر مفصلی نیز برای مدل‌سازی دو حرکت براونی قابل استفاده نیست.

۲ ساختار وابستگی گاوسی

فرایند تصادفی حقیقی مقدار $\{X_t : t \geq 0\}$ گاوسی (نرمال) نامیده می‌شود اگر به ازای هر دنباله $0 \leq t_1 < t_2 < \dots < t_n$ بردار تصادفی $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ دارای توزیع نرمال چند متغیره باشد.

در ادامه ویژگی گاوسی بودن بردار (B_t^1, B_t^2) ، $t \geq 0$ را بررسی می‌کنیم. فرض کنید $F_{j,t}(x)$ ، $j = 1, 2$ تابع توزیع B_t^i ، $i = 1, 2$ باشد. بنا به قضیه اسکالر تابع توزیع توأم B_t^1 و B_t^2 را می‌توان به صورت

$$H_t(x, y) = C_t(F_{1,t}(x), F_{2,t}(y)),$$

با تابع چگالی توأم

$$h_t(x, y) = f_{1,t}(x) f_{2,t}(y) c_t(F_{1,t}(x), F_{2,t}(y)), \quad (1)$$

نوشت، که در آن $c_t(u, v) = \frac{\partial^2}{\partial u \partial v} C_t(u, v)$ تابع چگالی مفصل است. با جایگذاری توابع توزیع $F_{i,t}$ و $f_{i,t}$ ، $i = 1, 2$ در (۱) خواهیم داشت

$$h_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2 + y^2}{2t}} c_t\left(\Phi\left(\frac{x}{\sqrt{t}}\right), \Phi\left(\frac{y}{\sqrt{t}}\right)\right). \quad (2)$$

قضیه ۱.۰۲. برای $t > 0$ ، بردار (B_t^1, B_t^2) دارای توزیع نرمال دو متغیره توأم است اگر و فقط اگر $c_t(u, v) = 1$ یا به طور معادل $C_t(u, v) = uv$ باشد یعنی B_t^1 و B_t^2 از هم مستقل باشند.

در بخش بعدی به حرکت براونی دو بعدی با ساختار وابستگی غیر گاوسی می‌پردازیم.

۳ ساختار وابستگی غیر گاوسی

مثال ۱.۳. فرض کنید مفصل C_t برای $t \geq 0$ برابر با M باشد به طوری که $M(u, v) = \min\{u, v\}$ ، بنابراین، $u, v \in [0, 1]$

$$\begin{aligned} H_t(x, y) &= \frac{1}{\sqrt{2\pi t}} \min\left\{ \int_{-\infty}^x \exp(-v^2/(2t)) du, \int_{-\infty}^y \exp(-u^2/(2t)) dv \right\} \quad (3) \\ &= \Phi\left(\frac{\min\{x, y\}}{\sqrt{t}}\right). \end{aligned}$$

با انجام محاسبات ساده خواهیم داشت،

$$\frac{\partial^2 H_t(x, y)}{\partial x \partial y} = 0 \quad a.s. \quad (4)$$

بنابراین H_t پیوسته مطلق نیست و در نتیجه نمی‌تواند یک توزیع گاوسی دو متغیره باشد هرچند حاشیه‌ای آن حرکت براونی است.

مثال ۲.۳. فرض کنید مفصل C_t برای $t \geq 0$ به صورت زیر باشد

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

آنگاه یک تابع توزیع با حاشیه‌ای‌های حرکت براونی به شکل زیر خواهد بود

$$H_t(x, y) = \max\left\{ \Phi\left(\frac{x}{\sqrt{t}}\right) + \Phi\left(\frac{y}{\sqrt{t}}\right) - 1, 0 \right\},$$

برای این توزیع دو متغیره نیز $a.s.$ $\frac{\partial^2 H_t(x, y)}{\partial x \partial y} = 0$ است و نمی‌تواند گاوسی باشد.

۴ مفصل قابل پذیرش برای حرکت براونی دو بعدی

فرض کنید مجموعه‌ی مفصل‌ها با \mathbb{C} و مجموعه‌ی مفصل گاوسی تکمیل شده توسط کران‌های بالا و پایین فرشه-هافدینگ $M(u, v) = \min\{u, v\}$ و $W(u, v) = \max\{u + v - 1, 0\}$ متناظر با حد $\rho = 1$ و $\rho = -1$ با \mathbb{C}_G نشان داده شود. $\mathbb{C}_G = \{C \in \mathbb{C} : \exists \rho \in (-1, 1), C = C_{G, \rho}\} \cup \{M, W\}$ که $C_{G, \rho}$ مفصل گاوسی با پارامتر ρ است. بنابراین،

$$C_{G, \rho}(u, v) = \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v)),$$

که Φ توزیع نرمال استاندارد و Φ_ρ توزیع نرمال دو متغیره با ضریب همبستگی ρ است

$$\Phi_\rho(x, y) = \int_{-\infty}^y \int_{-\infty}^x \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)} du dv.$$

گوئیم گردایه $\{C_t : t \geq 0\}$ از مفصل‌های C_t برای n فرایند مارکوفی حقیقی مقدار $\{X_t^{(i)} : t \geq 0, 1 \leq i \leq n\}$ تعریف شده روی فضای احتمال $(\Omega, \mathcal{F}, \mathbb{P})$ ، پذیرفتنی است اگر فرایند مارکوفی m بعدی $\{Z_t^{(i)} : t \geq 0, 1 \leq i \leq m\}$ روی فضای احتمال $(\Omega, \mathcal{F}, \mathbb{P})$ موجود باشد به طوری که

$$\begin{cases} \ell(Z_t^i) = \ell(X_t^i), 1 \leq i \leq n, \\ Z_t^i = X_t^i, 1 \leq i \leq n, \end{cases}$$

و مفصل آن برابر C_t باشد [۱].

مجموعه‌ی مفصل‌های پذیرفتنی برای حرکت‌های براونی با \mathbb{C}_B نشان داده می‌شود. می‌توان \mathbb{C}_G را به چارچوب پویا تعمیم داد با تعریف

$$\mathbb{C}_G^d = \{(C_t)_{t \geq 0} : \forall t \in \mathbb{R}^+, C_t \in \mathbb{C}_G\} \cap \mathbb{C}_B,$$

اشتراک گرفتن با \mathbb{C}_B لازم است، زیرا شرایط مورد نیاز روی C_t برای مفصل پذیرفتنی بودن، مشخص نیست. این اشتراک تهی نیست، زیرا $\{(C_t)_{t \geq 0} : \exists \rho \in (-1, 1), \forall t \in \mathbb{R}^+ C_t = C_{G,\rho}\} \subset \mathbb{C}_B$. مجموعه‌ی مفصل مجاز حرکت‌های براونی به مفصل گاوسی محدود نمی‌شود.

در ادامه دو خانواده از مفصل‌های پذیرفتنی برای حرکت براونی دو بعدی ارائه می‌شود. چون انعکاس حرکت براونی، مجدداً یک حرکت براونی است، مفصل بین حرکت براونی و انعکاس یافته آن کاندیدای خوبی برای متعلق بودن به \mathbb{C}_B است.

فضای احتمال فیلتر شده $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P})$ با فیلتر $(\mathcal{F}_t)_{t \geq 0}$ را در نظر بگیرید و همچنین فرض کنید $B = \{B_t : t \geq 0\}$ حرکت براونی نسبت به فیلتر $(\mathcal{F}_t)_{t \geq 0}$ باشد. انعکاس حرکت براونی B روی $x = h$ با \tilde{B}^h نشان داده می‌شود که $\tilde{B}_t^h = -B_t + 2(B_t - B_{\tau^h}) \mathbb{1}_{t \geq \tau^h}$ ، $\tau^h = \inf \{t \geq 0 : B_t = h\}$.

قضیه ۱.۴. فرض کنید $h > 0$ باشد. مفصل (B, \tilde{B}^h) را که با $(C_t^{ref,h})_{t \geq 0}$ نشان داده و مفصل براونی انعکاس یافته نامیده می‌شود به صورت زیر به دست می‌آید

$$C_t^{ref,h}(u, v) = \begin{cases} v & \Phi^{-1}(u) - \Phi^{-1}(v) \geq \frac{\gamma h}{\sqrt{t}}, \\ W(u, v) + \Phi(\Phi^{-1}(M(u, 1-v)) - \frac{\gamma h}{\sqrt{t}}) & \Phi^{-1}(u) - \Phi^{-1}(v) < \frac{\gamma h}{\sqrt{t}}, \end{cases} \quad (5)$$

که $(C_t^{ref,h})_{t \geq 0} \in \mathbb{C}_B$.

تعمیمی از مفصل براونی انعکاس یافته در قضیه قبل به صورت زیر است.

قضیه ۲.۴. فرض کنید $h > 0$ و $\rho \in (0, 1)$. تابع

$$C_t(u, v) = \begin{cases} \Phi_\rho(\Phi^{-1}(u), \Phi^{-1}(v) + \frac{\sqrt{\rho h}}{\sqrt{t}}) + v - \Phi(\Phi^{-1}(v) + \frac{\sqrt{\rho h}}{\sqrt{t}}) & u \geq \Phi(\frac{h}{\sqrt{t}}) \\ \Phi_{-\rho}(\Phi^{-1}(u), \Phi^{-1}(v)) + \Phi_\rho(\Phi^{-1}(u) - \frac{\sqrt{h}}{\sqrt{t}}, \Phi^{-1}(1-v) - \frac{\sqrt{\rho h}}{\sqrt{t}}) + \\ \Phi_\rho(\Phi^{-1}(u) - \frac{\sqrt{h}}{\sqrt{t}}, \Phi^{-1}(v)) - \Phi(\Phi^{-1}(u) - \frac{\sqrt{h}}{\sqrt{t}}) & u < \Phi(\frac{h}{\sqrt{t}}) \end{cases}$$

یک مفصل متعلق به \mathbb{C}_B است.

این مفصل نامتقارن است به این معنا که برای مقادیری از $u, v \in [0, 1]$ و $t > 0$ داریم: $C_t(u, v) \neq C_t(v, u)$. این مفصل شامل دو حالت از همبستگی است: وابستگی مثبت $\rho > 0$ و وابستگی منفی $\rho < 0$. حالتی که $\rho = 0$ است، مفصل تعریف شده در قضیه قبل تبدیل به مفصل استقلال می شود که متعلق به \mathbb{C}_B است.

قضیه زیر مفصل دیگری برای (B, \tilde{B}^ξ) ارائه می دهد که متعلق به \mathbb{C}_B است.

قضیه ۳.۴. فرض کنید ξ متغیر تصادفی مثبت با چگالی f^ξ و تابع بقا \bar{F}^ξ باشد. در این صورت تابع

$$C_t^\xi(u, v) = v - \int_{-\infty}^{\Phi^{-1}(M(1-u, v))} \frac{e^{-\frac{w^2}{t}}}{\sqrt{2\pi}} \bar{F}^\xi(\frac{\sqrt{t}}{2}(\Phi^{-1}(M(u, 1-v)) - w)) dw$$

یک مفصل متعلق به \mathbb{C}_B است.

مثال ۱.۴. فرض کنید X متغیر تصادفی نمایی با پارامتر λ و $Y \stackrel{d}{=} h + X$ باشد که $h \in \mathbb{R}$. در این صورت

$$\bar{F}^\xi(x) = \begin{cases} 1 & x \leq h \\ e^{-\lambda(x-h)} & x > h. \end{cases}$$

با قرار دادن این تابع در قضیه ۳,۴ داریم،

$$C_t^{h, \lambda}(u, v) = W(u, v) + \min [\Phi(\Phi^{-1}(M(1-u, v)) - \frac{\sqrt{h}}{\sqrt{t}}), M(u, 1-v)] - \Phi(\min[\Phi^{-1}(M(1-u, v)) - \frac{\sqrt{h}}{\sqrt{t}}, \Phi^{-1}(M(u, 1-v))] - \frac{\lambda\sqrt{t}}{2}) e^{\lambda h + \frac{\lambda^2 t}{4} + \frac{\lambda\sqrt{t}}{2} \Phi^{-1}(M(u, 1-v))}$$

که یک مفصل متعلق به \mathbb{C}_B است.

مفصل های بالا می توانند به طور همزمان برای ساخت حرکت براونی دو بعدی تعمیم یافته مورد استفاده قرار گیرند.

مراجع

- [1] Jaworski, P. and M. Krzywda (2013), Coupling of Wiener processes by using copulas, *Statist. Probab. Letter*, 83(9), 2027-2033.
- [2] Nelsen, R. B. (2006), *An Introduction to Copulas*, Second edition. Springer, New York.
- [3] Oksendal, B. (2003), Stochastic differential equations. *Stochastic differential equations*. Springer, Berlin, Heidelberg, PP. 56-84.
- [4] Sempi, C. (2010), Coupled brownian motion. *Combining Soft Computing and Statistical Methods in Data Analysis*, (PP. 569-574). Springer, Berlin, Heidelberg.
- [5] Shreve, S. E., and Karatzas, I. (1991), Brownian motion and stochastic calculus. *Newyork Berlin. Heidelberg. London Paris Tokyo*.



تحلیل دو متغیره خشکسالی با استفاده از شاخص SEPI بر مبنای توابع مفصل

وحدتی، ر^۱ داوری، ک^۲ قهرمان، ب^۳ رضایی پزند، ح^۴

^{۱،۲،۳} گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد

^۴ گروه مهندسی عمران، دانشکده مهندسی، دانشگاه آزاد اسلامی واحد مشهد

چکیده

تحلیل آماری خشکسالی یکی از نیازهای مهم هوا و اقلیم شناسی، هیدرولوژی و سیستم‌های برنامه‌ریزی و مدیریت منابع آب می‌باشد. از مشخصه‌های مهم خشکسالی می‌توان شدت و مدت را نام برد که شناخت و تجزیه و تحلیل آنها با استفاده از روش‌های آماری، باعث درک و شناخت بهتر خشکسالی می‌شود. از آنجا که این دو مولفه دارای همبستگی بالایی با یکدیگر می‌باشند، لذا تحلیل توأم آنها نتایج بهتری را نمایان می‌سازد. هدف از این مقاله استفاده از توابع مفصل دوبعدی برای ساخت توزیع آماری دومتغیره شدت و مدت برای تحلیل خصوصیات احتمالاتی خشکسالی از جمله احتمال وقوع و دوره بازگشت توأم می‌باشد که در تدوین صحیح و کارآمد طرح‌های مقابله با خشکسالی می‌تواند بسیار موثر واقع شود. بدین منظور شدت و مدت خشکسالی با استفاده از شاخص خشکسالی SPEI برای سری بارش ماهیانه شهر مشهد در دوره آماری ۲۰۱۷-۱۸۹۳ برآورد

^۱ rvahdati@um.ac.ir

^۲ k.davary@gmail.com

^۳

^۴

شده و سپس با استفاده از توابع مفصل، احتمال وقوع و دوره بازگشت توأم خشکسالی مورد مطالعه و بررسی قرار گرفت.

کلمات کلیدی: توابع مفصل، خشکسالی هواشناسی، شاخص SPEI، روش حداکثر درست‌نمایی، تحلیل دو متغیره.

۱ پیش‌گفتار

خشکسالی یک پدیده طبیعی و تکرار شونده از اقلیم است ولی مشخصات و تعاریف آن بسته به تفاوت‌های اقلیمی، نیازها و نظام‌های موجود و راهبردهای اتخاذ شده در نقاط مختلف، از یک منطقه به منطقه دیگر کاملاً متفاوت است [۳]. گرچه تعاریف متفاوتی برای خشکسالی ارائه شده لیکن این پدیده در کل حاصل کمبود بارش در طی یک دوره ممتد زمانی معمولاً یک فصل یا بیشتر و در منطقه‌ای وسیع می‌باشد که این شرایط تصادفی منجر به کاهش شدید دسترسی به آب (نسبت به حالت طبیعی) شود [۴]. در تحقیقات صورت گرفته در ارزیابی خشکسالی‌ها، بیشترین تمرکز بر خشکسالی هواشناسی است که از دلایل آن وجود اطلاعات وسیع‌تر و همچنین تفکر اینکه بارش عامل اصلی خشکسالی است، می‌باشد [۵]. برای اولین بار خسارات ناشی از بلایای طبیعی از سال ۱۹۰۰ تا ۲۰۱۵ بررسی شدند. در طول این دوره آماری بیش از ۳۰۰۰۰ رخداد خسارت اقتصادی در سطح جهانی تفکیک شدند که خسارتی بالغ بر ۷ تریلیون دلار در پی داشته‌اند و ۱۲٪ این خسارت به دلیل خشکسالی‌های رخ داده در سطح جهانی بوده است [۶]. خشکسالی در ایران نیز به گونه‌ی است که هیچ نقطه‌ای از کشور از این پدیده در امان نبوده و پیوسته در اغلب سال‌ها نقطه‌ای از کشور با این پدیده مواجه بوده است. در این میان، قسمت‌های مرکزی، جنوب و جنوب شرقی کشور، به علت نواسانات زیاد مقادیر بارندگی در سال‌های مختلف، در معرض آسیب بیشتری قرار دارند. بنابراین بررسی و شناخت دوره‌های مرطوب و خشک و پیش‌بینی آنها، خود باعث کاهش اثرات خشکسالی و خسارات وارده به محیط زیست و مردم منطقه خواهد شد [۱]. مدیریت مناسب خشکسالی، مستلزم شناخت فراوانی مورد انتظار یا احتمال وقوع شرایط خاص آب و هوایی است [۷]. بنابراین شناخت ویژگی‌های خشکسالی، یک عنصر ضروری در برنامه‌های مدیریت ریز خشکسالی است [۸]. از جمله روش‌های موجود در این زمینه می‌توان به روش تحلیل تک متغیره و تحلیل دو متغیره خشکسالی اشاره کرد. برای پایش خشکسالی نمی‌توان تنها از یک مشخصه استفاده کرد چرا که پایش خشکسالی بر اساس یک مشخصه نمی‌تواند شرایط منطقه را به خوبی نشان دهد. برای حل این مشکل می‌توان از توزیع‌های دو یا چندمتغیره استفاده نمود [۲]. توزیع‌های دو متغیره بطور معمول بیشتر در تحلیل سیلاب بکار رفته‌اند که شامل توزیع نرمال [۹]، نمایی [۱۰]، گاما [۱۱] و حدی [۱۲] می‌باشند. یکی از ضعف‌های اساسی توزیع‌های دو متغیره ذکر شده، استفاده از توزیع آماری

یکسان برای دو متغیر مورد نظر می‌باشد، به این معنی که با استفاده از این روش‌ها نمی‌توان دو متغیر همبسته که از توزیع‌های مختلفی (مثلاً یکی نرمال و دیگری نمایی) پیروی می‌کنند را مدل کرد. [۲] [۱۳] بنابراین، توزیع‌های چندمتغیره برای تحلیل خشکسالی‌ها مناسب نمی‌باشند، زیرا شدت و مدت خشکسالی اغلب از توزیع‌های متفاوت تبعیت می‌کنند. همچنین با پیچیدگی روابط ریاضی توزیع چند متغیره در نتیجه افزایش تعداد متغیرها، تشخیص رفتار توزیع‌های حاشیه‌ای و توأم امکان‌پذیر نمی‌باشد. [۱۳] [۱۴] برای بررسی رفتار آماری متغیرهای وابسته، اخیراً توابع مفصل^۱ به عنوان یک ابزار ریاضیاتی مفید برای کاربردهای هیدرولوژیکی مورد استفاده قرار گرفته‌اند [۱۳] [۱۵] توابع مفصل که ابتدا توسط اسکالر (۱۹۵۹) [۱۶] برای ایجاد توزیع‌های چند متغیره معرفی گردید، امکان ترکیب چند توزیع یک متغیره از خانواده‌های متفاوت را برای ساخت یک توزیع چند متغیره با در نظر گرفتن همبستگی بین متغیرها به وجود می‌آورد. دمیکله و سالوردی (۲۰۰۳) [۱۷] نخستین بار این توابع را در مطالعات هیدرولوژی برای ایجاد یک مدل دو متغیره توصیف کننده شدت و مدت رگبار به کار بردند. پس از آن پژوهشگران دیگر در سایر پدیده‌ها مانند سیلاب، خشکسالی، آب زیرزمینی و ... از توابع مفصل بطور گسترده استفاده نمودند. شیائو (۲۰۰۶) [۱۳] در پژوهشی توزیع احتمال توأم دو متغیره خصوصیات خشکسالی (مدت و بزرگی خشکسالی) در وسانتو واقع در جنوب تایوان را با استفاده از شاخص بارش استاندارد شده و تئوری توابع مفصل مورد ارزیابی قرار داد. در واقع او برای اولین بار توابع مفصل را بر متغیرهای مدت و سختی خشکسالی برازش داد و با رسم منحنی‌های احتمال توأم و منحنی‌های دوره بازگشت توأم، درباره مزیت‌های مدل سازی دو متغیره خشکسالی‌ها بر اساس توابع مفصل بحث کرد. کلیه مطالعات انجام شده در این زمینه به توانایی و قابلیت بسیار بالا تکنیک توابع مفصل در مدل سازی خصوصیات همبسته خشکسالی به صورت دو یا چند متغیره تاکید دارند. هدف از این مقاله استفاده از توابع مفصل دوبعدی برای ساخت توزیع آماری دو متغیره شدت و مدت برای تحلیل خصوصیات احتمالاتی خشکسالی دشت مشهد، از جمله احتمال وقوع و دوره بازگشت توأم می‌باشد. تحلیل احتمالاتی دو متغیر شدت و مدت خشکسالی در تدوین صحیح و کارآمد طرح‌های مقابله با خشکسالی می‌تواند بسیار موثر باشد.

۲ شاخص بارش- تبخیر و تعرق استاندارد شده SPEI

در این تحقیق جهت تعیین خشکسالی از شاخص SPEI^۲ استفاده شده است. جهت برآورد این شاخص ابتدا با برآورد مقدار تبخیر و تعرق پتانسیل از روش تورنت وایت، اختلاف بین بارش (P) و تبخیر و تعرق پتانسیل

^۱ Copula Functions

^۲ Standardized Precipitation Evapotranspiration Index

(PET)^۳ برای هر ماه بدست می‌آید. درنهایت با تبدیل مقادیر احتمال تجمعی به تابع نرمال استاندارد، می‌توان مقادیر شاخص SPEI را بدست آورد [۱۸]. شرح کامل محاسبات و شیوه برآورد SPEI در وینسنته سرانو (۲۰۱۰) [۱۸] یافت می‌شود.

جدول ۱: طبقه بندی شدت خشکسالی براساس شاخص SPEI

مقادیر SPEI	طبقه بندی
۲ یا بیشتر	ترسالی خیلی شدید
۱,۵ تا ۱,۹۹	ترسالی شدید
۱ تا ۱,۴۹	ترسالی متوسط
۰,۹۹ تا -۰,۹۹	نرمال
-۱ تا -۱,۴۹	خشکسالی متوسط
-۱,۵ تا -۱,۹۹	خشکسالی شدید
-۲ یا کمتر	خشکسالی خیلی شدید

۳ مشخصات خشکسالی

مشخصه‌های شدت و مدت خشکسالی بوسیله شاخص‌های خشکسالی، قابل اندازه‌گیری و کمی شدن هستند. با برآورد شاخص‌های مورد نظر، مقدار شدت خشکسالی به صورت تجمعی از رابطه (۱) بدست می‌آید که در آن D معرف مدت خشکسالی که به صورت مدت زمان وقوع مستمر خشکسالی تعریف می‌شود و S معرف شدت خشکسالی و Index شاخص خشکسالی در زمان منفی بودن آن می‌باشد [۱۳].

$$S = - \sum_{i=1}^D Index \quad (1)$$

۴ توزیع تک متغیره خشکسالی

در این تحقیق به منظور یافتن تابع توزیع مناسب بر متغیرهای شدت و مدت خشکسالی برآورد شده از شاخص SPEI از هفت تابع توزیع گاما دوپارامتری، نمایی، ویبول، گامبل، لوگ نرمال دوپارامتری، لوجیستیک و مقادیر حدی تعمیم یافته برای برازش و انتخاب بهترین توزیع حاشیه‌ای متغیر شدت و مدت خشکسالی

^۳ Potential Evapotranspiration

استفاده شده است. تابع توزیع مناسب برازش یافته بر شدت و مدت حاصل از شاخص SPEI، برای هر دو متغیر، مقادیر حدی تعمیم یافته تشخیص داده شد که تابع چگالی احتمال آن از رابطه (۲) بدست می آید. پارامترهای توابع حاشیه‌ای شامل موقعیت k ، مقیاس σ و شکل μ توسط روش حداکثر درست‌نمایی^۴ برآورد گردیدند.

$$f(x) = \begin{cases} \frac{1}{\sigma} \exp(-(1+kz)^{-\frac{1}{k}})(1+kz)^{-1-\frac{1}{k}} & K = 0 \\ z = \frac{x-\mu}{\sigma} & (2) \\ \frac{1}{\sigma} \exp(-z - \exp z) & K \neq 0 \end{cases}$$

۵ توابع مفصل

توابع مفصل که براساس تئوری اسکالر (۱۹۵۹) تعریف شدند، توابع توزیع تک متغیره را برای تشکیل توابع توزیع چند متغیره به هم پیوند می‌دهند [۱۳]. تحلیل با استفاده از دو متغیر وابسته که دارای توابع حاشیه‌ای مجزا هستند به جای تحلیل با استفاده از یک تابع توزیع دو متغیره، از مزایای بکارگیری توابع مفصل می‌باشد [۱۹]. طبق قضیه اسکالر با فرض اینکه S و D دو متغیر تصادفی وابسته مانند شدت و مدت خشکسالی با تابع توزیع دو متغیره $F_{D,S}$ و توابع توزیع حاشیه‌ای $F_S(s)$ و $F_D(d)$ باشند آنگاه تابع مفصل دو متغیره C را می‌توان برای آن تعریف کرد (رابطه ۳) که در آن $F_D(d)$ تابع حاشیه‌ای مدت و $F_C(c)$ تابع حاشیه‌ای شدت خشکسالی است.

$$F_{D,S}(d, s) = C(F_D(d), F_S(s)) \quad (3)$$

با قبول این فرض که توابع توزیع حاشیه‌ای پیوسته با تابع چگالی احتمال $f_D(d)$ و $f_S(s)$ باشند تابع چگالی احتمال توأم آنها از رابطه (۴) بدست می آید که در آن c تابع چگالی تابع مفصل C است و از رابطه (۵) استخراج می‌شود.

$$f_{D,S}(d, s) = c(F_D(d), F_S(s))f_D(d)f_S(s) \quad (4)$$

$$c(u, v) = \frac{\partial^2 c(u, v)}{\partial u \partial v} \quad (5)$$

در معادله فوق u و v توابع توزیع احتمال حاشیه‌ای دو متغیر مورد بررسی می‌باشد.

در این تحقیق چهار نوع از توابع مفصل دوبعدی یک پارامتری مرسوم گروه ارشمیدس شامل کلایتون، فرانک، گامبل و جو که بطور گسترده در تحقیقات هیدرولوژیکی مورد استفاده قرار گرفته‌اند به دلایلی همچون

^۴Maximum Likelihood

جدول ۲: توزیع تجمعی و چگالی احتمال برازش داده شده بر داده‌ها

نام تابع مفصل	تابع توزیع احتمال	تابع چگالی احتمال	فضای پارامتر
کلایتون	$(u^\theta + v^\theta - 1)^{-\frac{1}{\theta}}$	$(\theta + 1)(u^\theta + v^\theta - 1)^{\frac{2}{\theta}-2} (uv)^{\theta-1}$	$\theta \in [-1, \infty) \setminus 0$
فرانک	$-\frac{1}{\theta} \ln \left\{ \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{(e^{-\theta} - 1)} \right\}$	$-\frac{\theta e^{-\theta(u+v)}(e^{-\theta} - 1)}{(e^{-\theta(u+v)} - e^{-\theta u} - e^{-\theta v} + e^{-\theta})^2}$	$\theta \in [-\infty, \infty) \setminus 0$
گامبل	$\exp[-\{(-\ln u)^\theta + (-\ln v)^\theta\}^{\frac{2}{\theta}}]$	$\frac{[(\ln u)^\theta + (-\ln v)^\theta]^{\theta-1}}{uv} \{ [(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{2}{\theta}-2} \{ (\theta-1) [(-\ln u)^\theta + (-\ln v)^\theta]^{\frac{1}{\theta}-1} + 1 \}$	$\theta \in [1, \infty)$
جو	$1 - [(1-u)^\theta + (1-v)^\theta - (1-u)^\theta + (1-v)^\theta]^{\frac{1}{\theta}}$	$(1-u)^\theta + (1-v)^\theta - (1-u)^\theta + (1-v)^\theta]^{\frac{1}{\theta}-2} \times [(1-u)(1-v)]^{\theta-1} \times [(\theta+1) + (1-u)^\theta + (1-v)^\theta - ((1-u)(1-v))^\theta]$	$\theta \in [1, \infty)$

کاربرد رضایت بخش، ساختار آسان و انعطاف‌پذیری در انتخاب توزیع‌های حاشیه‌ای، جهت برازش بر داده‌ها و انتخاب مناسب‌ترین تابع از میان آنها به‌کار گرفته شد. توابع توزیع تجمعی^۵ و چگالی احتمال^۶ به همراه فضای پارامتر همبستگی مربوط به سه تابع مفصل ذکر شده در جدول ۲ نمایش داده شده است. در این تحقیق پارامترهای توابع مفصل با روش استنتاج حاشیه‌ای IFM^۷ برآورد شدند. روش IFM یک روش تماماً پارامتری می‌باشد که شامل دو مرحله مختلف است. در مرحله اول پارامترهای توزیع‌های حاشیه‌ای یک متغیره، به‌صورت جداگانه و با روش حداکثر درست‌نمایی محاسبه می‌شود، سپس در مرحله دوم پارامتر مفصل با دو توزیع حاشیه‌ای شدت $F_S(s)$ و مدت $F_D(d)$ که برای هر کدام دو پارامتر در نظر گرفته شده است، با استفاده از حداکثر نمودن تابع لگاریتم درست‌نمایی زیر بدست می‌آید [۱۳]:

$$\ln L(d, s; \alpha_1, \alpha_2, \beta_1, \beta_2, \theta) = \ln L_C(F_D(d), F_S(s); \theta) + \ln L_D(F_D(d); \alpha_1, \alpha_2) + \ln L_S(F_S(s); \beta_1, \beta_2) \quad (۶)$$

در رابطه فوق s و d به ترتیب بیانگر متغیرهای شدت و مدت خشکسالی، مقادیر α_1, α_2 پارامترهای توزیع تجمعی حاشیه‌ای مدت، β_1, β_2 پارامترهای توزیع تجمعی حاشیه‌ای شدت، θ پارامتر همبستگی تابع مفصل، L_C لگاریتم درست‌نمایی تابع مفصل، L_D لگاریتم درست‌نمایی تابع توزیع مدت و L_S لگاریتم درست‌نمایی تابع توزیع شدت می‌باشند.

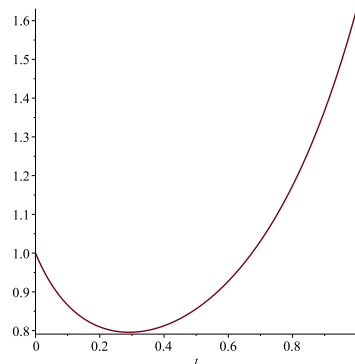
۶ احتمال تجاوز توأم و احتمال شرطی

با در اختیار داشتن توزیع دو متغیره، احتمال اینکه مقادیر شدت و مدت خشکسالی به طور همزمان از حد معینی بیشتر شود، اطلاعات مفیدی را برای مدیران منابع آب جهت ارتقای مدیریت سیستم‌های منابع آب در

^۵Cumulative Distribution Function

^۶Probability Distribution Function

^۷Inference Functions for Margins



شکل ۱: سری زمانی شاخص SPEI سالیانه در دوره آماری ۲۰۱۷-۱۸۹۳ برای ایستگاه سینوپتیک دشت مشهد

شرایط خشکسالی فراهم می‌کند. این احتمال به همراه توزیع احتمال شرطی خشکسالی را نیز می‌توان با استفاده از توابع مفصل تعیین و بر اساس آنها، احتمال چگونگی تغییر یک عامل در قبال تغییرات کنترل شده عامل دیگر، بررسی کرد. روابط احتمال توأم و توزیع احتمال شرطی شدت و مدت خشکسالی در [۱۳] به تفصیل بیان شده است.

۷ دوره بازگشت توأم و دوره بازگشت شرطی

برنامه‌ریزی و مدیریت سیستم‌های منابع آب که تحت تأثیر خشکسالی هستند، بیشتر نیاز به تعیین دوره بازگشت‌های خشکسالی بر اساس مشخصه‌های آن دارد [۲۰]. لذا دوره بازگشت خشکسالی باید بر اساس هر دو مشخصه شدت و مدت تعیین شود. برای تعیین دوره بازگشت دو متغیره خشکسالی در دو حالت متفاوت تجاوز همزمان دو متغیر از سطح آستانه خاص یا تجاوز یکی از دو متغیر از سطح آستانه خاص از روابطی که توسط شیائو (۲۰۰۶) ارائه شده، می‌توان استفاده کرد. این روابط در [۱۳] تشریح شده است.

۸ نتایج و بحث

در این تحقیق، شاخص خشکسالی SPEI در مقیاس سالیانه، که مقیاسی جهت برآورد خشکسالی متوسط و مرتبط در حوزه تصمیم‌گیری در زمینه منابع آب می‌باشد، با استفاده از تابع توزیع لوگ لجستیک برای دوره آماری ۲۰۱۷-۱۸۹۳ برآورد گردید. وضعیت خشکسالی بر مبنای شاخص مورد نظر برای طول دوره آماری در شکل ۱ ترسیم شده است. با توجه به نتایج خلاصه شده در جدول ۳، مقدار لگاریتم درست‌نمایی برای تابع جو

جدول ۳: مقادیر P_Value و آماره آزمون کرامرون مایسز، θ ، NS ML، RMSE، AIC برای توابع توزیع مفصل

P_Value	Sn	AIC	NS	RMSE	ML	θ	تابع توزیع
۰/۰۰۲۴	۰/۱۱۴۵	-۱۱۶/۵۱۹۱	۰/۹۵۷۹	۰/۰۵۶۴	۱/۰۷۲۵	۰/۳۹۴۲	کلایفون
۰/۵۲۵*	۰/۰۲۶۸	-۱۵۲/۱۸۲۸	۰/۹۹۲۷	۰/۰۲۳۷	۸/۱۱۸۶	۱/۷۴۷۷	کاهیل
۰/۰۵۴۴*	۰/۰۴۴۶	-۱۴۴/۶۵۱۲	۰/۸۸۶۴	۰/۰۲۳۸	۵/۱۰۱۱	۲/۹۷۰۸	فرلنک
۰/۹۴۶۶*	۰/۰۱۸۹	-۱۵۷/۸۳۸۳	۰/۹۹۶۱	۰/۰۱۸۴	۱۰/۲۰۸۱	۲/۴۷۵۸	جو

علامت * بیانگر رد نشدن آزمون در سطح ۱٪ می باشد.

از بقیه توابع بیشتر می باشد. همچنین مقادیر جذر میانگین مربعات خطا^۸ و آماره اطلاعات آکائیک^۹ کمتر و آماره نش-سانکلیف^{۱۰} بیشتر برای تابع جو در مقایسه با سایر توابع، نشان دهنده برتری برازش این تابع است. بنابراین در این تحقیق جهت ساخت تابع توزیع دومتغیره خشکسالی و مدل سازی ساختار همبستگی بین شدت و مدت خشکسالی، تابع مفصل جو انتخاب شد. یکی از پارامترهای اساسی که پس از ساخت توزیع دومتغیره می توان به آن دست یافت، احتمال توأم تجاوز دومتغیر از آستانه مشخصی است. با در اختیار داشتن توزیع دو متغیره، احتمال اینکه مقادیر شدت و مدت خشکسالی به طور همزمان از حد معینی بیشتر شود، اطلاعات مفیدی را برای مدیران منابع آب جهت ارتقای مدیریت سیستم های منابع آب در شرایط خشکسالی فراهم می کند. اگر بخواهیم احتمال وقوع خشکسالی براساس شاخص SPEI با شدت کمتر از ۲ ($F_S(2) = 0.8326$) و ۵ ($F_S(5) = 0.9533$) در صورتی که مدت بیشتر از ۴ ماه باشد را محاسبه کنیم، داریم:

$$F_D(4) = 0.8943, C(F_D(4), F_S(2)) = 0.8130, C(F_D(4), F_S(5)) = 0.8889$$

در نتیجه احتمال وقوع توأم خشکسالی با شدت های کمتر از ۲ و ۵ و با فرض بیشتر بودن خشکسالی از ۴ ماه به ترتیب $P(S \leq s | D \geq d') = 0.3443$ ، $P(S \leq s | D \geq d) = 0.1048$ می باشد. خطوط تراز احتمال وقوع توأم شدت و مدت خشکسالی در شکل ۲ نشان داده شده است، که در واقع در این شکل، خطوط تراز نشان دهنده احتمالات مختلف وقوع خشکسالی می باشد. از آنجا که متغیرهای شدت و مدت خشکسالی نقش بسزایی در مدیریت و ارزیابی خشکسالی دارند لذا برآورد دوره بازگشت توأم خشکسالی حائز اهمیت می باشد. برای مثال بیشترین شدت و مدت خشکسالی براساس شاخص SPEI برای دشت مشهد به ترتیب برابر ۲۳/۳۹۶۱ و ۱۹ سال می باشد. احتمال وقوع توأم خشکسالی با شدت و مدت مذکور بر اساس محاسبات

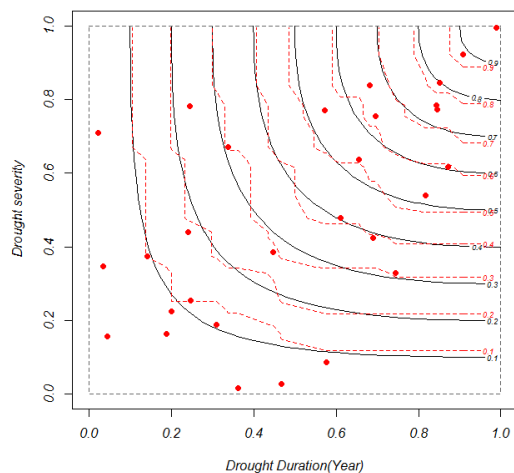
$$F_S(23/3961) = 0.9956, F_D(19) = 0.9884, C(F_D(19), F_S(23/3961)) = 0.9900$$

برابر است با $P(D \geq d, S \geq s) = 0.006$. همچنین برای این شدت و مدت خشکسالی دوره بازگشت در دو حالت تجاوز توأم یا یکی از دو متغیر از آستانه ای خاص، به ترتیب برابر $T_{DS} = 166/66$ و $T'_{DS} = 100$ می باشد. شکل ۳ نشان دهنده خطوط تراز دوره بازگشت در دو حالت تجاوز توأم یا یکی از دو متغیر از آستانه

^۸Root Mean Square Error

^۹Akaike Information Criterion

^{۱۰}Nash-Sutcliffe



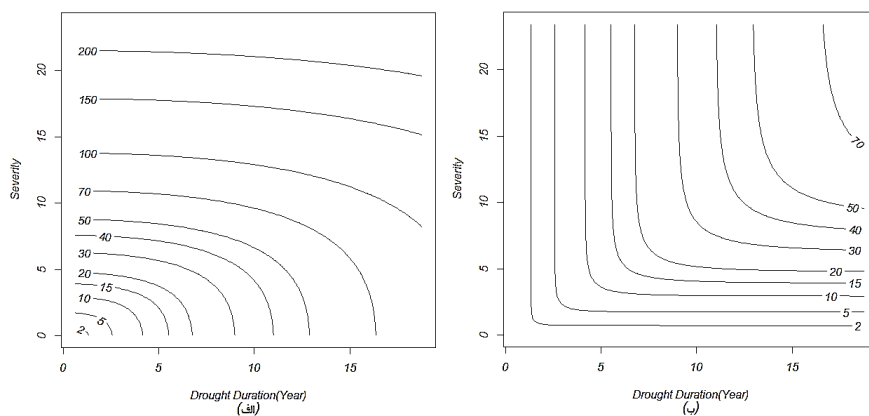
شکل ۲: پراکنش شدت و مدت خشکسالی همراه با خطوط تراز احتمال وقوع توأم خشکسالی با استفاده از تابع مفصل جو (خطوط ممتد مشکی) و تابع توزیع مفصل تجربی داده‌ها (خط چین قرمز)

خاص است که در واقع منحنی‌های شدت، مدت و فراوانی (SDF) ^{۱۱} برای منطقه مورد مطالعه می‌باشند. همچنین در شکل ۴ دوره بازگشت شرطی شدت و مدت خشکسالی در مقابل دوره بازگشت تک‌متغیره شدت و مدت ترسیم شده است.

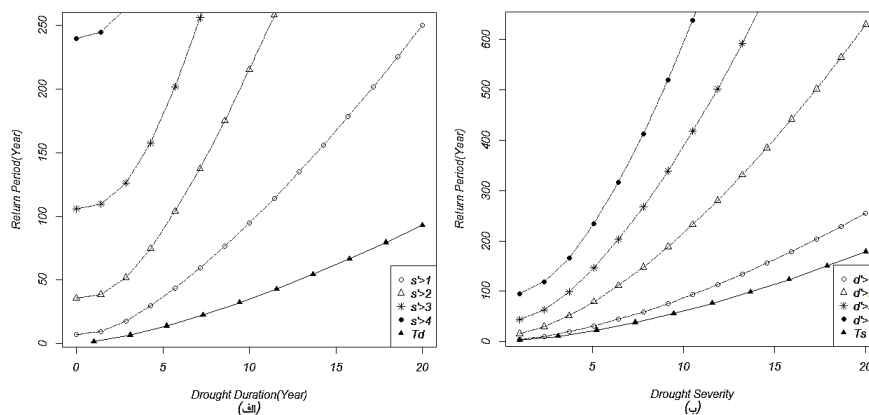
۹ نتیجه‌گیری

در این تحقیق تحلیل خشکسالی و اصول ساخت توزیع آماری دو متغیره از شدت و مدت خشکسالی بر مبنای توابع مفصل مورد مطالعه و بررسی قرار گرفت. تحلیل چند متغیره خشکسالی با استفاده از توابع مفصل، به جای تحلیل‌های تک متغیره این پدیده، نتایج دقیق‌تر و مطمئن‌تری را حاصل می‌شود. برای این منظور با استفاده از بارش و دمای ماهیانه در طول دوره آماری ۱۸۹۳-۲۰۱۷ ایستگاه سینوپتیک دشت مشهد، شاخص خشکسالی SPEI در مقیاس سالیانه برآورد شد. بر اساس آزمون نکویی برازش، تابع توزیع مقادیر حدی تعمیم یافته جهت ایجاد توزیع‌های حاشیه‌ای شدت و مدت خشکسالی مناسب تشخیص داده شدند. همچنین با استفاده از آزمون نکویی برازش کرامرون مایسز و دیگر آماره‌ها، تابع مفصل جو در مقایسه با سایر توابع بررسی شده نتایج بهتری را در برازش بر توزیع‌های حاشیه‌ای نمایان ساخت و سپس احتمال وقوع و دوره بازگشت توأم برای منطقه مورد مطالعه بدست آمد. بیشترین شدت خشکسالی در طول دوره بررسی شده برای منطقه مورد مطالعه برابر ۲۳,۳۹۶۱ و مدت متناظر با این خشکسالی برابر ۱۹ سال می‌باشد. لذا احتمال وقوع

^{۱۱} Severity-Duration-Ferquency



شکل ۳: خطوط تراز دوره بازگشت توأم شدت و مدت خشکسالی در حالت تجاوز همزمان دو متغیر از آستانه‌ای خاص (الف) و در حالت تجاوز یکی از دو متغیر از آستانه‌ای خاص (ب).



شکل ۴: دوره بازگشت شرطی شدت و مدت خشکسالی در مقابل دوره بازگشت تک متغیره برای حالت مقادیر مختلف شدت (الف) و برای حالت مقادیر مختلف مدت (ب).

توأم برای شدت و مدت مذکور برابر ۰,۰۰۶ خواهد بود که دوره بازگشت توأم این متغیرها در دو حالت تجاوز همزمان و یا تجاوز یکی از متغیرها از آستانه‌ای خاص، به ترتیب برابر ۱۶۶,۶۶ و ۱۰۰ سال برآورد شد. نتایج حاصل نشان می‌دهد با توجه به اینکه استفاده از دوره بازگشت توأم و شرطی غالباً منجر به تغییر در مقادیر دوره بازگشت می‌گردد، از این رو بررسی دوره بازگشت دو متغیر شدت و مدت خشکسالی که دارای همبستگی با یکدیگر هستند، به صورت تک متغیره توصیه نمی‌شود. همچنین با استفاده از توابع مفصل، به راحتی می‌توان همبستگی بین متغیرهای پدیده هیدرولوژیک را مدل‌سازی کرد که این امر باعث شناخت و درک بهتر رفتار پدیده‌های هیدرولوژیکی مانند خشکسالی می‌شود و اطلاعات مفیدی را برای مدیران منابع آب جهت ارتقای مدیریت سیستم‌های منابع آب در شرایط خشکسالی فراهم می‌کند.

مراجع

- [۱] دانشور، م، ر، تلوری، ع، توکلی، م، دانائیان، م، ر، (۱۳۸۵)، تحلیل دوره‌های بازگشت خشکسالی در شرق و جنوب شرق کشور. ۲۱ تا ۳۶.
- [۲] فرخ‌نیا، الف، مرید، س، (۱۳۸۷)، تحلیل شدت و مدت خشکسالی با استفاده از توابع مفصل. مجموعه مقالات چهارمین کنگره ملی مهندسی عمران، اردیبهشت ماه، تهران. ۱ تا ۸.
- [3] Lupu, A, B., Ionescu, F, C., Borza, L. 2010. The Phenomenon of Drought and it's Effects Within Romania. Research Journal of Agricultural Science, 42 (4): 102-109.
- [4] Rossi, G. 2000. Drought mitigation measures: a comprehensive framework. In: J.V. Vogt and F. Somma (eds.), Drought and Drought Mitigation in Europe, 233-246.
- [5] Mirakbari, M., Ganji, A., Fallah, S. R., 2010. "Regional bivariate frequency analysis of meteorological droughts", J Hydrol Eng, 15(12):985-1000.
- [6] Daniell, J. Wenzel, F. Schaefer, A. The economic costs of natural disasters globally from 1900-2015: historical and normalised floods, storms, earthquakes, volcanoes, bushfires, drought and other disasters. Geophysical Research Abstracts.
- [7] Hao, C., Zhang, J., Yao, F., Multivariate drought frequency estimation using copula method in Southwest China, Theoretical and Applied Climatology, February 2017, Volume 127, Issue 3-4, pp 977-991.
- [8] Shiau, J. T., and Modarres, R. 2009. Copula-based drought severity-duration frequency analysis in Iran. Meteorological Applications, 16: 481-489.

- [9] Goel, N. K., Seth, S. M. and Chandra, S. (1998) Multivariate modeling of flood flows. *Journal of Hydraulic Engineering ASCE*, 124(2), 146–155.
- [10] Bacchi, B., Becciu, G. and Kottegoda, N. T. (1994) Bivariate exponential model applied to intensities and durations of extreme rainfall. *Journal of Hydrology*, 155, 225–236.
- [11] Yue, S., Ouarda, T. B. M. J. and Bob´ee, B. (2001) A review of bivariate gamma distributions for hydrological application. *Journal of Hydrology*, 246, 1–18.
- [12] Shiau, J. T. (2003) Return period of bivariate distributed hydrological events. *Stochastic Environmental Research and Risk Assessment*, 17(1–2), 42–57.
- [13] Shiau J. T. 2006. Fitting Drought Duration and Severity with Two-Dimensional Copulas. *Water Resources Management*. 20:759-815.
- [14] Grimaldi, S. and Serinaldi, F. 2006. Asymmetric copula in multivariate flood frequency analysis. *Adv. Water Resour.* 29(8): 1115-1167.
- [15] Salvadori, G., De Michele, C., Kottegoda, N. T., and Rosso, R. 2007. *Extremes in Nature, An Approach Using Copulas*. Springer. Netherlands.
- [16] Sklar, A. 1959. Distribution functions, dimensions and margins. *Publications of the Institute of Statistics*, 1: 229-231.
- [17] De Michele, C., and Salvadori, G. 2003. A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. *Journal of Geophysical Research*, 108:1-10.
- [18] Vicente-Serrano, Sergio M., Santiago Beguería, and Juan I. López-Moreno. (2010) "A multi-scalar drought index sensitive to global warming: the standardized precipitation evapotranspiration index" *Journal of Climate* 23.7: 1696-1718.
- [19] Sriramula, S., Menon, D., Prasad, A. M. (2006). Multivariate simulation and multimodal dependence modeling of vehicle axle weights with copulas. *Journal of Transportation Engineering*, 132(12), 945-955.
- [20] Kim, T. W., Valdés, J. B., Yoo, C. 2003. Nonparametric approach for estimating return periods of droughts in arid regions. *Journal of Hydrologic Engineering*, 8: 237–246.



مدل سازی وابستگی ریسک‌های بیمه‌گری توسط کاپولا و کاربرد آن در محاسبه توانگری مالی

آراء، ش ۱ ملک‌پور، س.ا. ۲ جناتی کاشانی، ر ۳

مدیریت اکچوثری بیمه سامان

چکیده

در راستای اجرای طرح نظارت مالی، آیین نامه توانگری مالی که توسط بیمه مرکزی به شرکت‌های بیمه ابلاغ شده، شرکت‌ها را ملزم به محاسبه حاشیه توانگری مالی می‌کند. در حالیکه همبستگی ریسک‌ها و اثر متقابل آنها نقش مهمی در محاسبه ارزش در معرض خطر (VaR) دارد، در آیین نامه توانگری مالی فرض بر استقلال ریسک‌ها است. بنابراین ارائه مدلی که این همبستگی را در نظر بگیرد از اهمیت زیادی برخوردار است. کاپولا یکی از ابزارهای مفید در بررسی همبستگی بین ریسک‌ها می‌باشد که هم اکنون بسیاری از شرکت‌های معتبر دنیا از آن به عنوان روشی داخلی برای محاسبه VaR بهره می‌برند.

در این مقاله با استفاده از کاپولا، وابستگی بین ریسک بیمه‌گری در بیمه نامه‌های بدنه و شخص ثالث اتومبیل در شرکت بیمه سامان مدل شده است. سپس یک روش محاسباتی برای برآورد VaR با استفاده از کاپولای برازش شده در مرحله قبل، پیشنهاد شده است. همچنین، برآورد VaR با نتایج مبتنی بر آیین نامه ۶۹ مقایسه شده است.

۱ sh_ara@samaninsurance.ir

۲

۳

روش ارائه شده در این مقاله، قابل تعمیم به سایر ریسک‌های بیمه‌ای نیز می‌باشد و می‌تواند با مدل کردن وابستگی بین ریسک رشته‌های مختلف، VaR را به صورت دقیقی برآورد کند. با توجه به پیچیدگی‌های خاصی که این روش دارد، می‌توان از آن به عنوان روشی داخلی برای برآورد VaR بهره برد.

کلمات کلیدی: حاشیه توانگری مالی، ارزش در معرض خطر، همبستگی، کاپولا.

۱ مقدمه

حرکت از نظارت فنی به نظارت مالی در دنیای امروز از جمله تحولاتی است که تقریباً در صنعت بیمه تمامی کشورهای پیشرو دیده می‌شود. نظارتی که ضمن ایجاد فضای سالم رقابتی برای شرکت‌های بیمه، توانمندی مالی آنها را در برآوردن تعهدات آتی کنترل می‌کند و معیاری را در اختیار نهاد ناظر قرار می‌دهد تا از طریق آن بتواند اولین قدم را در رتبه بندی شرکت‌ها بردارد. این امر نه تنها نظارت بر فعالیت شرکت‌های بیمه را میسر می‌سازد بلکه به عنوان ابزار شفاف سازی مالی میتواند اعتماد افراد جامعه را به این صنعت دو چندان کند. در حقیقت افشاگری مالی، شرکت‌ها را به سمتی سوق می‌دهد که تمام تلاش خود را در جهت افزایش سطح توانگری مالی برای جذب مشتری و حضور فعال در بازار انجام دهند. این امر سبب تغییر و تحولاتی در مدیریت ریسک شرکت‌ها خواهد شد.

توانگری مالی یک شرکت بیمه که نشان دهنده توانایی آن شرکت در پوشش تعهدات است این امکان را فراهم می‌سازد تا مشخص شود دارایی‌ها، منابع و سرمایه شرکت‌های بیمه برای پوشش تعهدات بیمه‌گر، اعم از کوتاه‌مدت و بلندمدت کفایت می‌کند یا خیر. حاشیه توانگری مالی یکی از اصلی‌ترین و مهمترین شاخص‌ها در ارزیابی توانگری مالی یک شرکت است. این شاخص در شرکت‌های بیمه ایران مطابق آیین نامه شماره ۶۹ محاسبه و پس از تایید بیمه مرکزی جمهوری اسلامی ایران منتشر می‌شود. در این آیین نامه ریسک‌های نقدینگی، اعتباری، بازار و بیمه‌گری مورد توجه قرار گرفته و روش محاسبه ارزش در معرض خطر (VaR) هر کدام از این ریسک‌ها بیان شده است. اما یکی از نکات مهمی که در آیین نامه ۶۹ در نظر گرفته نشده است ارتباط میان ریسک‌ها و ضریب همبستگی آنها است.

پر واضح است که در عمل ریسک‌ها از یکدیگر مستقل نیستند و فرض عدم وابستگی، عدم دقت در محاسبه حاشیه توانگری را در پی خواهد داشت. به طور مثال وقوع آتش سوزی در یک منطقه می‌تواند بر نرخ مرگ و میر در بیمه‌های زندگی تاثیرگذار باشد و یا وقوع خسارت در بیمه‌های ثالث ارتباط مستقیمی با بیمه نامه‌های بدنه دارد. علاوه بر اینکه در هر زیر مجموعه از ریسک‌های عنوان شده همبستگی وجود دارد، ریسک‌های یک گروه با گروه دیگر نیز ارتباط دارند به عنوان مثال تغییر نرخ بهره می‌تواند بر نرخ ریزش در بیمه‌های عمر

و پس انداز تاثیر گذار باشد.

یک روش برای در نظر گرفتن رابطه میان ریسک‌ها، همبستگی خطی است که در استاندارد Solvency II نیز از آن استفاده شده است. ولی این روش نمی‌تواند رابطه غیرخطی میان ریسک‌ها و به خصوص ریسک‌های رشته‌های مختلف بیمه‌ای را به خوبی بیان کند. به همین جهت، اخیراً در بسیاری از مطالعات اکچوئری از توابع کاپولا برای مدل کردن همبستگی میان ریسک‌ها استفاده شده است.

[۱۴] توابع کاپولا را برای اندازه‌گیری همبستگی میان ریسک‌های بیمه‌ای، برآورد زیان ناشی از مرگ و میر، تعدیل هزینه‌ها و قیمت گذاری قراردادهای اتکایی پیشنهاد دادند. علاوه بر این [۱۴] از توابع کاپولا برای تخمین اعتبار خسارتهای تجمعی استفاده کردند. [۱۵] اهمیت استفاده از کاپولا را در قراردادهای اتکایی نشان دادند. همچنین [۵] از کاپولاها برای ارزیابی مقدار ذخائر استفاده کردند.

[۶]، [۱۳] [۸] نشان دادند که یک مدل بر مبنای کاپولا می‌تواند تجمع ریسک‌ها را بیان و سرمایه‌ای بالاتر از حالتی که در آن ریسک‌ها مستقل فرض می‌شوند را برآورد کند. به طور مشابه [۱۲] از سه روش بوت استرپ، شوک عمومی و کاپولا برای مدل کردن همبستگی‌ها استفاده کرده و آن را برای تعیین ذخائر بیمه‌ای به کار برد.

[۷] سرمایه قابل تخصیص به کل پرتفولیو را با استفاده از معیار اندازه‌گیری ریسک $Tvar$ (متوسط خسارت در وضعیت عدم توانگری) و کاپولای فارلی — گامبل — مورگنشرن (FGM) برآورد کردند. اخیراً نیز [۲۰] مدل کاپولای نیمه پارامتریک را روی داده‌های خسارت بیمه‌های انفرادی بکار برده و بوسیله آن مقدار ذخائر خسارت را پیش‌بینی کردند. [۱۷] با استفاده از کاپولای دو متغیره و مدل‌های خطی ذخائر خسارتهای بیمه‌ای را بدست آوردند. از طرفی [۹] انعطاف کاپولای برنتین را در مدل کردن محصولات مختلف بیمه‌ای نشان داد.

در این تحقیق با استفاده از روش کاپولا، مدلی داخلی برای برآورد VaR با در نظر گرفتن همبستگی بین خسارت‌های بیمه بدنه و شخص ثالث اتومبیل — در شرکت بیمه سامان — ارائه می‌شود. در واقع، پس از مدل کردن توزیع توام خسارات بیمه بدنه و شخص ثالث اتومبیل توسط کاپولا، یک روش نیز برای برآورد VaR که مبتنی بر کاپولای برازش شده می‌باشد، پیشنهاد شده است. برای بیان بهتر مسئله نتایج این روش با نتایج استخراج شده از آئین نامه ۶۹ مقایسه شده است.

در بخش شرح مقاله و روش تحقیق، ابتدا انواع ریسک‌هایی که بیمه‌گران با آن مواجه هستند شرح داده شده است. سپس، مفهوم احتمالاتی ارزش در معرض خطر (VaR) و اهمیت برآورد این مقدار برای هر ریسک بیمه‌ای بیان شده است. در ادامه، خاصیت انواع مختلفی از کاپولا و توانایی آنها در مدل کردن وابستگی بین ریسک‌های بیمه‌ای مورد بحث قرار گرفته است.

در بخش نتایج نیز، ابتدا بر اساس روش ماکزیمم درستنمایی توزیع پارامتریک مناسبی به هر یک از داده‌های

خسارت بیمه بدنه و شخص ثالث اتومبیل برآزش شده است. سپس از این توزیع‌ها برای مدل کردن وابستگی بین خسارت‌ها توسط کاپولا استفاده شده است و سرانجام مقدار VaR برآورد شده است.

۲ شرح مقاله و روش تحقیق

در این بخش، ابتدا انواع مختلف ریسک‌های بیمه‌ای که یک بیمه‌گر با آن مواجه است توضیح داده شده‌اند و مفهوم ارزش در معرض خطر مربوط به هر ریسک بیان شده است. در ادامه نیز، ویژگی هر یک از انواع کاپولا در مدل کردن همبستگی بین ریسک‌های مختلف بیمه‌ای مورد تحلیل قرار گرفته است.

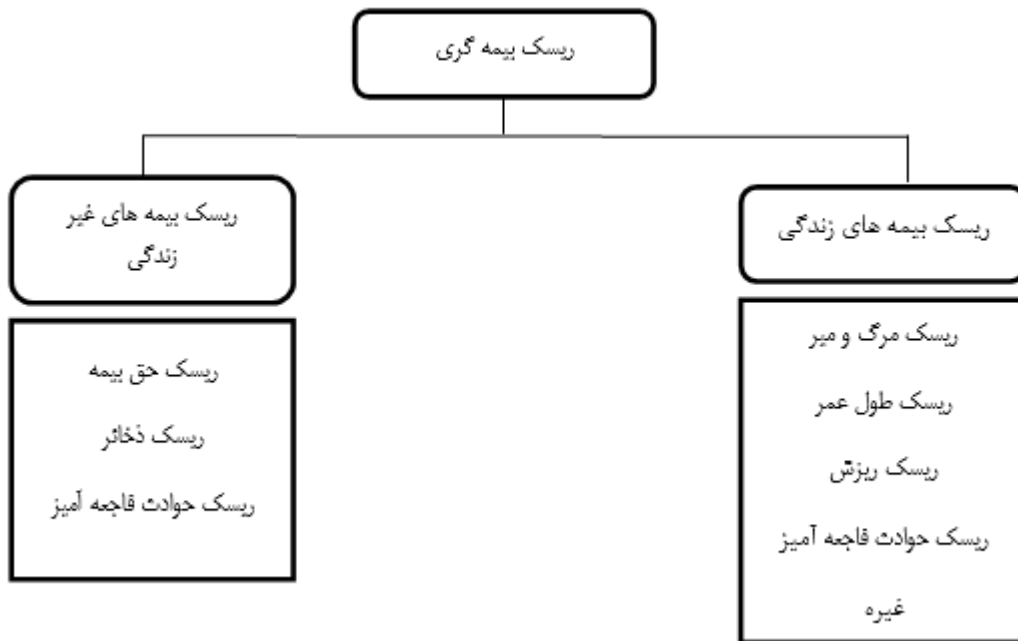
ریسک‌های بیمه‌گری، ریسک‌های بازار، ریسک‌های اعتباری، ریسک‌های عملیاتی، ریسک‌های نقدینگی از جمله ریسک‌هایی هستند که یک بیمه‌گر با آن مواجه است. این ریسک‌ها که در آیین نامه ۶۹ نیز آمده‌اند، در زیر تعریف و شرح داده شده‌اند.

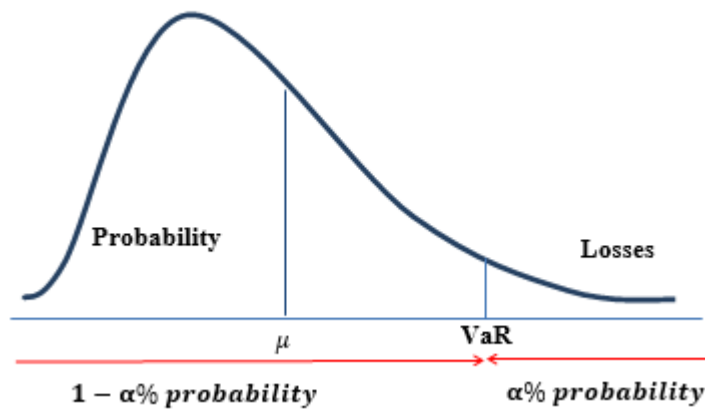
- ریسک‌های بیمه‌گری (ریسک صدور): این ریسک مربوط به ریسک‌های کسب و کار بیمه‌ای است که انواع مختلفی دارد. معمولاً ریسک‌های بیمه‌های زندگی و غیر زندگی به خاطر ماهیت متفاوت از هم جدا بررسی می‌شوند. دسته‌بندی این ریسک‌ها در شکل ۱ نشان داده شده‌اند. البته در برخی از استانداردهای توانگری از جمله Solvency II ریسک بیمه‌های درمان به طور جداگانه در نظر گرفته و محاسبه می‌شود.
- ریسک‌های بازار: این ریسک انواع مختلفی دارد که از جمله آنها می‌توان ریسک‌های نرخ بهره، سهام، املاک و مستغلات و ارز را نام برد.
- ریسک‌های اعتباری: ریسکی است که از قصور طرف قرار داد به وجود می‌آید. در واقع ریسک اعتباری ناشی از این است که طرف قرارداد نخواهد و یا نتواند تعهدات قرارداد را انجام دهد.
- ریسک‌های عملیاتی: این ریسک از نقصان فرآیندهای داخلی سازمان ناشی می‌شود. این ریسک شامل تقلب، اشتباهات مدیریتی، خطاهای انسانی و... می‌شود.
- ریسک‌های نقدینگی: این ریسک به قدرت نقد شوندگی دارایی‌ها و خالص جریان نقدینگی (دارایی‌ها جاری منهای بدهی‌های جاری) مربوط می‌شود.

۱.۲ مفاهیم مربوط به VaR

برای محاسبه حاشیه توانگری مالی لازم است ساختار دقیق تعهدات و ریسک‌هایی را که یک شرکت بیمه با آن مواجه است بشناسیم و بدقت آنها را محاسبه نماییم. برای اینکار از مفهوم ارزش در معرض خطر (VaR) استفاده می‌شود. ارزش در معرض خطر که در دهه ۱۹۹۰ معرفی شده، یک معیار بسیار مهم برای اندازه‌گیری

شکل ۱: اجزای مختلف ریسک بیمه‌گری





شکل ۲: مقدار VaR در روی تابع چگالی داده‌ها مشخص شده است

ریسک است و آن را به وسیله یک عدد بیان میکند. ارزش در معرض خطر بیشترین زیان مورد انتظار را برای یک بازه زمانی مشخصی در آینده (افق زمانی) — و در سطح اطمینان معینی — اندازه‌گیری می‌نماید. برای تعیین ارزش در معرض خطر دو پارامتر در نظر گرفته می‌شود:

(آ) سطح اطمینان $1 - \alpha$

(ب) افق زمانی مورد نظر

در واقع اندازه‌گیری ارزش در معرض خطر به احتمال زیان و افق زمانی مورد نظر بستگی دارد. به عنوان مثال، VaR را می‌توان به صورت فرمول زیر نیز توصیف کرد:

$$P(\Delta p_t \leq -Var(\alpha)) = \alpha$$

که در آن P_t ارزش پرتفو در زمان t و $\Delta p_t = p_t - p_{t-1}$ زیان یا سود پرتفو در فاصله زمانی $t-1$ و t است. رابطه فوق بیان می‌کند که احتمال اینکه کاهش ارزش پرتفو در زمان آتی بیشتر از VaR باشد حداکثر برابر α است. بنابراین اگر L متغیر تصادفی زیان (خسارت) پرتفو در بازه زمانی آتی و $F_L(l)$ تابع توزیع احتمال زیان (خسارت) باشد، VaR از فرمول زیر به دست می‌آید:

$$VaR_{1-\alpha} = \inf\{l \in \mathbb{R} : P(L > l) \leq \alpha\} = \inf\{l \in \mathbb{R} : F_L(l) \geq 1 - \alpha\}$$

و اگر تابع توزیع احتمال پیوسته باشد داریم:

$$VaR_{1-\alpha} = F^{-1}(1 - \alpha)$$

که در آن، $F^{-1}(1-\alpha)$ چندک $1-\alpha$ م تابع توزیع احتمال F است و با q نمایش داده می‌شود. در حقیقت ارزش در معرض خطر، چندک تابع توزیع احتمال زیان (خسارت) در سطوح بحرانی α است. روش‌های متعددی برای محاسبه ارزش در معرض خطر وجود دارد که می‌توان آنها را به دسته‌های زیر تقسیم نمود:

- روش‌های واریانس- کوواریانس
- روش‌های شبیه سازی تاریخی
- روش شبیه‌سازی مونت- کارلو

روش واریانس- کوواریانس، تغییرات، نوسانات و ارتباطات متغیرها را با استفاده از اطلاعات تاریخی اندازه گیری کرده و یک مدل آماری برای برآورد VaR ارائه می‌دهد. روش شبیه سازی تاریخی از اطلاعات تاریخی استفاده کرده و از آنها به طور مستقیم برای پیش بینی و شبیه سازی در آینده بهره می‌گیرد. در واقع استفاده از این روش بدین معنا است که روند گذشته در آینده تکرار خواهد شد. شبیه‌سازی مونت- کارلو نیز به روش شبیه سازی تاریخی شباهت زیادی دارد. تفاوت اصلی این دو روش در این است که بجای استفاده از داده های گذشته و فرض تکرار روند گذشته در آینده، در روش مونت- کارلو از نمونه‌های تصادفی برای برآورد ارزش در معرض خطر استفاده می‌شود.

اما در داده‌های چند متغیره، وجود همبستگی بین داده‌ها باعث می‌شود که برآورد VaR که برای هر یک از متغیرها به طور جداگانه انجام می‌گیرد کارایی لازم را نداشته باشد. به عبارت دیگر، برای اینکه برآورد دقیقی از VaR داشته باشیم، لازم است که توزیع توام داده‌ها و وابستگی بین آنها را نیز مدل کنیم. در این راستا، کاپولا ابزاری توانمند برای یافتن توزیع توام چند متغیره و استنتاج مقدار VaR بر اساس آن می‌باشد. ادامه، انواع مختلفی از کاپولا تعریف شده و ویژگی هر یک در مدل کردن همبستگی بین داده‌های غایی مورد تحلیل قرار گرفته است.

۲.۲ کاپولا

کاپولاروشی ریاضی است که با ترکیب توزیهای یک متغیره، توزیع چند متغیره داده‌ها را می‌سازد و به این ترتیب برای مدل کردن همبستگی داده‌ها مناسب است. به بیان ریاضی، یک تابع $C: [0, 1]^2 \rightarrow [0, 1]$ زمانی یک تابع کاپولا است که ویژگیهای زیر را دارد:

$$(آ) \quad C(u, 0) = C(0, v) = 0 \quad \text{برای هر } 0 \leq u, v \leq 1 \quad \text{عبارت های روبرو برقرار هستند}$$

$$(ب) \quad C(u, 1) = u, C(1, v) = v$$

$$(ج) \quad \text{همچنین اگر } u_1, v_1, u_2, v_2 \in [0, 1]; u_2 \geq u_1, v_2 \geq v_1 \text{ باشد، در اینصورت}$$

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

بر اساس تعریف‌های بالا مشخص است که دامنه تعریف هر یک از متغیرها در کاپولا در بازه $[0, 1]$ می‌باشد و چگالی حاشیه‌ای هر یک از متغیرها یکنواخت می‌باشد. به کمک این ویژگی می‌توانیم ساختار وابستگی بین چگالی‌های حاشیه‌ای یک تابع توزیع چند متغیره را تعیین کنیم. اسکالر (۱۹۵۹) نشان داد که اگر $F(X_1, X_2)$ تابع توزیع توأم دو متغیره با توابع حاشیه‌ای $F_1(X_1)$ و $F_2(X_2)$ باشد، آنگاه یک تابع کاپولای C بصورت زیر وجود دارد:

$$F(X_1, X_2) = C(F_1(X_1), F_2(X_2))$$

که با مشتق‌گیری از دو طرف معادله بالا داریم:

$$\frac{\partial^2 F(X_1, X_2)}{\partial X_1 \partial X_2} = \frac{\partial^2 C(F_1(X_1), F_2(X_2))}{\partial F_1 \partial F_2} \times f_1(X_1) f_2(X_2) = C(\tilde{u}) \times \prod_{i=1}^2 f_i(X_i)$$

که در آن f_i تابع چگالی F_i و $F_i = F_i(X_i)$ و $u_i = F_i(X_i)$ و $\tilde{u} = (u_1, u_2)$ و $C(\tilde{u})$

تابع چگالی کاپولاست. هنگامی که متغیرها پیوسته باشند، قضیه اسکالر نشان می‌دهد که هر تابع توزیع احتمال چند متغیره می‌تواند با یک تابع حاشیه‌ای و یک ساختار وابستگی به فرم معادله بالا نشان داده شود. اگر همه توزیع‌های حاشیه‌ای پیوسته باشند، آنگاه کاپولا یکتاست و می‌توان آن را به صورت یکتا با استفاده از توزیع‌های حاشیه‌ای بیان کرد. اما کاپولا انواع مختلفی دارد که هر یک از آنها دارای ویژگی‌های منحصر به فردی در مدل کردن وابستگی بین داده‌های غایبی می‌باشد. در زیر برخی از انواع مهم کاپولا به اختصار شرح داده شده است.

- کاپولای نرمال

[۱۹] تابع توزیع کاپولای نرمال را بصورت زیر بیان کرده است:

$$C^{Ga}(u_1, u_2; \rho) = \Phi_p(\psi^{-1}(u_1), \psi^{-1}(u_2)) \quad \rho \in (-1, 1)$$

که در آن ψ تابع توزیع نرمال استاندارد تک متغیره و Φ_p تابع توزیع نرمال استاندارد دو متغیره با ضریب همبستگی خطی ρ بین متغیرها است. این کاپولا توانایی مدل کردن همبستگی بین داده‌های غایبی (بزرگ و یا کوچک) را ندارد (به شکل ۳ مراجعه شود).

- کاپولای تی-استیودنت

[۱۱] تابع توزیع کاپولای تی استیودنت را بصورت زیر تعریف کرده اند:

$$C_{\nu\rho}^t(u_1, u_2) = T_{\nu\rho}(t_{\nu}^{-1}(u_1), t_{\nu}^{-1}(u_2)) \quad \rho \in (-1, 1), \nu > 0$$

که در آن T_{ν} تابع توزیع استاندارد دو متغیره تی استیودنت، ρ مقدار ضریب همبستگی خطی بین داده‌ها، ν درجه آزادی و t_{ν}^{-1} تابع توزیع معکوس در توزیع تک متغیره تی استیودنت است. [۱۱] نشان دادند

که تابع کاپولای تی استیودنت می تواند همبستگی بین مقادیرهای غایی بزرگ و همچنین همبستگی بین مقادیرهای غایی کوچک را به طور همزمان مدل کند (شکل ۳).

- کاپولای کلایتون

این تابع کاپولا دارای توزیع نامتقارن است به نحوی که در آن وابستگی در داده‌های غایی کوچک بیشتر از وابستگی در دنباله مقادیرهای غایی بزرگ می‌باشد.

$$C_{\theta}(u_1, u_2) = [Max(u_1^{-\theta} + u_2^{-\theta} - 1, 0)]^{-\frac{1}{\theta}}$$

که در آن θ پارامتر توزیع می باشد و تابع مولد آن بصورت زیر تعریف می شود.

$$\Psi(t) = (1/\theta)(t^{-\theta} - 1) \quad \theta \in [-1, \infty)$$

- کاپولای گامبل

این کاپولا مانند تابع کاپولای کلایتون دارای توزیع نامتقارن است و برخلاف کاپولای کلایتون وابستگی در مقادیرهای غایی بزرگ را به خوبی مدل می‌کند.

$$C_{\theta}(u_1, u_2) = \exp \left\{ - [(-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}]^{\frac{1}{\theta}} \right\} \quad \theta \in [1, \infty)$$

این تابع کاپولا دارای تابع مولد $\Psi(t) = (-\ln t)^{\theta}$ می‌باشد.

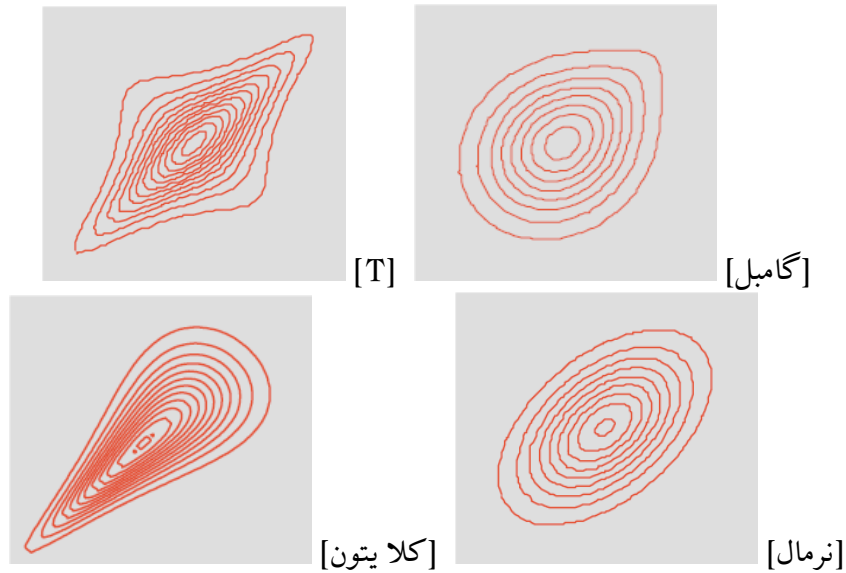
بنابراین، کاپولای تی استیودنت و نرمال هر دو متقارن هستند ولی اشکال عمده کاپولای نرمال این است که مقادیر غایی را مدل نمی‌کند و به همین جهت در دنیای واقعی کمتر از کاپولای نرمال استفاده می‌شود. کاپولای τ وابستگی بین متغیرها در مقادیرهای بزرگ و یا مقادیرهای کوچک را به خوبی مدل می‌کند. ولی ایراد عمده کاپولای تی استیودنت این است علاوه بر ضریب همبستگی خطی ρ ، یک پارامتر درجه آزادی نیز دارد که باید برآورد شود. کاپولاهای گامبل و کلایتون نیز به ترتیب وابستگی بین داده‌های غایی بزرگ و کوچک را به خوبی مدل می‌کند.

اما پارامترهای گفته شده در توابع کاپولا با معیارهای سنجش همراستایی متغیرها نظیر ضریب همراستایی τ کندال و ضریب همبستگی ρ اسپیرمن ارتباط دارند که در ادامه به آنها اشاره شده است.

τ کندال: برای هر جفت متغیر تصادفی (X, Y) ، ضریب همراستایی τ کندال به صورت زیر تعریف می‌شود:

$$\rho_{\tau}(X, Y) = P\{(X - \bar{X})(Y - \bar{Y}) > 0\} - P\{(X - \bar{X})(Y - \bar{Y}) < 0\}$$

که در آن (\bar{X}, \bar{Y}) میانگین متغیرها می باشد. همانطور که از تعریف پیداست، τ کندال احتمال همراستایی متغیرها را از احتمال غیر همراستایی آنها کم می‌کند. حال فرض کنیم (X, Y) متغیرهای تصادفی پیوسته



شکل ۳: توانایی انواع مختلفی از کاپولا در مدل کردن همبستگی بین داده‌ها نشان داده شده است. کاپولای نرمال برای مدل کردن وابستگی بین داده‌های غایبی مناسب نیست. کاپولای T وابستگی در مقدارهای بزرگ و همچنین مقدارهای کوچک را به خوبی مدل می‌کند. کاپولای گامبل و کلا یتون به ترتیب برای مدل کردن وابستگی در داده‌های بزرگ و داده‌های کوچک کاربرد دارند.

باشند، در این صورت τ کندال به صورت زیر تعریف می‌شود.

$$\rho_{\tau}(X, Y) = 4 \int \int_{[0,1]^2} C(u, v) dC(u, v) - 1$$

که در آن $u = F(X)$ و $v = G(Y)$ هستند که F و G به ترتیب تابع توزیع تجمعی متغیرهای X و Y می‌باشند. ρ اسپیرمن: این مقدار برای متغیرهای تصادفی X و Y با توزیع‌های تجمعی حاشیه‌ای F و G به صورت زیر محاسبه می‌شود: $\rho_S(X, Y) = \rho(F(X), G(Y))$ که در آن ρ نشان دهنده ضریب همبستگی خطی است. به بیان دیگر ρ_S (اسپیرمن) ضریب همبستگی خطی بین تابع توزیع متغیرهای تصادفی X و Y را اندازه می‌گیرد. و برای متغیرهای تصادفی پیوسته به صورت زیر محاسبه می‌شود:

$$\rho_S(X, Y) = 12 \int \int_{[0,1]^2} uv dC(u, v) - 3$$

در جدول ۱ رابطه τ کندال و ρ اسپیرمن با کاپولا ذکر شده‌اند.

۳ نتایج

برای اینکه ارتباط بین خسارت بیمه بدنه و شخص ثالث اتومبیل بهتر مشخص شود، لگاریتم این داده‌ها را در مبنای ۱۰ محاسبه کرده و در شکل ۴ نمودار نقطه‌ای-هیستوگرام آن نمایش داده شده است. بر اساس شکل

جدول ۱: رابطه بین ضریب همبستگی خطی ρ با ضریب همبستگی ρ اسپیرمن و ضریب همراستایی τ کندال در انواع مختلفی از کاپولا

کاپولا	ρ اسپیرمن	τ کندال
کاپولای نرمال	$\frac{\rho}{\pi} \arcsin(\rho/\sqrt{2})$	$\frac{\rho}{\pi} \arcsin \rho$
کاپولای تی استیودنت	-	$\frac{\rho}{\pi} \arcsin \rho$
کاپولای کلایتون	-	$\frac{\theta}{\theta+2}$
کاپولای گامبل	-	$1 - \frac{1}{\theta}$

۴، هیستوگرام لگاریتم مقدار خسارات در بیمه شخص ثالث و بدنه اتومبیل دارای چولگی به سمت راست (مقدارهای بزرگ) می باشد. همچنین، وجود همبستگی مثبت بین لگاریتم خسارت‌های بیمه بدنه و ثالث مشخص است. در واقع لگاریتم خسارت‌های بیمه شخص ثالث (Y) در برابر لگاریتم خسارت‌های بیمه بدنه (X) دارای معادله رگرسیونی زیر است:

$$Y = 4/3 + 0.37X$$

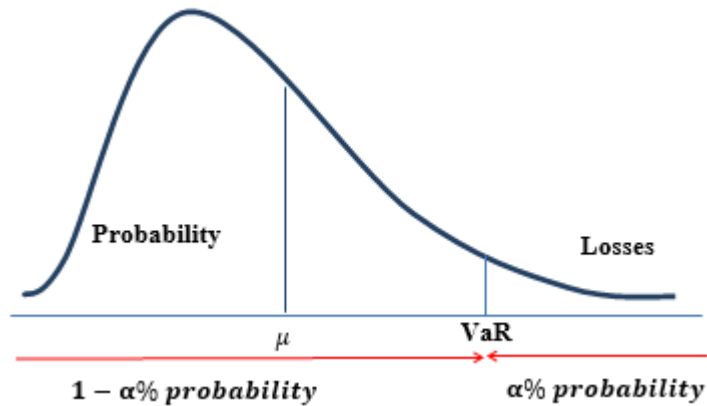
و مقدار R^2 برابر با ۰.۳۶ است. برای برازش کاپولا به لگاریتم داده‌های خسارت مراحل زیر انجام می‌شود:

- برازش یک توزیع پارامتری به داده‌های خسارت
- محاسبه مقدار تابع توزیع تجمعی برای هر یک از خسارت‌ها، اگر (X, Y) نشان دهنده یک زوج مرتب از خسارت‌های بیمه بدنه و شخص ثالث اتومبیل برای یک فرد بیمه شده باشند و F و G به ترتیب تابع توزیع تجمعی برازش شده به خسارت بیمه بدنه و شخص ثالث اتومبیل باشند، آنگاه $(u, v) = (F(X), G(Y))$ از بین توزیع‌های مختلفی که به لگاریتم داده‌های خسارت بدنه و شخص ثالث برازش شده است، توزیع مقادیر غایی گسترش یافته بر اساس معیار ماکزیمم درستنمایی بهترین برازش را داشته است. این توزیع که با $GEV(\xi, \mu, \sigma)$ نشان داده می‌شود دارای تابع چگالی احتمال زیر است:

$$f(x) = \frac{1}{\sigma} t(x)^{\xi+1} e^{-t(x)}$$

$$t(x) = \begin{cases} \left(1 + \left(\frac{x-\mu}{\sigma}\right) \xi\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ e^{-(x-\mu)/\sigma} & \text{if } \xi = 0 \end{cases}$$

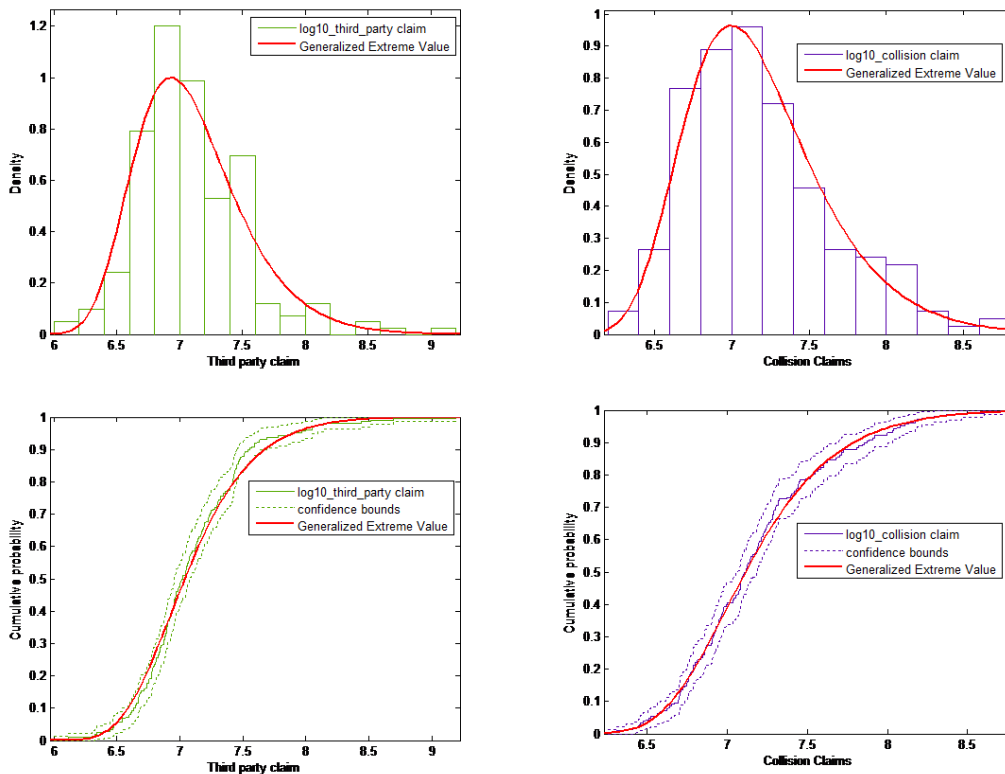
در جدول ۲ پارامترهای برآورد شده به همراه انحراف معیار برآوردها آمده است. نمودار تابع توزیع تجربی لگاریتم داده‌های خسارت به همراه تابع چگالی و تابع توزیع تجمعی GEV برازش شده به آنها در



شکل ۴: نمودار نقطه‌ای لگاریتم خسارت بیمه بدنه در برابر لگاریتم خسارت بیمه شخص ثالث رسم شده است.

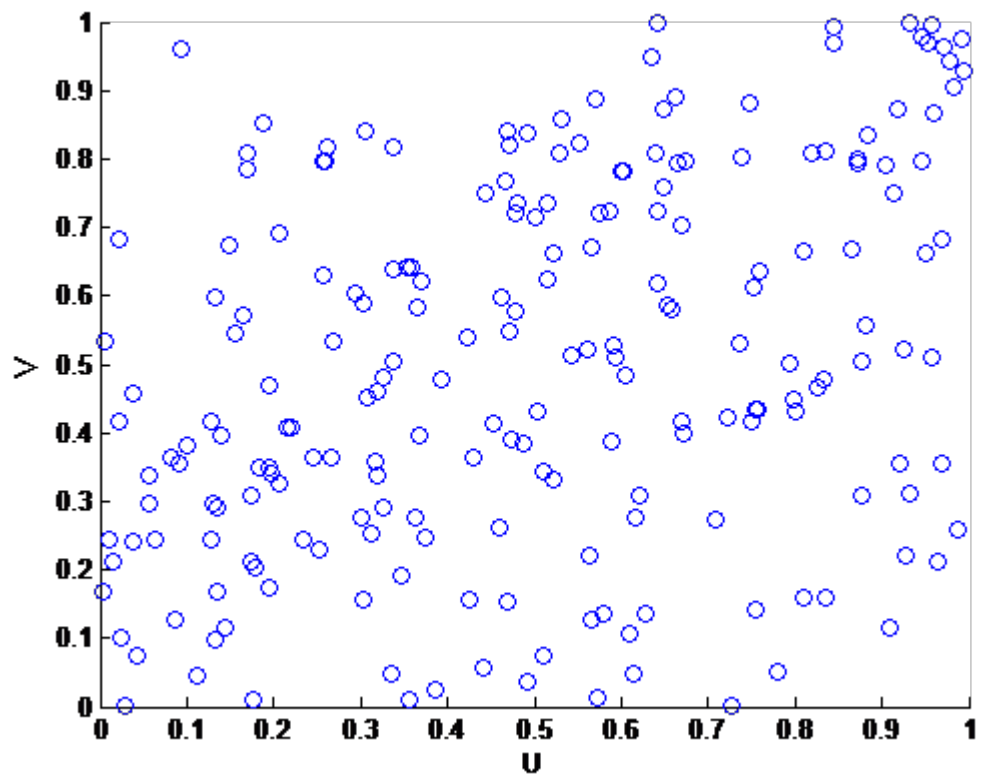
جدول ۲: توزیع‌های برازش شده به لگاریتم داده‌های خسارت به همراه پارامترهای برآورد شده

توزیع GEV برازش شده به لگاریتم داده‌های خسارت بدنه - لگاریتم درست‌نمایی ۱۲۲/۱۹۸ -		
پارامتر	برآورد	انحراف معیار برآورد
ξ	-۰/۰۵۱	۰/۰۰۴۸
σ	۰/۳۸۲۰	۰/۰۲۱
μ	۶/۹۷۶۲	۰/۰۲۹
توزیع GEV برازش شده به لگاریتم داده‌های خسارت شخص ثالث - لگاریتم درست‌نمایی ۱۰۹/۳۳۵ -		
پارامتر	برآورد	انحراف معیار برآورد
ξ	-۰/۰۷۰	۰/۰۳۱۹
σ	۰/۳۶۸	۰/۰۱۸۷
μ	۶/۹۰۵	۰/۰۲۷۷

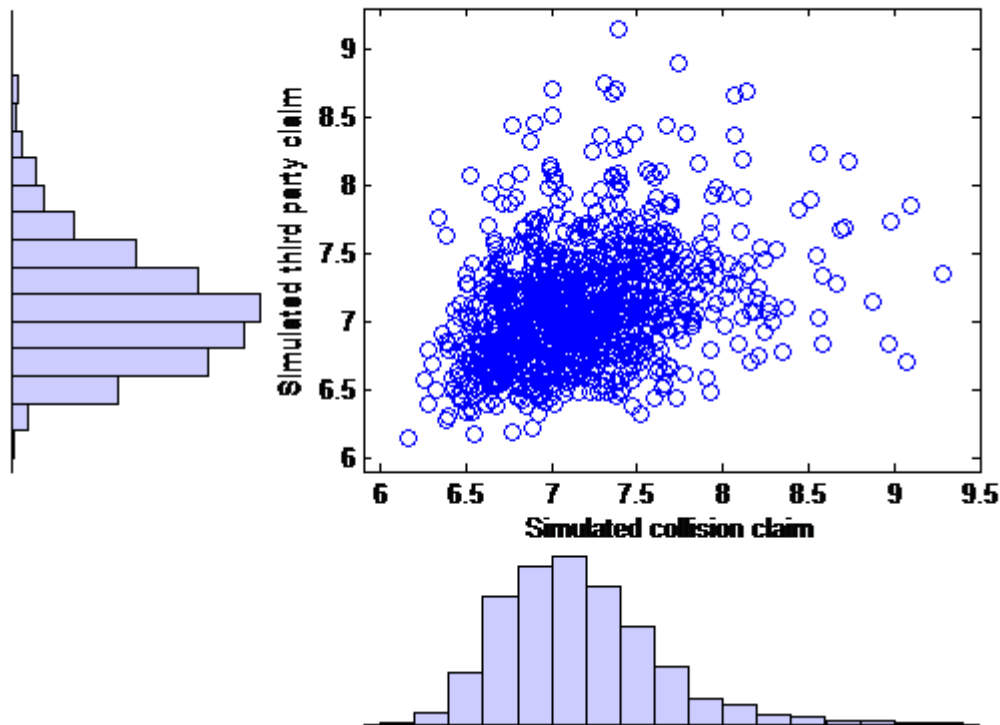


شکل ۵: تابع چگالی و تابع توزیع برازش شده به داده های خسارت. الف و ب به ترتیب تابع چگالی GEV برازش شده به لگاریتم داده های خسارت بیمه بدنه و شخص ثالث اتومبیل می باشند. نمودار های ج و د نیز تابع توزیع های GEV برازش شده را نشان می دهند.

شکل ۵ رسم شده است. از مقایسه تابع های برازش شده و تابع توزیع تجربی لگاریتم داده های خسارت مشخص است که توزیع برازش شده و توزیع تجربی لگاریتم خسارات انطباق زیادی با یکدیگر دارند. پس از تعیین توزیع لگاریتم داده های خسارت، مقدار این تابع توزیع برای هر یک از لگاریتم داده های خسارت محاسبه شده است. یعنی اگر (X, Y) به ترتیب نشان دهنده لگاریتم خسارت های بیمه بدنه و شخص ثالث باشد، آنگاه $(u, v) = (F(X), G(Y))$ است. که $F(X)$ و $G(Y)$ به ترتیب تابع توزیع تجمعی برازش شده به لگاریتم خسارت های بیمه بدنه و شخص ثالث اتومبیل می باشد. پس از تعیین زوج های مرتب (u, v) برای افراد خسارت دیده، کاپولا به آنها برازش می شود. در کاپولای T برازش شده به داده ها بر اساس معیار ماکزیمم درستمایی، پارامتر $\rho = ۰.۳۷$ و درجه آزادی آن $\nu = ۲۵/۶۲$ برآورد شده است. لازم به یادآوری است که ρ ضریب همبستگی خطی بین زوج های مرتب (u, v) است. در شکل ۶ نیز ۱۰۰۰ نمونه تصادفی شبیه سازی شده از کاپولای T برازش شده به داده ها نمایش داده شده است. پس از



شکل ۶: نمونه‌های شبیه‌سازی شده از کاپولای T



شکل ۷: نمونه‌های شبیه‌سازی شده از توزیع توام خسارت‌های بیمه بدنه و شخص ثالث

شبیه‌سازی زوج‌های مرتب (u, v) از تابع کاپولا، این زوج‌های مرتب را به کمک تابع توزیع معکوس‌های F^{-1} و G^{-1} به نمونه‌هایی از توزیع خسارت بیمه بدنه و شخص ثالث اتومبیل تبدیل می‌کنیم. یعنی:

$$x = F^{-1}(u), \quad y = G^{-1}(v)$$

و به این ترتیب یک نمونه از توزیع توام خسارت‌های بیمه بدنه و شخص ثالث شبیه‌سازی می‌شود. در شکل ۷ نمودار نقطه‌ای و هیستوگرام این داده‌ها نشان داده شده‌اند.

– محاسبه ارزش در معرض خطر (VaR)

برای برآورد در VaR یک فرد بیمه شده، تعداد m نمونه از کاپولا شبیه‌سازی شده و بر اساس هر نمونه برآوردی از مجموع خسارت بیمه بدنه و شخص ثالث اتومبیل به دست آمده است (جدول ۱۱). سپس چندانک ۹۹/۵٪ این نمونه‌ها به عنوان برآورد VaR در نظر گرفته می‌شود. با توجه به اینکه کاپولا وابستگی بین خسارات بدنه و شخص ثالث را در محاسبه VaR در نظر می‌گیرد، طبیعی است که — در مقایسه با II Solvency و آیین نامه ۶۹ — مقدار VaR را دقیق‌تر برآورد کند.

اما باید توجه داشت که در آیین نامه ۶۹ مقدار VaR روی مجموع خسارات رخ داده در یک سال برآورد می‌شود (مجموع خسارات سالیانه در فاکتور ریسک ضرب می‌شود). در نتیجه، برآورد VaR شده بر اساس آیین نامه ۶۹ به طور مستقیم با VaR برآورد شده بر اساس کاپولا — که برای یک فرد خسارت دیده است — قابل مقایسه نمی‌باشد. برای مقایسه نتایج روش‌های مبتنی بر کاپولا و آیین نامه ۶۹، متوسط خسارت بدنه و شخص ثالث اتومبیل برای افراد مختلف توسط کاپولا شبیه‌سازی شده است و شبیه‌سازی تا جایی که مجموع متوسط خسارت افراد مختلف به اندازه خسارت‌های آیین نامه ۶۹ شود ادامه پیدا کرده است (جدول ۳). یعنی

$$\text{خساراتی که در آیین نامه ۶۹ آمده و VaR برای آن برآورد شده است} = \sum_{i=1}^n \bar{z}^t$$

پس از شبیه‌سازی خسارت‌ها بر اساس الگویی که در جدول ۳ آمده، مقدارهای $\sum_{j=1}^n (x_i^j + y_i^j)$ برای $i = 1, 2, \dots, m$ نیز محاسبه شده‌اند و بر اساس این m نمونه، مقدار VaR برای مجموع خسارت‌ها بر اساس کاپولا برآورد شده است.

در شرکت بیمه سامان، مجموع خسارت پرداختی سهم نگهداری بیمه‌نامه‌های ثالث و بدنه — در سال ۹۳ — مبلغ ۳۴۷,۰۵۱,۲۳۸,۷۶۰ ریال بوده است. بر اساس آیین نامه ۶۹ نیز مقدار VaR برای این خسارت مبلغ ۴۲۹,۶۲۶,۴۹۱,۵۹۰ ریال برآورد شده است. برآورد VaR بر اساس کاپولا نیز در جدول ۴ آورده شده است. همانطور که جدول ۴ نشان می‌دهد، کاپولای T، نرمال و کلایتون برآوردهای تقریباً مشابهی را برای VaR دارند. ولی نتایج کاپولای گامبل بزرگتر از برآوردهای به دست آمده از انواع دیگر کاپولا است. البته با توجه به آزمون‌های نیکویی برازش انجام شده، کاپولای گامبل برازش مناسب‌تری را به داده‌ها داشته است و برآورد VaR به دست آمده توسط آن معتبرتر می‌باشد.

لازم به ذکر است که برازش کاپولا و برنامه نویسی‌های کامپیوتری برای محاسبه VaR توسط نرم افزار متلب انجام شده است. همچنین آزمون نیکویی برازش کاپولا توسط نرم افزار R انجام شده است.

۴ نتیجه‌گیری و پیشنهادات

باید در نظر داشت که داده‌های مورد استفاده در این تحقیق مربوط به خسارات بیمه بدنه و شخص ثالث اتومبیل — در سال ۱۳۹۳ — در شرکت بیمه سامان می‌باشد. طبیعی است که با توجه به سیاست‌های مناسب شرکت سامان در کنترل خسارات اتومبیل و به ویژه در بیمه‌های شخص ثالث، خسارت رخ داده در این بیمه نامه‌ها از نوسان کمتری برخوردار باشد. در نتیجه، ارزش در معرض خطر این بیمه نامه‌ها در شرکت بیمه سامان کمتر از ارزش در معرض خطر صنعت بیمه برآورد می‌گردد.

جدول ۳: نمونه های گرفته شده برای محاسبه VaR. از هر فرد تعداد m نمونه از خسارت های بیمه بدنه و شخص ثالث شبیه سازی شده است. i امین نمونه گرفته شده از فرد j ام با زوج مرتب (x_i^j, y_i^j) نمایش داده شده است که $1 \leq i \leq m, 1 \leq j \leq n$ می باشد و x و y به ترتیب نشان دهنده خسارت رخ داده در بیمه بدنه و شخص ثالث هستند. میانگین داده های ستون j ام با \bar{z}^j نشان داده شده است و $\sum_{j=1}^n \bar{z}^j$ برابر با خساراتی است که در آیین نامه ۶۹ آمده و VaR برای آن برآورد شده است.

افراد

		فرد ۱	فرد ۲	...	فرد $n-1$	فرد n ام
نمونه های شبیه سازی شده *۴	۱	(x_1^1, y_1^1)	(x_1^2, y_1^2)	...	(x_1^{n-1}, y_1^{n-1})	(x_1^n, y_1^n)
	۲	(x_2^1, y_2^1)	(x_2^2, y_2^2)	...	(x_2^{n-1}, y_2^{n-1})	(x_2^n, y_2^n)

	m	(x_m^1, y_m^1)	(x_m^2, y_m^2)	...	(x_m^{n-1}, y_m^{n-1})	(x_m^n, y_m^n)

$$\bar{z}^1 = \bar{x}^1 + \bar{y}^1 \quad \bar{z}^2 = \bar{x}^2 + \bar{y}^2 \quad \dots \quad \bar{z}^{n-1} = \bar{x}^{n-1} + \bar{y}^{n-1} \quad \bar{z}^n = \bar{x}^n + \bar{y}^n$$

جدول ۴: VaR برای مجموع خسارت بیمه بدنه و شخص ثالث بر اساس دو روش کاپولا و آیین نامه ۶۹ برآورد شده است. پارامترهای هر کاپولا و τ کندال برآورد شده برای آن در ستون های مربوطه آورده شده اند. مقدار p-value برای آزمون نیکویی برازش کاپولا نیز در ستون آخر آمده است.

روش	VaR	پارامتر	کندال τ برآورد	p-value
بر اساس آیین نامه ۶۹	۴۹۱,۵۹۰,۴۲۹,۶۲۶			
بر اساس کاپولای T	۴۰۴,۴۰۶,۳۷۱,۱۱۲	$\nu = 25/6, \rho = 0.37$	۰/۲۳۵۵	۰/۱۱۶۴
بر اساس کاپولای نرمال	۴۰۹,۹۹۹,۱۶۰,۲۳۷	$\rho = 0.37$	۰/۲۳۵۵	۰/۲۸۵۲
بر اساس کاپولای کلایتون	۴۰۳,۳۰۵,۸۰۵,۷۳۱	$\theta = 0.3$	۰/۱۳۰۴	۰/۰۰۱۴
بر اساس کاپولای گامبل	۴۱۹,۶۷۷,۷۶۰,۲۸۳	$\theta = 1/3$	۰/۲۳۰۷	۰/۳۶۳۱

این در حالی است که در آیین نامه ۶۹، ارزش در معرض خطر بیمه نامه‌های بدنه و شخص ثالث اتومبیل بر اساس ریسک کل صنعت بیمه برآورد می‌شود. با توجه به وجود نوسان زیاد در صنعت بیمه و سیاست‌های متفاوت در شرکت‌های بیمه مختلف، فاکتورهای ریسک در نظر گرفته شده در آیین نامه ۶۹ به گونه‌ای است که مقدار ارزش در معرض خطر در شرکت‌های پر ریسک را نیز به نحو مطلوبی برآورد می‌کند. دلایل گفته شده در بالا توجیه می‌کند که چرا در مقایسه با آیین نامه ۶۹، استفاده از کاپولا به برآوردهای کوچکتری از ارزش در معرض خطر منجر می‌شود.

در حالت کلی باید توجه داشت که در آیین نامه ۶۹ ضریب همبستگی بین ریسک‌ها در نظر گرفته نشده است. بنابراین، در صورت وجود ضریب همبستگی خطی مثبت بین خسارت در رشته‌های مختلف بیمه‌ای، انتظار داریم که ارزش در معرض خطر برآورد شده در آیین نامه ۶۹ کوچکتر از مقدار این برآورد بر اساس کاپولا باشد. همچنین، وجود همبستگی منفی بین خسارت در رشته‌های مختلف بیمه‌ای موجب می‌شود که کاپولا نسبت به آیین نامه ۶۹ برآوردهای کوچکتری از ارزش در معرض خطر را داشته باشد. البته عبارات گفته شده در صورتی برقرار است که از داده‌های یکسانی برای برآورد فاکتورهای ریسک آیین نامه ۶۹ و پارامترهای کاپولا استفاده شود.

در مجموع می‌توان گفت که کاپولا می‌تواند به عنوان ابزاری قدرتمند و به عنوان یک روش داخلی برای برآورد ارزش در معرض خطر هر شرکت مورد استفاده قرار گیرد. کاپولا با در نظر گرفتن همبستگی بین ریسک‌های مختلف، برآوردهای دقیقی از ارزش در معرض خطر منحصر به هر شرکت را ارائه می‌دهد. ارزش در معرض خطر منحصر به هر بیمه‌گر می‌تواند بیشتر یا کمتر از ارزش در معرض خطر برآورد شده بر اساس آیین نامه ۶۹ باشد که در آن فاکتور ریسک رشته‌های بیمه‌ای بر اساس نوسانات کل صنعت بیمه به دست آمده‌اند.

مراجع

- [۱] سجادی زینب، فتحی سعید، (۱۳۹۲)، تبیین فرآیند چهار گامی محاسبه ارزش در معرض خطر به عنوان معیار ی برای اندازه گیری ریسک و پیاده سازی آن در یک مدل بهینه سازی سرمایه گذاری، فصل نامه علمی پژوهشی دانش مالی تحلیل اوراق بهادار، سال ششم، شماره بیستم، صفحه ۱-۱۳.
- [۲] شهریار بهنام، (۱۳۹۳)، مبانی مدیریت ریسک و نظارت بر توانگری مالی در شرکت های بیمه، انتشارات پژوهشکده بیمه، ۴۲۵ صفحه.
- [۳] کشاورز حداد غلامرضا، حیرانی مهرداد، برآورد ارزش در معرض ریسک با وجود ساختار وابستگی بین بازدهی های مالی: رهیافت مبتنی بر توابع کاپولا، تحقیقات اقتصادی، دوره ۴۹، شماره ۴، زمستان ۱۳۹۳، صفحه ۸۶۹-۹۰۲.

- [4] Araichi S., Belkacem L., (2014), Solvency capital for insurance company: modelling dependence using copula.
- [5] Antonin G. and Benjamin S., (2001), Effets de la dépendance entre différentes branches sur le calcul des provisions, *Astin Bulletin*, 1–28.
- [6] Belguise O., Levi C., (2002), Tempêtes: Etude des dépendances entre les branches automobile et incendie laide de la thorie des copulas, *Bulletin Franais dActuariat*, 1–30.
- [7] Barges M., Cossette H., and Marceau, E., (2009), Tvar-based capital allocation with copulas, *Insurance: Mathematics and Economics*, 1–33.
- [8] Cadoux D. and Loizeau J.-M., (2004), Copules et dépendances: application pratique la dtermination du besoin en fonds propres dun assureur non vie, *Astin Bulletin*, 1–30.
- [9] Diers D., Eling M., Marek S. D., (2012), Dependence modeling in non-life insurance using the bernstein copula, *Insurance: Mathematics and Economics*, **50**(3), 430–436.
- [10] Dacorogna M. M., Canestraro D.,(2010), The influence of risk measures and tail dependencies on capital allocation, *SCOR papers*.
- [11] Embrechts P., McNeil A.J., Straumann, D., (2002), Correlation and dependence in risk management: properties and pitfalls, In: Dempster, M.A.H. (Ed.), *Risk Management: Value at Risk and Beyond*, CambridgeUniversity Press, Cambridge.
- [12] Krauth G., (2007),Provisionnement et corrlation entre branches, *Master's thesis*, Conservatoire National Des Arts Et Metiers.
- [13] Faivre F., (2002),Copula: A new vision for economic capital and application to a four line of business company, *Astin Bulletin*, 1–22.
- [14] Frees E. W., Valdez E. A.,(1998),Understanding Relationships Using Copulas, *North American Actuarial Journal*, **2**(1), 1–25.
- [15] Kaishev V. K., Dimitrova D. S.,(2006), Excess of loss reinsurance under joint survival optimality, *Insurance: Mathematics and Economics*, 1–26.
- [16] Sklar A., (1959), Fonctions de repartition a n dimensions et leurs marge, *Publications de l'Institut Statistique de l'Universite de Paris* 8, 229–231.
- [17] Shi, P. and Frees, E. W., (2011), Dependent loss reserving using copulas, *Astin Bulletin Journal of the International Actuarial Association*.

- [18] Staudt A., (2010), Tail risk, systemic risk and copulas, *Casualty actuarial society E-Forum*, **2**.
- [19] Song P.X.-K. (2000), Multivariate dispersion models generated from Gaussian copula, *Scandinavian Journal of Statistics*, **27**(2), 305–320.
- [20] Zhao, X. and Zhou, X., (2010), Applying copula models to individual claim loss reserving methods, *Insurance: Mathematics and Economics*.