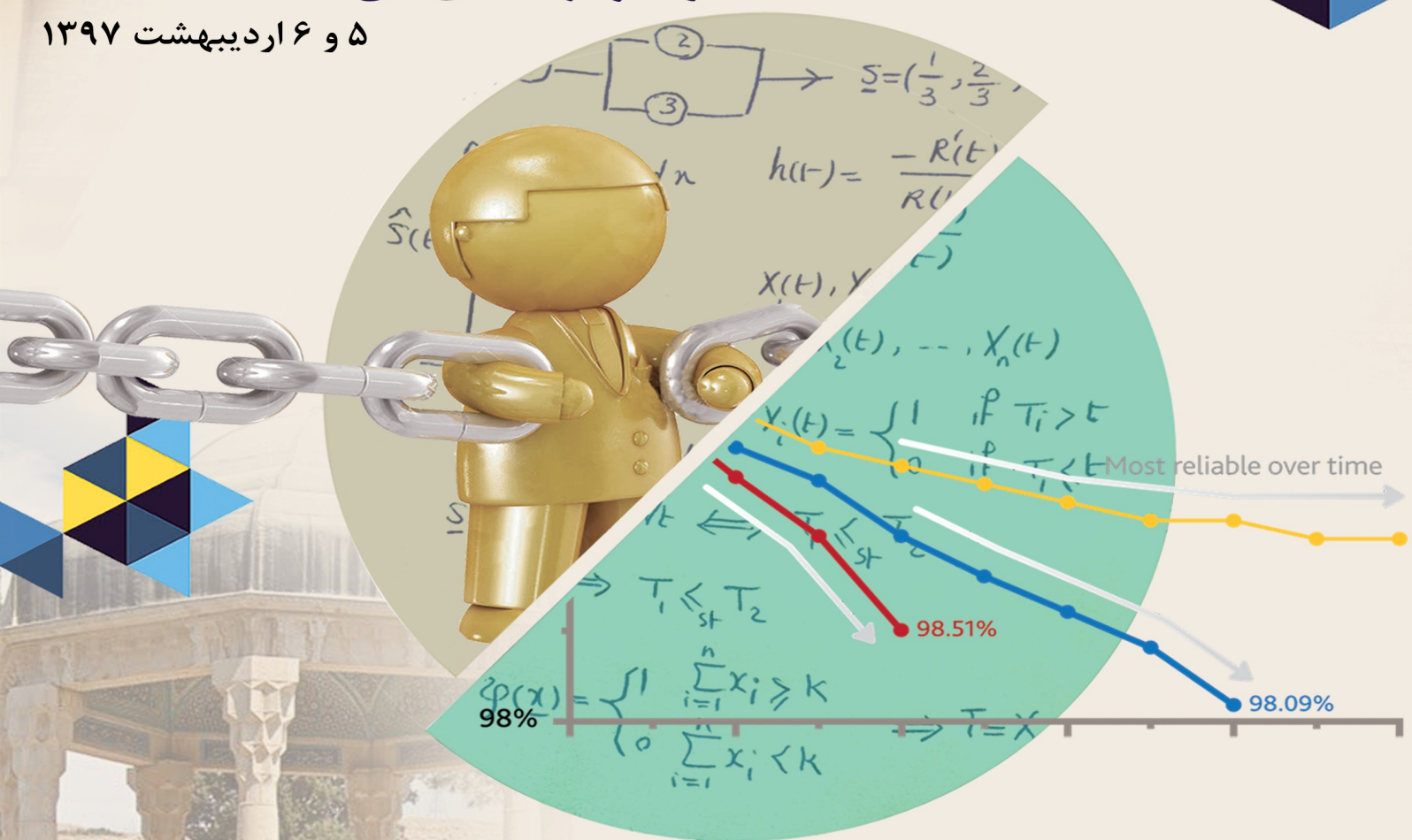


مجموعه مقالات

چهارمین سمینار تخصصی

نظریه قابلیت اعتماد و کاربردهای آن

۵ و ۶ اردیبهشت ۱۳۹۷



باسمه تعالی



مجموعه مقالات

چهارمین سمینار تخصصی

نظریه قابلیت اعتماد و کاربردهای آن

بخش آمار، دانشگاه شیراز

۵ و ۶ اردیبهشت ماه ۱۳۹۷

مقدمه

پیرو برگزاری سه دوره کارگاه آموزشی «قابلیت اعتماد» در دانشگاه‌های فردوسی مشهد (۱۳۷۸، ۱۳۹۱) و اصفهان (۱۳۹۰) و سه دوره سمینار تخصصی «نظریه قابلیت اعتماد و کاربردهای آن» در دانشگاه‌های اصفهان (۱۳۹۴)، تهران (۱۳۹۵) و مشهد (۱۳۹۶)، خداوند منان را سپاسگزاریم که توفیق برگزاری چهارمین دوره از این سمینار تخصصی را نصیب بخش آمار دانشگاه شیراز کرد. این سمینار دو روزه با حمایت قطب علمی داده‌های ترتیبی و فضایی، مرکز اطلاع رسانی و پایگاه استنادی جهان اسلام و همکاری انجمن آمار ایران، با هدف فراهم نمودن امکانات مناسب برای تبادل اطلاعات پژوهش‌های اعضای هیأت علمی، دانشجویان تحصیلات تکمیلی و کارشناسان و کاربران در محورهای سمینار، روزهای چهارشنبه و پنجشنبه ۵ و ۶ اردیبهشت ۱۳۹۷ در دانشگاه شیراز برگزار می‌شود. امیدواریم که تلاش کمیته‌های 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 ۵۰ مقاله به صورت ارائه شفاهی و ۹ مقاله به صورت ارائه پوستر مورد پذیرش قرار گرفت. لازم به ذکر است که شرکت‌کنندگان این سمینار ۹۰ عضو هیئت علمی و دانشجوی تحصیلات تکمیلی از ۲۲ دانشگاه معتبر کشور هستند. در پایان لازم می‌دانیم از کمیته‌های علمی، اجرایی و کمیته داوران سمینار نهایت تشکر و سپاسگزاری را داشته باشیم. همچنین مراتب قدردانی خود را از حوزه معاونت پژوهشی دانشگاه شیراز و دانشکده علوم، ریاست محترم دانشکده علوم، قطب علمی داده‌های ترتیبی و فضایی، پایگاه استنادی علوم جهان اسلام، رئیس محترم پارک علم و فناوری استان فارس، انجمن آمار ایران و همکاران بخش آمار دانشگاه شیراز اعلام نماییم.

از خداوند منان آرزوی توفیق تمامی شرکت‌کنندگان محترم در این سمینار را داریم. امیدواریم اقامت دلپذیری در شهر شیراز داشته باشند.

سمیه زارع‌زاده (دبیر)

اردیبهشت ۱۳۹۷

محورهای سمینار

فرآیندهای تصادفی در قابلیت اعتماد	آزمون‌های طول عمر تسریع‌یافته
قابلیت اعتماد سیستم‌های منسجم	الگوریتم‌ها و روش‌های محاسباتی در قابلیت اعتماد
قابلیت اعتماد شبکه‌ها	الگوهای تعمیر و نگهداری سیستم‌ها
قابلیت اعتماد نرم‌افزار	بهینه‌سازی در قابلیت اعتماد
مدل‌های تنش-مقاومت	تحلیل بقاء
مدل‌های فرسایشی	تحلیل داده‌های طول عمر
مطالعات موردی در قابلیت اعتماد	تحلیل ریسک در قابلیت اعتماد
مفاهیم سالخوردگی	ترتیب‌های تصادفی در قابلیت اعتماد
نظریه توزیع‌های طول عمر	داده کاوی در قابلیت اعتماد
وابستگی تصادفی در قابلیت اعتماد	روش‌های بیزی در قابلیت اعتماد

اعضای کمیته علمی (به ترتیب حروف الفبا):

- دکتر جعفر احمدی، دانشگاه فردوسی مشهد
- دکتر مجید اسدی، دانشگاه اصفهان
- دکتر اکبر اصغرزاده، دانشگاه مازندران
- دکتر جواد بهبودیان، دانشگاه شیراز
- دکتر مینا توحیدی، دانشگاه شیراز (دبیر علمی سمینار)
- دکتر فیروزه حقیقی، دانشگاه تهران
- دکتر بهاءالدین خالدی، دانشگاه رازی
- دکتر احمد خدادادی، دانشگاه شهید بهشتی

- دکتر محمود خراتی کوپایی، دانشگاه شیراز
- دکتر اسماعیل خرم، دانشگاه صنعتی امیرکبیر
- دکتر محمد خنجری صادق، دانشگاه بیرجند
- دکتر سمیه زارعزاده، دانشگاه شیراز
- دکتر علیرضا نعمت الهی، دانشگاه شیراز
- دکتر علی زینل همدانی، دانشگاه صنعتی اصفهان

اعضای کمیته علمی خارجی (به ترتیب حروف الفبا):

- Prof. Gaudoin, O., Université Grenoble Alpes (France)
- Prof. Soofi, E. S., University of Wisconsin-Milwaukee (USA)
- Prof. Ross, S., University Of Southern California (USA)

اعضای کمیته برگزار کننده (به ترتیب حروف الفبا):

۱. دکتر عبدالرسول برهانی حقیقی، دانشگاه شیراز
۲. دکتر هادی جباری نوقابی، دانشگاه فردوسی مشهد
۳. دکتر محمود خراتی کوپایی، دانشگاه شیراز
۴. دکتر سمیه زارعزاده، دانشگاه شیراز (دبیر اجرایی سمینار)
۵. سرکار خانم بهناز میرزایی، دانشگاه شیراز
۶. دکتر علیرضا نعمت الهی، دانشگاه شیراز

کادراجرایی سمینار (به ترتیب حروف الفبا):

عرفانه ایزدی (دانشجوی کارشناسی)، سمانه باقری زاده (دانشجوی کارشناسی)، کیمیا فروزانی (دانشجوی کارشناسی)، زهرا فیلی (دانشجوی کارشناسی)، رهام کمندی (فارغ التحصیل کارشناسی ارشد)، حامد نیرومندفرد (دانشجوی کارشناسی)، امیرحسین هوشمند (دانشجوی کارشناسی)

فهرست

بهینه سازی احتمال بقا در برنامه ریزی تعویض کلیه

احمدی، م.ب.، امیدی، ا.، نژادحقیقی، م.ح. ۱

طراحی بهینه ی سیستم های k از n وزنی چندوضعیتی

اورنگی، س. 11

بررسی قابلیت اعتماد یک سیستم با استفاده از مفهوم تنش مقاومت

بازیاری، ا.، کرمی، م. 24

مرور بر پارامتر تنش-مقاومت در سیستم های منسجم تحت توزیع نرمال چوله

حسنعلی پور، پ.، رزمخواه، م. 37

برآورد قابلیت اطمینان تنش-مقاومت چند مؤلفه ای در توزیع رایلی

رحیمی ده چراغی، ا.، طاهری، س. 46

برآورد قابلیت اطمینان مدت سفر در جریان شهری مبتنی بر روش تابع مفصل

علیزاده، س.، زال زاده، س. 57

مشخص سازی های توزیع هندسی از طریق باقیمانده طول عمر

گودرزی، ف. 70

کاربرد ترتیب های تصادفی در بیمه: چالش ها و نیازها

محمودوند، ر. 82

بررسی تأثیر ساختار وابستگی و توزیع‌های حاشیه‌ای در محاسبه قابلیت اعتماد مدل‌های تنش-مقاومت

91 محمودی‌فرد، ف.، دست‌برآورده، ع.، میرحسینی، س. م.

مقایسه دو برآوردگر ناپارامتری قابلیت اعتماد

104 مهرآور، ف.، مهدی‌زاده، م.

مدل رگرسیون نمایی قطعه‌ای در تحلیل داده‌های بقاء بیماران مبتلا به انفارکتوس حاد و مقایسه این مدل با مدل رگرسیون کاکسی

115 هادی‌پور، م.، منوچهری، ط.، بانشی، م. ر.



بهینه سازی احتمال بقا در برنامه ریزی تعویض کلیه

احمدی، م.ب.^۱ امید، ا.^۲ نژادحقیقی، م.ح.^۳

^{۱،۲} بخش ریاضی، دانشکده علوم پایه، دانشگاه شیراز

^۳ بخش آمار، دانشکده علوم پایه، دانشگاه شیراز

چکیده

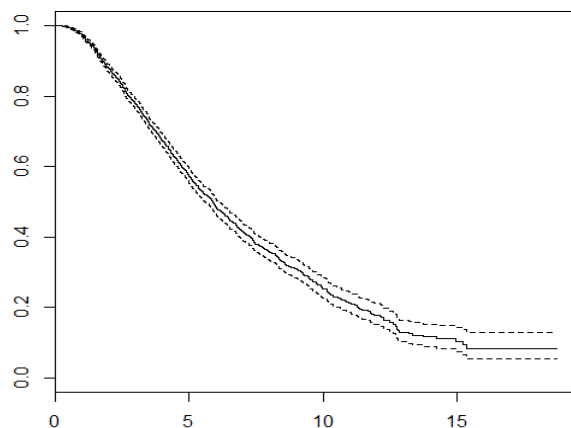
پیوند کلیه غالباً از اهدا کننده فوت شده (مرگ مغزی) به فرد نیازمند صورت می‌گیرد. بدلیل افزایش قابل ملاحظه لیست انتظار، تمامی بیماران کلیوی نیازمند به پیوند، نمی‌توانند از دریافت کلیه بهره‌مند گردند. از این رو می‌توان با استفاده از اهدا کننده زنده و تعویض کلیه مابین جفت‌های دهنده-گیرنده‌ی ناسازگار افراد بیشتری را از خطر مرگ نجات داد. در این مقاله قصد داریم با شبیه‌سازی استخری از جفت‌ها و محاسبه احتمال بقای پس از پیوند آن‌ها، یک روش امتیازدهی مناسب به هر عمل پیوند عرضه کنیم؛ سپس با مدل‌سازی ریاضی، سیاستی بهینه برای انتخاب تعویض‌های شدنی ارائه دهیم.

کلمات کلیدی: احتمال بقا، برنامه‌ریزی تعویض کلیه، مدل دوسطحی، بهینه‌سازی در سلامت.

^۱mbahmadi@shirazu.ac.ir

^۲atomidi@shirazu.ac.ir

^۳nhaghighi.mh@gmail.com



شکل ۱: تعویض کلیه با مشارکت دو جفت

۱ مقدمه

در اغلب موارد مناسب ترین درمان برای بیماران نارسایی کلیوی، پیوند کلیه است. با توجه به رشد روز افزون بیماران در لیست انتظار پیوند کلیه و کمبود کلیه، بسیاری از این بیماران جان خود را از دست می‌دهند. یکی از بهترین راه حل‌ها برای برطرف ساختن شکاف عرضه و تقاضا مابین کلیه‌های موجود و بیماران نیازمند، استفاده از دهنده‌ی زنده می‌باشد؛ اما متأسفانه اکثر بیمارانی که از اهداکننده‌ی زنده‌ی داوطلب (معمولاً خویشاوند) برخوردارند به دلیل ناسازگاری فیزیولوژیکی با فرد داوطلب خود، قادر به انجام عمل پیوند نیستند. این مشکل را می‌توان تا حدودی با تعویض کلیه بر طرف ساخت. بدین گونه که هر جفت دهنده‌گیرنده‌ی (PDP) ناسازگار، با تعویض کلیه‌ی اهداکننده‌ی خود با جفتی دیگر، عمل پیوند را با فردی که از نظر فیزیولوژیکی با آن سازگار است، انجام می‌دهد. شکل ۲ ساده ترین حالت را با تنها دو جفت گیرنده-دهنده نمایش می‌دهد [۱]. راپاپورت در سال ۱۹۸۶ استخری از جفت‌های ناسازگار را به منظور تعویض کلیه مابین آنها ایجاد کرد [۲]. چالش اصلی که در تعویض کلیه با آن روبه‌رو هستیم، انتخاب سیاستی عادلانه برای تخصیص کلیه است به طوریکه کارآمدی پیوند را نیز دربرداشته باشد. راث و همکاران [۴] و جنتری و همکاران [۳] از الگوریتم ماکزیمم تطابق ادموند برای انتخاب تعویض‌های دوتایی استفاده کردند که تنها ماکزیمم تعداد پیوندهای شدنی را به‌همراه داشت. در این‌جا تعویض کلیه مابین جفت‌های دهنده-گیرنده‌ی ناسازگار یا سازگار به منظور دریافت کلیه‌ای سازگارتر انجام می‌گیرد. در بخش ۳ استخری از ۱۰ هزار جفت با مشخصات متفاوت، شبیه سازی شده است؛ و احتمال بقای پیوند هر بیمار را در بخش ۴ برآورد می‌کنیم؛ و کارآمدی هر پیوند را میزان بقا پس از پیوند، قرار می‌دهیم. سپس با استفاده از بهینه سازی، مؤثرترین پیوندها را برای انجام عمل کلیه انتخاب می‌کنیم. در بخش ۵ مسئله تعویض کلیه، به صورت یک برنامه‌ریزی بهینه‌سازی دو سطحی مدلسازی شده است. در سطح پیشرو هدف بهینه سازی احتمال بقای پیوند است و

در سطح پسرو هدف ماکزیمم کردن مجموع پیوندها با دهنده- گیرنده‌ی دارای گروه خونی یکسان بعلاوه‌ی مجموع زمان‌های سپری شده هر بیمار تا دریافت پیوند است؛ به عبارت دیگر با استفاده از بهینه‌سازی، کارآمدی پیوند در سطح پیشرو و عدالت را در سطح پسرو برقرار می‌کنیم.

۲ پیش زمینه

بیماری نارسایی کلیوی اگر با انجام دیالیز و یا پیوند کلیه درمان نشود می‌تواند مرگ‌بار باشد. کیفیت زندگی در بیماران دریافت کننده پیوند نسبت به بیماران دیالیزی در حد مطلوب تری است. لذا درمان پیشنهادی به چنین بیمارانی، پیوند کلیه است. دو عامل اساسی که نقش کلیدی در شدنی بودن و موفقیت پیوند ایفا می‌کنند به ترتیب تطابق گروه خونی و تطابق آنتی ژن‌های HLA هستند [۴]؛ و نیز با توجه به آنچه در [۵] آمده است سن، جنسیت، میزان حساسیت بیمار (PRA) و غیره در موفقیت پیوند تأثیر دارند. چهار نوع گروه خونی A، B، AB و O وجود دارند. دهنده با گروه خونی O می‌تواند به هر چهار نوع کلیه اهدا کند. A و B به هم نوع خود یا AB کلیه می‌دهد و AB تنها می‌تواند با AB پیوند انجام دهد. (جدول ۱ را ببینید).

HLA ترکیبی از شش پروتئین است. هر چه تفاوت HLA فرد دهنده با گیرنده زیادتر باشد احتمال موفقیت پیوند کاهش می‌یابد.

میزان حساسیت یک بیمار در ایمن شناسی با آزمایش PRA سنجیده می‌شود بدین صورت که PRA درصدی از %۰ تا %۹۹ را به خود اختصاص می‌دهد و هر چه به %۹۹ درصد نزدیکتر باشد فرد حساسیت بیشتری دارد و احتمال بقای پیوند آن فرد کم‌تر است.

۳ شبیه سازی

پس از تلاش برای دریافت داده‌های واقعی از برخی بیمارستان‌هایی که پیوند کلیه در ایران انجام می‌دهند و بی نتیجه ماندن تلاش‌هایمان برای بدست آوردن داده‌های واقعی، مجبور به انجام شبیه‌سازی جفت پیوندهای دهنده- گیرنده کلیه شدیم، تا بتوانیم مطالعه‌ای روی سیستم برنامه ریزی جורسازی پیوندها انجام دهیم.

جدول ۱: جدول فراوانی گروه خونی دریافت کننده و اهداکننده

گروه خونی اهداکننده				
AB	B	A	O	گروه خونی دریافت کننده
۲,۹۹	۹,۳۲	۱۱,۶۲	۱۴,۰۹	O
۲,۲۹	۷,۲۷	۸,۹۹	۱۱,۱۳	A
۱,۸۵	۶,۱۲	۷,۲۷	۹,۲۷	B
۰,۵۰	۱,۹۳	۲,۲۰	۳,۱۶	AB

در این مطالعه تعداد ۱۰ هزار جفت گیرنده-دهنده با مشخصات گروه خونی، سن، جنسیت، اختلاف HLA، با در نظر داشتن سابقه پیوند برای گیرنده‌ی کلیه و دیابتی بودن گیرنده‌ی کلیه و درصد حساسیت (PRA)، شبیه سازی شده است.

زمان پس زدن پیوند (بر حسب سال) را یک رویداد در نظر می‌گیریم و زمان آن را در دوره زمانی مورد مطالعه بر اساس مشخصات پزشکی‌شان و با در نظر گرفتن سانسورشدگی با بهره‌گیری از روش‌های آنالیز بقا و توزیع های احتمالی و براساس منابع موجود [۳، ۵، ۸] در این زمینه شبیه سازی کردیم. زمانی فرد مورد مطالعه ما سانسور می‌شود که یکی از حالات زیر رخ دهد:

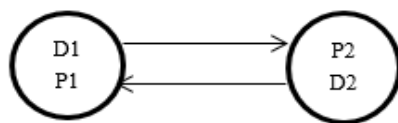
(الف) مرگ بیمار به هر دلیل دیگری که با پیوند کلیه مرتبط نباشد.

(ب) از دسترس خارج شدن بیمار برای مثال انتقال به بیمارستانی دیگر

(ج) تمام شدن دوره زمانی که فرد مورد مطالعه قرار گرفته است.

ناسازگاری های پیوند تاثیر بسزایی در کارایی پیوند دارد. کارایی پیوند با احتمال بقای بیشتر یا نرخ خطر کمتر تناسب دارد. این ناسازگاری ها می‌توانند در مرحله اول گروه خونی گیرنده و دهنده باشد و همچنین دیابتی بودن و پس از آن میزان حساسیت (PRA) و تعداد میس‌مچ ها یا اختلاف HLA آنها و همچنین اختلاف سنی و در آخر، جنسیت باشند.

تمامی این فاکتورها به میزانی روی کارایی و کیفیت پیوند تاثیرگذار هستند. به این صورت که اگر گروه خونی ها ناسازگار باشند، پیوند در همان لحظه‌های نخست پس زده می‌شود. پس زمان رخداد رویداد در پیوند



شکل ۲: نمودار تابع بقای پیوندهای شدنی

در ناسازگاری‌های گروه خونی صفر است. که در دسته‌ی نشدنی بودن قرار می‌گیرد. همچنین اگر میزان حساسیت بیشتر یا مساوی ۸۰ درصد و همزمان دارای اختلاف HLA بزرگتر یا مساوی ۲ باشد نیز، پیوند نشدنی تلقی خواهد شد.

به علاوه، هرچه مدت زمان دیالیز شدن فرد گیرنده و اختلاف سنی بین فرد گیرنده و دهنده بیشتر باشد کیفیت پیوند کاهش می‌یابد به جز در مواردی که سن دهنده عضو کمتر از سن گیرنده باشد. همچنین طبق مطالعات قبلی نشان داده شده است که پیوند در جنسیت گیرنده و دهنده‌ی مرد به مرد بهتر بوده است. جدول فراوانی داده‌ها در جدول ۱ آمده است:

۴ آنالیز بقای داده‌های شبیه‌سازی شده

در این بخش با بکارگیری روش کاپلان-مه-یر، نمودار بقای جفت‌های شبیه‌سازی شده در بخش قبل را محاسبه می‌کنیم. این نمودار با در نظر نگرفتن پیوندهای نشدنی، در ذیل آمده است. توجه شود که این مدل دارای نقص‌هایی است چرا که سیستم بقای چنین مدلی شیب زیادی دارد. علاوه بر این، جفت‌های ناسازگار باید تا زمان یافتن دهنده‌ی مناسب دیالیز شوند. لذا با بکارگیری بهینه‌سازی، تصمیم گرفته می‌شود که کدام گیرنده از کدام اهداکننده، کلیه دریافت نماید به طوری که بقای سیستم (با شیب کمتر) حداکثر شود. لذا در ادامه مدلی به این منظور ارائه می‌دهیم.

برای ارائه مدل نیاز داریم که بدانیم میزان بقای هر جفت با جفت‌های دیگر چه میزان است. برای این کار از ۱۰ هزار نمونه قبلی به دلیل محدودیت در سرعت نرم‌افزار R هزار نمونه هزارتایی را همراه با مشخصات متناظر با آن‌ها (مانند جنسیت، گروه خونی، سن، سابقه بیماری و غیره)، نمونه‌گیری می‌کنیم. این نمونه در یک ماتریس هزار در هزار نمایش داده می‌شود. سپس بقا را برای هر ستون را محاسبه می‌کنیم اما به دلیل بزرگ شدن ماتریس مذکور تنها اطلاعات مربوط به یک ستون آن را در ذیل می‌آوریم. برای هر حالت احتمال بقا را با روش کاپلان-مه-یر محاسبه می‌کنیم.

جدول ۲: جدول فراوانی

فاکتورها	سطوح فاکتورها	درصد فراوانی در دوره زمانی رویداد کمتر از ۴ سال	درصد فراوانی در دوره زمانی رویداد بیشتر از ۴ سال	درصد فراوانی کل
سن اهدانکنده	کمتر از ۱۸	—	—	—
	۱۸ الی ۲۹	۴,۰۹	۰,۹۶	۵,۰۵
	۲۹ الی ۳۹	۲۱,۳۷	۴,۳۶	۲۵,۷۳
	۳۹ الی ۴۹	۳۲,۹۱	۵,۰۵	۳۷,۹۶
	۴۹ الی ۵۹	۲۱,۳۰	۲,۸۱	۲۴,۱۱
سن دریافت کننده	۵۹ به بالا	۶,۵۱	۰,۶۴	۷,۱۵
	کمتر از ۱۳	۱,۰۰	۰,۰۸	۱,۰۸
	۱۳ الی ۱۸	۰,۹۱	۰,۰۶	۰,۹۷
	۱۸ الی ۲۹	۷,۶۲	۰,۷۴	۸,۳۶
	۲۹ الی ۳۹	۱۵,۷۰	۱,۴۹	۱۷,۱۹
جنسیت	۳۹ الی ۴۹	۲۲,۷۵	۲,۳۴	۲۵,۰۹
	۴۹ الی ۵۹	۲۰,۵۹	۳,۶۹	۲۴,۲۸
	۵۹ به بالا	۱۷,۶۱	۵,۴۲	۲۳,۰۳
	مرد به مرد	۱۴,۲۷	۳,۷۳	۱۸,۰۰
	زن به مرد	۴۰,۶۹	۵,۶۶	۴۶,۳۵
دیابتی بودن دریافت کننده	مرد به زن	۱۵,۶۳	۲,۲۲	۱۷,۸۵
	زن به زن	۱۵,۵۹	۲,۲۱	۱۷,۸۰
	غیر دیابتی	۶۵,۷۹	۱۲,۲۶	۷۸,۰۵
	دیابتی	۲۰,۳۹	۱,۵۶	۲۱,۹۵
	ندارد	۷۴,۹۲	۱۳,۴۰	۸۸,۳۲
اختلاف HLA	دارد	۱۱,۲۶	۰,۴۲	۱۱,۶۸
	۰	۰,۱۲	۰,۰۲	۰,۱۴
	۱-۲	۴,۴۴	۱,۳۱	۵,۷۵
	۳-۴	۳۷,۶۴	۵,۸۹	۴۳,۵۳
	۵-۶	۴۳,۹۸	۶,۶	۵۰,۵۸
حساسیت	کمتر از ۰,۱	۷۳,۸۶	۱۱,۶۲	۸۵,۴۸
	۰,۱ الی ۰,۷۹	۱۲,۳۰	۲,۲۰	۱۴,۵۰
	بیشتر از ۰,۷۹	۰,۰۲	۰,۰۰	۰,۰۲
	سانسور شده	۵۲,۴۶	۸,۲۰	۶۰,۶۶
	رویداد	۳۳,۷۲	۵,۶۲	۳۹,۳۴

جدول ۳: جدول طول عمر کاپلان مدیر

LCL	UCL	$s.e.$	w_{ij}	f_{ij}	n_{ij}	t_{ij}
۰,۷۷۶	۰,۷۲۲۶	۰,۱۳۷	۰,۷۴۹۰	۲۵۱	۱۰۰۰	۰,۰۰۰
۰,۷۷۶	۰,۷۲۲۶	۰,۱۳۷	۰,۷۴۹۰	۰	۳۹۵	۰,۱۴۲
۰,۷۷۵	۰,۷۲۰۵	۰,۱۳۸	۰,۷۴۷۱	۱	۳۹۴	۰,۲۵۹
۰,۷۷۳	۰,۷۱۸۴	۰,۱۳۹	۰,۷۴۵۲	۱	۳۹۳	۰,۲۶۷
۰,۷۷۱	۰,۷۱۶۳	۰,۱۴۰	۰,۷۴۳۳	۱	۳۹۲	۰,۲۷۶
⋮	⋮	⋮	⋮	⋮	⋮	⋮
۰,۱۵۲	۰,۰۱۳۴	۰,۰۲۷۹	۰,۰۴۵۱	۱	۳	۱۵,۲۵۱
۰,۱۴۲	۰,۰۳۵۷	۰,۰۲۱۲	۰,۰۲۲۶	۱	۲	۱۵,۳۸۳
۰,۱۴۲	۰,۰۳۵۷	۰,۰۲۱۲	۰,۲۲۶	۰	۱	۱۸,۶۵۸

برای $j = ۱$ جدول طول عمر کاپلان مدیر جدول ۳ است:

در بخش بعد مولفه های احتمال بقای حالات مختلف را با w_{ij} نشان می دهیم.

۵ ارائه یک مدل ریاضی برای بهینه سازی احتمال پیوند در مسئله تعویض کلیه

مدل تعویض کلیه را می توان با استفاده از گراف جهت دار نمایش داد. گراف $G(V, E)$ را که رئوس آن V مجموعه جفت های دهنده-گیرنده ی ناسازگار می باشد، در نظر بگیرید. دو رأس i و j به وسیله یک یال زمانی به هم متصل می شوند که دهنده جفت i ام با بیمار جفت j ام سازگار باشد. به هر یال (i, j) وزن w_{ij} را نسبت می دهیم که نشان دهنده ی میزان موفقیت پیوند کلیه ی دهنده ی i ام با بیمار جفت j ام می باشد. در بخش ۴ اطلاعات لازم برای محاسبه w_{ij} برای هر یال (i, j) آورده شده است. به گراف $G(V, E)$ گراف تعویض کلیه گفته می شود.

تعریف ۱. در گراف $G(V, E)$ یک دور به طول k دنباله ای به شکل $v_1 e_1 v_2, v_2 e_2 v_3, \dots, v_{k-1} e_{k-1} v_k$ است که $v_k = v_1$ و به جز این دو رأس در دنباله، همه رئوس و همه یال ها غیر تکراری می باشند.

تعریف ۲. به مجموعه‌ای از دوره‌های رأس مجزا در گراف تعویض کلیه $G(V, E)$ یک تعویض گفته می‌شود.

تعریف ۳. برای گراف تعویض کلیه $G(V, E)$ و $k \in N$ ، Ck مجموعه تمامی تعویض‌های گراف $G(V, E)$ است که دارای حداکثر طول k هستند. قرار دهید:

الف) y_{ij} برابر ۱ اگر یال ij برای انجام پیوند انتخاب شود و صفر در غیر این صورت؛

ب) x_{ij} برابر ۱ اگر گروه خونی دهنده‌ی جفت i ام و گیرنده‌ی جفت j ام یکسان باشد و صفر در غیر این صورت؛

ج) t_i مدت زمانی است که بیمار i ام تا دریافت کلیه سپری می‌کند؛

د) b_i و b'_i به ترتیب گروه خونی گیرنده و دهنده‌ی i ام؛

ه) M عددی بسیار بزرگ؛

مدل دوسطحی برای استخری از جفت‌های سازگار یا ناسازگار:

$$\text{Max} \sum_{(i,j) \in E} w_{ij} y_{ij} \quad (۱)$$

$$\text{s.t.} \sum_{j:(i,j) \in E} y_{ji} \leq 1 \quad \forall i \in V \quad (۲)$$

$$\sum_{j:(i,j) \in E} y_{ij} = \sum_{j:(i,j) \in E} y_{ji} \quad \forall i \in V \quad (۳)$$

$$x_{i_1 i_2} + x_{i_2 i_3} + \dots + x_{i_k i_{k+1}} \leq k - 1 \quad \forall \{i_1, i_2, \dots, i_{k+1}\} \subseteq V \quad (۴)$$

$$\text{max} \sum_{(i,j) \in E} x_{ij} y_{ij} + \sum_{i \in V} \sum_{j \in V} t_j y_{ij} \quad (۵)$$

$$\text{s.t.} b_i - b'_j \leq M(1 - x_{ij}) \quad (۶)$$

$$b'_j - b_i \leq M(1 - x_{ij}) \quad \forall i, j \in V \quad (۷)$$

$$x_{ij}, y_{ij} \in \{0, 1\}, \forall i, j \in V \quad (۸)$$

قید ۲ تضمین می‌کند که حداکثر یک یال به رأس i ام وارد شود. قید ۳ تعادل جریان در رأس i را بیان می‌کند؛ بدین گونه که اگر گیرنده‌ی i ام کلیه دریافت کند دهنده‌ی آن موظف است کلیه‌ای اهدا کند. قید ۴ طول هر تعویض را به حداکثر k (به دلیل ضرورت انجام عمل‌های پیوند به صورت همزمان و کمبود امکانات بیمارستان مورد نظر) محدود می‌کند.

قیود ۶ و ۷ تعداد پیوندها با گروه خونی یکسان را می‌شمارد. تابع هدف پیشرو، تعویض‌ها را به‌گونه‌ای انتخاب می‌کند که مجموع کارآمدی پیوندها بیشینه گردد؛ و تابع هدف پسرو دو هدف را دنبال می‌کند. ۱- ماکزیمم کردن تعداد پیوندها با گروه خونی یکسان ۲- ماکزیمم مجموع زمان‌های سپری شده تا دریافت کلیه؛ دهنده‌ی دارای گروه خونی O قادر است به دیگر گروه‌ها کلیه دهد این در حالی است که گیرنده‌ی نوع O تنها می‌تواند از O کلیه بگیرد. پس با ماکزیمم کردن مجموع پیوندهایی که گروه خونی دهنده-گیرنده مشابه دارند و مدت زمان بیشتری در انتظار بوده‌اند، می‌توان انتخابی منصفانه‌تر داشت.

این مدل را می‌توان با استفاده از الگوریتم ارائه شده در مقاله [۹] حل کرد.

۶ نتیجه‌گیری

با بکارگیری احتمال بقا در مسئله‌ی تعویض کلیه، که جواب حاصله را به واقعیت نزدیکتر می‌کند، مدلی ریاضی ارائه داده ایم. این مدل از جهت اینکه کیفیت پیوند (احتمال بقا) را بهینه می‌کند و همچنین عدالت را در قبال بیماران منتظر برقرار می‌سازد چرا که با انتخاب پیوند هایی که زمان بیشتری را برای دریافت در انتظار بوده اند و آنهایی که گروه خونی یکسانی دارند را انتخاب می‌کند، از این رو نسبت به مدل های موجود [۶] اثربخش تر است.

مراجع

- [1] Klimentova, X., Pedroso, J.P., Viana, A.(2016), *Maximising expectation of the number of transplants in kidney exchange programmes*. Computers and Operations Research, 73 1-11.
- [2] Rapaport, Felix T.(1986), *The Case for a Living Emotionally Related International Kidney Donor Exchange Registry*. Transplantation Proceedings, XVIII 5-9.
- [3] Segev D, Gentry S, Warren D, Reeb B, Montgomery R.(2005), *Kidney paired donation and optimizing the use of live donor organs*. J Am Med Assoc; 293 (15) 1883-90.
- [4] Roth A, Sönmez T, Ünver M.(2004), *Kidney exchange*. Q J Econ 119(2) 457-88.

- [5] Valarie B. Ashby, Alan B. Leichtman, Michael A. Rees, Peter X.-K. Song, Mathieu Bray, Wen Wang, and John D. Kalbfleisch.(2017), *A Kidney Graft Survival Calculator that Accounts for Mismatches in Age, Sex, HLA, and Body Size*. Clinical Journal of the American Society of Nephrology 12(7) 1148-1160.
- [6] Vicky Mak-Hau,(2017),*On the kidney exchange problem: cardinality constrained cycle and chain problems on directed graphs: a survey of integer programming approaches*. Journal of Combinatorial Optimization 33(1) 35-59
- [7] Edmonds J.(1965),*Paths, trees, and flowers*. CanJMath. 17:449.
- [8] Wolfe RA, McCullough KP, Schaubel DE, Kalbfleisch JD, Murray S, Stegall MD, Leichtman AB.(2008),*Calculating life years from transplant (lyft): Methods for kidney and kidney-pancreas candidates*. American Journal of Transplantation. 8(4p2):997–1011. [PubMed: 18336702]
- [9] Jonathan F. Bard, James T. Moore (1992),*An algorithm for the discrete bilevel programming problem* Naval Research Logistics (NRL)(39) 419-435

طراحی بهینه‌ی سیستم‌های k از n وزنی چندوضعیتی

اورنگی، س^۱

گروه آمار، دانشکده ریاضی و آمار، دانشگاه اصفهان

چکیده

در سال‌های اخیر، سیستم‌های k از n وزنی چندوضعیتی که حالت تعمیم‌یافته‌ی سیستم‌های k از n قدیمی هستند، به دلیل کاربرد فراوان در صنعت، مخابرات، حمل و نقل و ترافیک، مدارهای الکتریکی و ... بسیار مورد توجه پژوهشگران قرار گرفته‌اند. این سیستم‌ها اغلب با ساختار پیچیده، دوره‌ی کاری طولانی‌مدت، قابلیت اعتماد و طراحی بالا و هزینه‌های زیاد تعمیر و نگهداری همراه هستند. در همین راستا، تضمین کیفیت و بهینه‌سازی چنین سیستم‌هایی بسیار حائز اهمیت است. در این مقاله ابتدا رابطه‌ی بین قابلیت اعتماد و هزینه‌ی مؤلفه را از سیستم دودویی به سیستم k از n وزنی چندوضعیتی بسط می‌دهیم و سپس رابطه‌ی بین وزن و هزینه‌ی مؤلفه را براساس تحلیل مقادیر تجربی ارتقا می‌دهیم. پس از آن، توابع این روابط را با هم ترکیب می‌کنیم تا هزینه‌ی مؤلفه را به‌عنوان تابعی از قابلیت اعتماد مؤلفه و وزن آن نشان دهیم. سرانجام، از تابع به‌دست‌آمده استفاده می‌کنیم تا طراحی بهینه‌ی سیستم‌های k از n وزنی چندوضعیتی را ارائه دهیم.

کلمات کلیدی: سیستم چندوضعیتی، سیستم k از n وزنی، طراحی بهینه، الگوریتم ژنتیک.

^۱sz.orangi@sci.ui.ac.ir

۱ پیش‌گفتار

در مهندسی قابلیت اعتماد، یکی از مسائل بهینه‌سازی که پژوهشگران به‌طور گسترده‌ای به آن پرداخته‌اند، مسأله‌ی انتخاب مؤلفه به گونه‌ای است که ویژگی‌های هزینه و قابلیت اعتماد سیستم در نظر گرفته شده باشد. بنابراین هدف اکثر مسائل بهینه‌سازی، مینیم کردن کل هزینه‌ی مورد انتظار با توجه به نیازمندی‌های قابلیت اعتماد سیستم و یا ماکسیم کردن قابلیت اعتماد با توجه به محدودیت هزینه‌ی کل سیستم است. برای حل چنین مسائلی، باید قابلیت اعتماد سیستم را به‌طور مکرر ارزیابی کنیم و پس از مقایسه‌ی نتایج و انتخاب کاراترین روش محاسبه‌ی قابلیت اعتماد (تابع مولد عمومی یا الگوریتم‌های بازگشتی) به مدل‌سازی بهینه بپردازیم.

در این زمینه، لی و ژو [۵] مدل‌های بهینه برای قابلیت اعتماد را با توجه به محدودیت‌های هزینه‌ی سیستم ارائه داده‌اند. فقیه روحی و همکاران [۲] مدل قابلیت دسترسی پویا را توسعه و دو مدل بهینه براساس قابلیت دسترسی و محدودیت‌های هزینه‌ی طراحی به کمک الگوریتم ژنتیک ارائه داده‌اند و اریلماز [۱] الگوریتم قابلیت اعتماد پویا را برای سیستم k از n وزنی چندوضعیتی تعمیرناپذیر مورد مطالعه قرار داده‌است.

در بخش دوم، تعاریف اولیه‌ی موردنیاز برای این سیستم‌ها، مدل سیستم k از n وزنی دودویی و سپس مدل سیستم k از n وزنی چندوضعیتی ارائه شده‌است. در بخش سوم مدل‌های طراحی این سیستم‌ها و توابع هزینه و قابلیت اعتماد مربوطه معرفی شده‌اند. یک مثال کاربردی برای روشن‌تر شدن هدف این پژوهش در بخش چهارم ارائه شده‌است. بخش پنجم به بیان یک راهکار مناسب برای حل معادلات بهینه‌سازی در سیستم‌های k از n وزنی چندوضعیتی پرداخته‌است و سرانجام در بخش ششم نتیجه‌ی کلی پژوهش به همراه فراهم آوردن بستری برای فعالیت‌های آتی ارائه گردیده‌است.

۲ تعاریف اولیه

در این بخش تعاریف اولیه را به‌منظور بررسی سیستم‌های k از n وزنی، ابتدا در حالت ساده‌ی دودویی و سپس در حالت چندوضعیتی ارائه می‌دهیم. همچنین وزن‌ها و قابلیت‌های اعتماد مؤلفه‌ها را به‌عنوان متغیرهای طراحی و هزینه‌ی سیستم را تابعی از پیش تعریف‌شده برحسب آنها در نظر می‌گیریم و به جزئیات بهینه‌سازی نوع سیستم می‌پردازیم. سیستم k از n

سیستمی است که زمانی کار می‌کند که حداقل k مؤلفه از n مؤلفه‌ی آن کار کنند. بنابراین اگر φ تابع ساختار سیستم باشد و $\sum_{i=1}^n x_i \geq k$ آنگاه $\varphi(x) = 1$ و در غیراین صورت $\varphi(x) = 0$. سیستم k از n وزنی تعمیم سیستم k از n است که به هر کدام از مؤلفه‌های سیستم، وزن صحیح و مثبت w_{ij} نسبت می‌دهیم تا نشان دهیم برای مؤلفه‌های سیستم سطوح عملکرد متفاوتی را قائل می‌شویم. سیستم k از n وزنی دودویی در یک سیستم k از n وزنی دودویی G ، سیستم کار می‌کند اگر و تنها اگر وزن تمام مؤلفه‌های در حال کار، حداقل k (مقداری از پیش تعیین شده) باشد. این سیستم معادل است با یک سیستم $w - k + 1$ از n وزنی دودویی F که سیستم از کار می‌افتد اگر و تنها اگر وزن کل مؤلفه‌های از کار افتاده حداقل $w - k + 1$ باشد. مدل سیستم k از n دودویی حالت خاصی از مدل سیستم k از n وزنی دودویی است، زیرا می‌توان به هر مؤلفه‌ی آن وزن یک تخصیص داد. بعلاوه منظور از وزن هر مؤلفه، میزان اهمیتی است که آن مؤلفه در کارایی سیستم دارد. برای مثال، در یک هواپیمای جت که معمولاً با چندین موتور کار می‌کند، موتوری که فشار بیشتری را تولید کند، وزن بالاتری خواهد داشت. سیستم‌های k از n وزنی چندوضعیتی

در این سیستم‌ها، مؤلفه‌ها بیش از دو وضعیت دارند و نسبت به هر وضعیت، یک میزان اهمیت برای کارایی سیستم خواهند داشت و در صورتی که مؤلفه‌ای کاملاً از کار بیفتد، وزن آن در سیستم صفر است. فرض می‌کنیم n مؤلفه در سیستم داریم که هر کدام در $M+1$ وضعیت ممکن می‌توانند قرار داشته باشند و $\{0, 1, 2, \dots, M\}$. به مؤلفه‌ی i در وضعیت j مقدار w_{ij} را تخصیص می‌دهیم. برای این سیستم‌ها توسط لی و ژو [۴] دو تعریف متفاوت براساس دو مدل مختلف ارائه شده است: **مدل (I)**: سیستم در وضعیت j یا بیشتر است، هرگاه وزن همه‌ی مؤلفه‌ها برابر با یا بیشتر از یک مقدار از پیش تعیین شده‌ی k_j باشد. به این معنا که اگر φ تابع ساختار سیستم باشد، داریم

$$Pr(W_{total} \geq k_j) = Pr(\varphi \geq j)$$

و از آنجا که وضعیت صفر بدترین وضعیت است، $Pr(\varphi \geq 0) = 1$. **مدل (II)**: سیستم در وضعیت j یا بالاتر است اگر وزن کل مؤلفه‌هایی که در وضعیت j یا بیشترند، برابر با یا بیشتر از مقدار از پیش تعیین شده‌ی k_j باشد. براین اساس داریم

$$Pr(\varphi \geq j) = Pr(w_j \geq k_j)$$

که w_j مجموع وزن‌های مؤلفه‌هایی است که در وضعیت j یا بیشترند.

۳ مدل‌های طراحی

در بحث طراحی بهینه‌ی سیستم‌های k از n دودویی، اغلب پژوهشگران، مقدار بهینه برای مرتبه‌ی سیستم n ، مقدار بهینه‌ی k و یا مقادیر بهینه‌ی همزمان برای n و k را تعیین می‌کنند. فرم رایجی که به‌عنوان تابع هدف در نظر گرفته می‌شود و تلاش می‌شود تا آن را مینیمم کرد به صورت $cn + d(1 - R_s)$ است. به‌طوری‌که؛ c هزینه‌ی هر مؤلفه، d هزینه‌ی یک سیستم از کارافتاده و R_s قابلیت اعتماد سیستم، تابعی برحسب ماتریس‌های توزیع وزن و قابلیت اعتماد مؤلفه‌ها است.

$$R_s = R(W, P)$$

ماتریس قابلیت اعتماد

$$P = \begin{bmatrix} p_{10} & p_{11} & \dots p_{1M} \\ p_{20} & p_{21} & \dots p_{2M} \\ \vdots & \vdots & \vdots \\ p_{n0} & p_{n1} & \dots p_{nM} \end{bmatrix}$$

است که در آن p_{ij} احتمال مؤلفه‌ی i در وضعیت j می‌باشد. همچنین ماتریس وزن

$$W = \begin{bmatrix} w_{10} & w_{11} & \dots w_{1M} \\ w_{20} & w_{21} & \dots w_{2M} \\ \vdots & \vdots & \vdots \\ w_{n0} & w_{n1} & \dots w_{nM} \end{bmatrix}$$

است که در آن w_{ij} وزن مؤلفه‌ی i در وضعیت j و $i \in \{1, 2, \dots, n\}$ و $j \in \{1, 2, \dots, M\}$ می‌باشد.

بنابراین U و P را متغیرهای تصمیم و مقادیر k_j و n را مقادیر ثابت در نظر می‌گیریم. این کار به این دلیل است که می‌خواهیم تعیین کنیم کدام مؤلفه‌ها در طراحی سیستم بهینه مورد استفاده قرار می‌گیرند. برای شروع به تابعی احتیاج داریم که رابطه‌ی بین هزینه، قابلیت اعتماد و وزن مؤلفه‌ها را توصیف کند. می‌تاس و همکاران [۳] تابع هزینه‌ی مؤلفه را برحسب قابلیت اعتماد آن برای مؤلفه‌ی دودویی به‌صورت زیر ارائه دادند.

$$c = \exp[(1 - f) \frac{p - p_{min}}{p_{max} - p}] \quad (1)$$

که در آن؛ p_{min} کمترین قابلیت اعتماد، p_{max} بیشترین قابلیت اعتماد، c هزینه و p قابلیت اعتماد مؤلفه می‌باشند و f مقدار ثابت و مثبتی است که میزان دشواری افزایش قابلیت اعتماد مؤلفه را بیان می‌کند. زمانی که بخواهیم این

معادله را به i -امین مؤلفه‌ی سیستم k از n وزنی چندوضعیتی در مدل (I) بسط بدهیم، مقدار p را با $\sum_{j=1}^M p_{ij}$ باید جایگزین کنیم. در این صورت خواهیم داشت

$$c_i^I = \exp[(1 - f_i^I) \frac{\sum_{j=1}^M p_{ij} - P_{i_{min}}^I}{P_{i_{max}}^I - \sum_{j=1}^M p_{ij}}] \quad (2)$$

به‌طوری‌که؛ بیشترین احتمال مؤلفه‌ی i وقتی‌که در وضعیتی بالاتر از صفر باشد، c_i^I هزینه‌ی مؤلفه‌ی i در مدل (I) با توجه به رابطه‌ی بین قابلیت اعتماد و هزینه‌ی مؤلفه و f_i^I پارامتری است که میزان دشواری افزایش احتمال مؤلفه‌ی i در مدل (I) را وقتی بخواهد در وضعیتی بالاتر از وضعیت صفر باشد نشان می‌دهد، به طوری‌که هرچه این مقدار پایین‌تر باشد، تابع هزینه با سرعت بیشتری به بی‌نهایت میل می‌کند.

اگر در معادله‌ی دوم قراردهیم $M = 2$ به معادله‌ی اول می‌رسیم که حالت خاصی از معادله‌ی دوم است. همچنین احتیاج داریم که رابطه‌ی بین هزینه‌ی یک مؤلفه و وزن آن را به‌دست آوریم. اما از آنجا که در اغلب اوقات، یک رابطه‌ی تجربی مشخص در دسترس نیست و یا داده‌های تجربی برای برازش چنین تابعی کافی نیستند، به جای اینکه در پی برازش تابع هزینه به تابع قابلیت اعتماد و وزن براساس داده‌های تجربی باشیم، تلاش می‌کنیم برخی گرایش‌های موجود در داده‌ها را بازتاب دهیم و از معادله‌های ریاضی برای توصیف این گرایش‌ها استفاده کنیم. واضح است که هر محصولی که برای مشتری احساس رضایت‌مندی بیشتری را برآورده سازد، قیمت بالاتری خواهد داشت. بعلاوه با بالاتر رفتن ویژگی‌های مطلوب محصول، قیمت آن هم با سرعت بیشتری بالا می‌رود، بنابراین انتخاب یک تابع صعودی مثل تابع نمایی می‌تواند برای بیان تقریبی رابطه‌ی بین هزینه و وزن مؤلفه به صورت زیر مناسب به نظر برسد.

$$c = g[\exp(w - w_{min})] \quad (3)$$

که w_{min} کمترین وزن مؤلفه، c هزینه‌ی مؤلفه و w وزن مؤلفه می‌باشند و g مقدار ثابت و مثبتی است که میزان دشواری افزایش وزن مؤلفه را بیان می‌کند. هرچه میزان دشواری افزایش وزن مؤلفه‌ای بالاتر باشد، هزینه‌ی آن هم بالاتر می‌رود و از آنجا که معقول نیست محصولی وزن صفر داشته‌باشد، معمولاً مقدار w برابر یا بیشتر از یک مقدار داده‌شده‌ی w_{min} در نظر گرفته می‌شود.

وقتی این تابع را به i -امین مؤلفه‌ی سیستم k از n وزنی چندوضعیتی مدل (I) بسط دهیم، وزن مؤلفه‌ی دودویی را

باید با وزن مورد انتظار مؤلفه‌ی چندوضعیتی که میانگین وزنی وزن‌های مؤلفه‌ها در وضعیت‌های مختلف است یعنی $\sum_{j=0}^M p_{ij} w_{ij}$ جایگزین کنیم. بنابراین معادله‌ی ۳ به صورت زیر بازنویسی می‌شود.

$$c_i^I = g_i^I [\sum_{j=0}^M p_{ij} w_{ij} - w_{i_{min}}^I] \quad (۴)$$

به‌طوری‌که؛ $w_{i_{min}}^I$ کمترین مقدار برای میانگین وزنی وزن‌های i -امین مؤلفه‌ها در تمام وضعیت‌های ممکن، c_i^I هزینه‌ی مؤلفه‌ی i در مدل (I) برحسب رابطه‌ی بین هزینه و وزن مؤلفه‌ها، g_i^I پارامتری است که میزان دشواری برای افزایش وزن مؤلفه‌ی i در مدل (I) را نشان می‌دهد. هرچه این مقدار بالاتر باشد، تابع هزینه با سرعت بیشتری به بی‌نهایت میل می‌کند.

همانطور که مشاهده می‌شود، معادله ۴ تأثیر احتمال وضعیت مؤلفه را روی هزینه‌ی آن نادیده می‌گیرد. واضح است که روش معقول‌تر این است که وزن و احتمال وضعیت مؤلفه را همزمان در نظر بگیریم. در این صورت به تابع زیر خواهیم رسید که در واقع معادله‌های ۲ و ۴ را با یکدیگر ادغام می‌کند.

$$c_i^I = g_i^I [\exp(\sum_{j=0}^M p_{ij} w_{ij} - W_{i_{min}}^I) + \exp[(1 - f_i^I) \frac{\sum_{j=0}^M p_{ij} - P_{i_{min}}^I}{P_{i_{max}}^I - \sum_{j=0}^M p_{ij}}]] \quad (۵)$$

قرار می‌دهیم $P_i^I = \sum_{j=0}^M p_{ij}$ و $W_i^I = \sum_{j=0}^M p_{ij} w_{ij}$ و معادله ۵ به صورت زیر بازنویسی می‌شود.

$$c_i^I = g_i^I [\exp(W_i^I - W_{i_{min}}^I)] + \exp[(1 - f_i^I) \frac{P_i^I - P_{i_{min}}^I}{P_{i_{max}}^I - P_i^I}] \quad (۶)$$

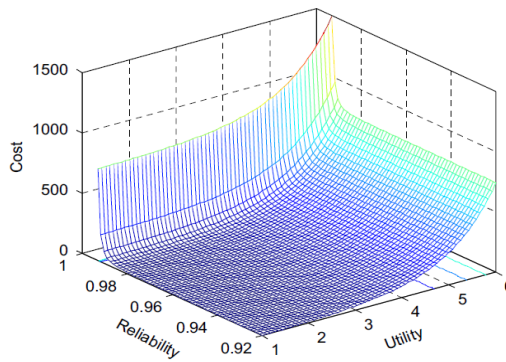
اگر مقادیر داده‌شده‌ی زیر را داشته‌باشیم، تابع ارائه‌شده در ۶ دارای شکل ۱ خواهدبود.

$$P_{i_{max}}^I = ۰.۸۹۹۹, P_{i_{min}}^I = ۰.۸۳, W_{i_{min}}^I = ۱, f_i^I = ۰.۴, g_i^I = ۵,$$

$$P_i^I \in [۰.۸۳, ۰.۴۴۴۴], W_i^I \in [۱, ۶]$$

در صورتی‌که بخواهیم براساس مدل (II) تابع هزینه را به‌دست آوریم، داریم

$$c_i^{II} = g_i^{II} [\exp(W_i^{II} - W_{i_{min}}^{II})] + \exp[(1 - f_i^{II}) \frac{P_i^{II} - P_{i_{min}}^{II}}{P_{i_{max}}^{II} - P_i^{II}}] \quad (۷)$$



شکل ۱: تابع هزینه برحسب قابلیت اعتماد و وزن مؤلفه

که در آن؛ $W_i^{II_j} = \sum_{r=j}^M p_{ir} w_{ir}$ و $P_i^{II_j} = \sum_{r=j}^M p_{ir}$ ؛ کمترین احتمال مؤلفه‌ی i است وقتی در وضعیت j یا بالاتر باشد؛ $P_{i_{max}}^{II_j}$ بیشترین احتمال مؤلفه‌ی i است وقتی در وضعیت j یا بالاتر باشد؛ $f_i^{II_j}$ پارامتری است که میزان دشواری افزایش احتمال مؤلفه‌ی i را وقتی که در وضعیت j یا بالاتر باشد نشان می‌دهد؛ $W_{i_{min}}^{II_j}$ کمترین میانگین وزنی وزن مؤلفه‌ها در تمام وضعیت‌هاست؛ $g_i^{II_j}$ پارامتری است که میزان دشواری افزایش وزن مؤلفه‌ی i را برای مدل (II) نشان می‌دهد و $c_i^{II_j}$ هزینه‌ی مؤلفه‌ی i برای مدل (II) است. تفاوت بین مدل (I) و مدل (II) این است که دریابیم آیا مؤلفه‌هایی که وضعیت آنها کمتر از j است، کمکی به سیستم می‌کنند که به وضعیت j یا بالاتر برود یا خیر. در مدل دوم تنها مؤلفه‌هایی که وضعیت آنها j یا بالاتر باشد می‌توانند به سیستم کمک کنند تا به وضعیت j و بالاتر برسد. به همین دلیل داشتیم $W_i^{II_j} = \sum_{r=j}^M p_{ir} w_{ir}$ و $P_i^{II_j} = \sum_{r=j}^M p_{ir}$. توجه می‌کنیم که ممکن است پژوهشگران مختلف، فرم‌های کلی‌تر و پارامترهای متفاوتی را برای سیستم‌های وزنی چندوضعیتی ارائه دهند، اما ایده‌ی اصلی برای این توابع (به منظور توصیف رفتار کلی هزینه به‌عنوان تابعی از قابلیت اعتماد و وزن مؤلفه) مشابه خواهد بود.

با روابط داده‌شده در معادله‌های ۶ و ۷ مسائل بهینه‌سازی را به‌صورت زیر می‌توان فرموله کرد.

مسئله یک. مینیمم سازی:

$$C_s^{T_j} = \sum_{i=1}^n c_i^X + (1 - R_j^T(k_j, n)) C_j^T \quad (۸)$$

با در نظر گرفتن:

$$R_j^T(k_j, n) \geq \hat{R}_j^T$$

$$\sum_{j=\circ}^M p_{ij} = 1, \circ \leq p_{ij} \leq 1 (i = 1, 2, \dots, n; j = \circ, 1, 2, \dots, M)$$

$$w_{i\circ} = \circ, w_{ij} \geq \circ$$

مسئله دو. ماکسیمم سازی:

$$R_j^T(k_j, n) \quad (9)$$

با در نظر گرفتن:

$$\begin{aligned} C_s^{T_j} &= \sum_{i=1}^n c_i^X + (1 - R_j^T(k_j, n)) C_j^T \leq \hat{C}_s^{T_j} \\ \sum_{j=0}^M p_{ij} &= 1, 0 \leq p_{ij} \leq 1 (i = 1, 2, \dots, n; j = 0, 1, 2, \dots, M) \\ w_{i^*} &= 0, w_{ij} \geq 0 \end{aligned}$$

که در آن؛ $C_s^{T_j}$ کل هزینه‌ی موردانتظار وقتی سیستم در وضعیت j یا بالاتر باشد؛ $R_j^T(k_j, n)$ احتمال اینکه سیستم در وضعیت j یا بالاتر باشد و k_j ، j و n مقادیر داده شده‌اند؛ C_j^T هزینه‌ی سیستم وقتی در وضعیت کمتر از j باشد؛ $\hat{R}_j^T$ کمترین احتمال موردنیاز برای سیستم که به وضعیتی بالاتر از j برسد؛ $\hat{C}_s^{T_j}$ حد بالا برای کل هزینه وقتی که سیستم به وضعیتی بالاتر از j برسد؛ $T \in \{I, II\}$ ، $X \in \{I, II_j\}$ در حالت خاص، وقتی $T = I$ ، X باید برابر با ۱ باشد. تحت این شرایط، دو مدل بالا (مسأله‌های یک و دو) برای یک سیستم k از n وزنی چندوضعیتی مدل (I) مناسبند و تابع هزینه‌ی مؤلفه به صورت معادله ۶ خواهد بود. وقتی $T = II$ ، X باید برابر با II_j باشد. تحت این شرایط، دو مدل فوق برای یک سیستم k از n وزنی چندوضعیتی مدل (II) مناسبند و تابع هزینه‌ی مؤلفه به صورت معادله ۷ خواهد بود.

در مدل‌های فوق، C_j^T ، $\hat{R}_j^T$ ، $\hat{C}_s^{T_j}$ ، j ، k_j ، f_i^X ، g_i^X ، W_{imin}^X ، P_{imax}^X ، P_{imin}^X ، M و n همگی پارامترهای مدل هستند و متغیرهای تصمیم P و W می‌باشند. علیرغم اینکه در عمل بسیار سخت است که بگوییم توزیع وزن مؤلفه‌ها در هر شرایطی تحت کنترل است، اما در این مدل‌ها فرض می‌کنیم که چنین کنترلی مقدور باشد. زیرا با در نظر گرفتن سرعت پیشرفت تکنولوژی در دنیای صنعتی امروزی چنین فرضی نامعقول هم نخواهد بود.

۴ مثال کاربردی

کاربردهای این مدل‌های بهینه‌سازی را می‌توان به‌طور گسترده در صنایع مهندسی شامل نفت و گاز، سیستم‌های حمل و نقل و مخابرات دید. برای روشن‌تر شدن این موضوع، مثال زیر را بررسی می‌کنیم.

فرض کنید چندین حامل داریم که به صورت موازی به یکدیگر متصل شده‌اند. در طول مدت عمر حامل، وضعیت آن از "بی نقص" به "ازکارافتاده" رو به زوال می‌رود. زمانی که وضعیت حامل تنزل پیدا کند ممکن است دیگر نتواند یک بار کامل را حمل کند. بنابراین این حامل در وضعیت‌های متفاوت، وزن متفاوتی خواهد داشت. در صورتی که مایل باشیم تمام حامل‌های این سیستم بتوانند سوخت کافی برای تولید حداقل ۱۰۰ کیلووات را انتقال دهند و اهمیتی نداشته باشد هر حامل در چه وضعیتی قرار دارد، می‌توانیم سیستم k از n وزنی چندوضعیتی مدل (I) را مورد استفاده قرار دهیم. اگر بخواهیم کل سیستم بتواند یک بار کامل (یا بیشتر از سطح مشخصی) را حمل کند، در این صورت آن حامل‌هایی که قادر نباشند مقدار بار موردنظر را حمل کنند، برای ما سودی ندارند. در این حالت می‌توانیم سیستم k از n وزنی چندوضعیتی مدل (II) را مورد استفاده قرار دهیم. اگر طراحان بخواهند هزینه را کاهش دهند و همزمان قابلیت اعتماد را در سطح مشخصی حفظ کنند، از آنجا که سیستم حامل یک سیستم k از n وزنی چندوضعیتی است، مدل بهینه‌سازی پیشنهادی براساس **مسئله یک** در این مورد مناسب است.

در عمل، انتخاب بین مدل (I) و (II) بستگی به نیازمندی‌های سیستم حامل دارد. برای مثال، اگر نیاز باشد که تمام حامل‌ها یک بار کامل را حمل کنند، **مسئله یک** برای مدل (II) باید استفاده شود. اگر هیچ نیاز خاصی به وضعیت حامل‌ها نداشته باشیم، **مسئله دو** برای مدل (I) باید استفاده شود.

به‌طور مشابه، اگر بودجه‌ی ثابتی برای ساخت یک سیستم حامل در نظر گرفته شده باشد و هدف به‌دست آوردن بالاترین قابلیت اعتماد ممکن تحت این بودجه باشد، **مسئله دو** مناسب خواهد بود.

۵ راهکار مناسب برای حل معادلات بهینه‌سازی

خانواده‌ی تکنیک‌های بهینه‌سازی اکتشافی شامل؛ شبیه‌سازی آنیل، الگوریتم ژنتیک (GA)، استراتژی‌های تکاملی، کلونی مورچه و روش‌های دیگر است، که از آنها برای حل مسائل بهینه‌سازی پیچیده و بزرگ مقیاس استفاده می‌شود. این روش‌ها بسیار بهتر و کاراتر از روش‌های بهینه‌سازی قدیمی عمل می‌کنند. به‌طور خاص، الگوریتم ژنتیک روشی است که به‌طور گسترده‌تری مورد استفاده‌ی پژوهشگران است. این روش دو برتری کلیدی نسبت به دیگر روش‌ها دارد:

(۱) انعطاف‌پذیری در مدل‌سازی مسائل: الگوریتم ژنتیک به ملزومات ریاضی پیچیده مانند محدودیت‌های

مشتق‌گیری برای توابع هدف و شرایط بسیار احتیاجی ندارد. تنها احتیاج داریم که توابع هدف و شرایط آنها به طریقی برآورد شوند.

(۲) توانایی بهینه‌سازی عمومی

در این مقاله، از روش الگوریتم ژنتیک برای حل مسائل طراحی شده در بخش ۲ استفاده می‌کنیم. به‌منظور محاسبه‌ی آسان‌تر، توابع هدف و شرایط مسئله‌های یک و دو را به‌صورت تابع جریمه در زیر می‌آوریم.

مسئله یک. مینیم سازی:

$$\sum_{i=1}^n c_i^T + (1 - R_j^T(k_j, n))C_j^T + \max(\hat{R}_j^T - R_j^T(k_j, n), 0)\eta. \quad (10)$$

مسئله دو. مینیم سازی:

$$-R_j^T(k_j, n) + \max(\sum_{i=1}^n c_i^T + (1 - R_j^T(k_j, n))C_j^T - \hat{C}_s^{Tj}, 0)\eta. \quad (11)$$

در این توابع هدف، η مقداری بسیار بزرگ است. (ما در برنامه ۹۹۹۹۹۹ را قرار می‌دهیم). سیستمی با دو مؤلفه ($n = 2$) که هر مؤلفه چهار حالت داشته باشد ($M = 3$) را در نظر می‌گیریم. مقادیر زیر داده‌شده‌اند

$$j = 2, k_T = 3, R_T^T = 0.8, C_T^T = 10, \hat{C}_s^{Tj} = 9$$

دیگر پارامترهای مورد استفاده در این مثال در جدول ۱ نمایش داده‌شده‌اند.

پس از اجرای این برنامه در نرم‌افزار *MATLAB* و با در نظر گرفتن حجم ۲۰ برای اندازه‌ی جامعه، معیار توقف ۱۰۰۰ و دیگر مقادیر مورد نیاز، نتایج بهینه‌سازی برای مدل‌های (I) و (II) را در جدول‌های ۲ و ۳ خلاصه کردیم. در هر یک از این مسائل بهینه‌سازی، چهارده متغیر تصمیم داریم و هر دو مسئله غیرخطی هستند. بنابراین روش‌های بهینه‌سازی پیشین برای حل این مسائل بسیار دشوار خواهند بود، اما با استفاده از الگوریتم ژنتیک زمان لازم برای به جواب رسیدن، هفت ثانیه است. اگر مقدار n از ۲ به ۱۰ افزایش پیدا کند و کلیه پارامترهای دیگر بدون تغییر

جدول ۴: پارامترهای مورد استفاده در مثال

$P_{max}^I = P_{max}^I$	$P_{min}^I = P_{min}^I$	$W_{min}^{Ij} = W_{min}^{Ij}$	$f_1^I = f_1^I$	$g_1^I = g_1^I$
۰,۹۹۹۹۹	۰,۹۳	۱	۰,۹۹	۱
$P_{max}^{IIj} = P_{max}^{IIj}$	$P_{min}^{IIj} = P_{min}^{IIj}$	$W_{min}^{IIj} = W_{min}^{IIj}$	$f_1^{IIj} = f_1^{IIj}$	$g_1^{IIj} = g_1^{IIj}$
۰,۹۹۹۹۹	۰,۹۳	۱	۰,۹۹	۱

جدول ۵: نتایج بهینه‌سازی در مدل (I)

مسئله دو				مسئله یک			
۰,۹۹۲۷				۰,۹۷۸۱			
۸,۸۵۲۵				۵,۹۷۵۳			
۳,۰۰۰۰	۳,۰۰۰۰	۱,۱۱۸۵	۰	۲,۰۰۰۱	۲,۰۰۰۰	۲,۰۰۰۰	۰
۲,۰۰۰۰	۱,۹۹۹۶	۱,۹۹۹۴	۰	۱,۰۰۲۶	۱,۰۰۰۰	۱,۰۰۰۰	۰
۰,۳۳۱۴	۰,۳۳۱۴	۰,۳۳۲۶	۰,۰۰۵۶	۰,۳۲۹۱	۰,۳۲۹۱	۰,۳۲۹۱	۰,۰۱۲۷
۰,۳۳۰۲	۰,۳۲۹۹	۰,۳۳۰۲	۰,۰۱۰۷	۰,۳۳۰۳	۰,۳۲۹۹	۰,۳۳۰۵	۰,۰۰۸۷

باقی بمانند، تعداد هفتاد متغیر در هر مدل داریم و کل زمان پردازش چهار دقیقه می‌شود. اگر مقدار n را به ۲۰ ارتقا دهیم و دیگر پارامترها بدون تغییر باقی بمانند، ۱۴۰ متغیر در هر مدل داریم و زمان پردازش در حدود سی دقیقه طول خواهد کشید.

نتایج نشان می‌دهند که مسائل بهینه‌سازی که کل هزینه را حداقل می‌کنند، نسبت به مسائلی که احتمال اینکه سیستم در وضعیتی مشخص یا بالاتر از آن باشد را ماکسیم کنند، متحمل هزینه‌ی کلی پایین‌تری هستند. از طرف دیگر، مسائل بهینه‌سازی که احتمال اینکه یک سیستم در وضعیتی مشخص یا بالاتر از آن باشد را ماکسیم می‌کنند، در رسیدن به هدفشان نسبت به مسائلی که هزینه‌ی کل را حداقل می‌کنند، توانا ترند. چنین نتیجه‌ای منطقی به نظر می‌رسد. زیرا زمانی که هدف، تنها حداقل نمودن کل هزینه باشد، ترجیح می‌دهیم برخی احتیاجات برای احتمال اینکه سیستم در وضعیتی مشخص یا بالاتر از آن باشد را فقط با یک شرط روی آن احتمال ارائه دهیم. به طور مشابه، زمانی که هدف، تنها ماکسیم نمودن این احتمال باشد که سیستم در وضعیتی مشخص یا بالاتر از آن باشد، برای محدود کردن کل هزینه شرایطی قرار می‌دهیم. در نهایت فرد تصمیم‌گیرنده باید با در نظر گرفتن این نتایج تصمیم بگیرد قابلیت اعتماد را تا چه آستانه‌ای می‌تواند بهبود ببخشد.

جدول ۶: نتایج بهینه‌سازی در مدل (II)

مسئله دو				مسئله یک				
۰,۹۷۷۰				۰,۹۶۹۰				قابلیت اعتماد
۸,۰۹۰۵				۵,۹۴۷۳				هزینه
۳,۰۰۰۰	۱,۹۴۲۵	۰,۰۰۰۰	۰	۲,۰۴۵۶	۲,۰۴۲۳	۰,۰۰۰۰	۰	W
۱,۷۶۱۰	۱,۲۵۰۸	۰,۰۰۰۰	۰	۰,۹۶۱۶	۰,۹۶۰۴	۰,۰۰۰۰	۰	
۰,۴۹۸۶	۰,۵۰۰۳	۰,۰۰۰۶	۰,۰۰۱۲	۰,۴۹۰۸	۰,۴۸۴۴	۰,۰۰۹۵	۰,۰۱۶۲	P
۰,۴۶۴۵	۰,۴۹۰۷	۰,۰۴۳۱	۰,۰۰۲۷	۰,۴۹۶۸	۰,۴۹۶۸	۰,۰۰۳۴	۰,۰۰۳۹	

۶ نتیجه‌گیری و پژوهش‌های آتی

در این مقاله، دو مسئله‌ی بهینه‌سازی برای سیستم k از n وزنی چندوضعیتی براساس دو مدل (I) و (II) ارائه شد. تمامی این مسائل یا به‌صورت حداقل کردن کل هزینه‌ی مورد انتظار سیستم با توجه به احتیاجات قابلیت اعتماد و یا به‌صورت حداکثر کردن قابلیت اعتماد سیستم با توجه به محدودیت‌های هزینه، مدل‌سازی می‌شوند. از الگوریتم ژنتیک برای حل معادلات بهینه‌سازی استفاده شده‌است که در حال حاضر ابزار قدرتمندی برای این محاسبات پیچیده است. پس از اینکه نتایج حاصل می‌شوند، اغلب سعی می‌کنیم در شرایط هر مسئله تغییرات اندکی را اعمال کنیم تا بتوانیم مقادیر موردنظرمان را سبک و سنگین کنیم و مقادیر بهینه را جایگزین کنیم. یافتن مناسب‌ترین شرایط، یک فرآیند آزمون و خطاست که بسیار وقت‌گیر است. به همین دلیل، امید است که در پژوهش‌های آتی از ابزار بهینه‌سازی چندمنظوره‌ای برای بررسی همزمان چندین هدف به‌طور همزمان استفاده کنیم.

مراجع

- [1] Eryilmaz, S., and Bazbulut, A. R. (2014). *An algorithmic approach for the dynamic reliability analysis of non-repairable multi-state weighted k-out-of-n systems*, Reliability Engineering and system safety. 131, 61–65.

- [2] Faghieh-Roohi, S., MingNg, K., Xie, M., Richard, C. M. Y. (2013), *Dynamic availability assessment and optimal component design of multi-state weighted k-out-of-n systems*, Reliability Engineering and system safety. 123, 57–62.
- [3] Mettas, A. (2000), ReliaSoft corporation and Tucson, *reliability allocation and optimization for complex systems*, In: Proceedings of the annual reliability and maintainability symposium, Los Angeles, CA, USA [January 24-27].
- [4] Li, W., Zuo, M. J. (2008), *Reliability evaluation of multi-state weighted k-out-of-n systems*, Reliability Engineering and system safety. 93(1), 161–168.
- [5] Li, W., Zuo, M. J. (2008), *Optimal design of multi-state weighted k-out-of-n systems based on component design*, Reliability Engineering and system safety. 93, 1673–1681.



بررسی قابلیت اعتماد یک سیستم با استفاده از مفهوم تنش مقاومت

بازیاری، ا^۱ کرمی، م^۲

^۱ گروه آمار، دانشکده علوم پایه، دانشگاه خلیج فارس بوشهر

^۲ دانشجوی کارشناسی ارشد دانشگاه شیراز

چکیده

در این مقاله، مفهوم تنش مقاومت بطور کامل توضیح داده شده و قابلیت اعتماد یک سیستم از دیدگاه تنش مقاومت مورد بررسی قرار گرفته است. پارامتر تنش مقاومت در توزیع لامبدای تعمیم یافته محاسبه شده و نیز به ازای مقادیر مختلف پارامترهای توزیع لامبدای تعمیم یافته، پارامتر تنش مقاومت محاسبه شده است. با مثالی واقعی کاربرد نتایج نشان داده شده است.

کلمات کلیدی: پارامتر تنش مقاومت، قابلیت اعتماد.

^۱ ab.bazaryari@gmail.com

^۲ mahvashkarami@yahoo.com

۱ پیش‌گفتار

قابلیت اعتماد در سیستم‌های مهندسی از اهمیت بالایی برخوردار هستند. همچنین دانستن توزیع آماری مولفه‌هایی که در سیستم در حال کار می‌باشند نیز یکی از موارد قابل لحاظ در محاسبه قابلیت اعتماد آن سیستم است. سیستم‌های مختلفی از جمله سیستم‌های موازی، سیستم‌های سری و بسیاری دیگر از سیستم‌ها در شاخه‌های مهندسی و صنایع کاربرد فراوان دارند. یک سیستم در حال کار همواره مولفه‌های آن با فشارها و تنش‌هایی مواجه می‌شوند که این مولفه‌ها در برابر فشارهای وارده و تهاجمات مقاومت می‌کنند تا سیستم به فعالیت خود ادامه دهد. توزیع لامبدای تعمیم یافته^۳ یک توزیع آماری^۴ مفید برای آزمون و برازش داده‌ها است. برازش داده‌ها به یک توزیع احتمالی یکی از مسایل اساسی و مهم در تحلیل داده‌های آماری است که بایستی به اندازه کافی به داده‌های تجربی جمع آوری شده متناسب و نزدیک باشد. در این زمینه مطالب قابل توجهی وجود دارد و تحقیقات زیادی از طرف نویسندگان و محققان انجام شده است. یکی از توزیع‌های کاربردی که اخیراً مورد توجه آماردانان قرار گرفته، توزیع لامبدای تعمیم یافته می‌باشد. این توزیع یک توزیع پارامتری انعطاف‌پذیر برای داده‌های تک متغیره است. انعطاف‌پذیر بودن آن با وجود چهار پارامتر در این توزیع مشخص می‌شود. به‌طوری که می‌توان داده‌هایی از جوامعی با توزیع‌هایی مانند توزیع نرمال، نمایی، بتا، یکنواخت و فشر را توسط این توزیع برازش داد. اصل توزیع لامبدا توسط توکی (۱۹۶۲) مشخص و مورد بررسی قرار گرفت. از اوایل سال ۱۹۷۰ میلادی توزیع لامبدا تعمیم یافته برای برازش داده‌ها مورد توجه قرار گرفت. توزیع لامبدا توسط جویئر و روسنبلا (۱۹۷۱) و فیلیپ (۱۹۷۵) تعمیم داده شد. رامبرگ و اشمیس (۱۹۷۴) و رامبرگ و همکاران (۱۹۷۹) ویژگی‌های توزیع لامبدا تعمیم یافته را مورد بررسی قرار دادند. ریسر (۱۹۸۰) از توزیع لامبدا تعمیم یافته برای برازش داده‌های صنعتی با توجه به انحراف داده‌ها از حالت نرمال استفاده کرد. رامبرگ و همکاران (۱۹۷۹)، برگوین (۱۹۹۳) و کاریان و دودیچ (۲۰۰۰) برای برآورد پارامترهای توزیع لامبدای تعمیم یافته از روش گشتارها استفاده کردند. توزیع لامبدای تعمیم یافته دو متغیره همراه با مثال‌های کاربردی آن توسط بکویت و دودیچ (۱۹۹۶) مورد مطالعه قرار گرفت. سو (۲۰۰۶) از توزیع لامبدای تعمیم یافته برای برازش روی داده‌های دو مدی استفاده و با روش ماکسیم درستنمایی^۵ پارامترهای آن را برآورد کرد. سو (۲۰۰۷b) توزیع لامبدای تعمیم یافته را در نرم افزار مورد تجزیه و تحلیل قرار داد و از روش

Generalized lambda distribution^۳
Statistical distribution^۴
Maximum likelihood method^۵

ماکسیم درستمایی برای برآورد پارامترهای آن استفاده کرد. سو (۲۰۰۷a) از توزیع لامبدای تعمیم یافته برای برازش روی داده‌های تک مدی استفاده و با روش ماکسیم درستمایی پارامترهای آن را برآورد کرد. کینک و مککیلوری (۲۰۰۷) روش‌های مناسب را برای برازش داده‌ها با استفاده از توزیع لامبدای تعمیم یافته مکانی و مقیاسی مورد بررسی قرار دادند. برای اطلاع بیشتر در مورد کاربرد توزیع لامبدای تعمیم یافته به سو (۲۰۱۰) رجوع شود. چالابی و همکاران (۲۰۱۲) از توزیع لامبدای تعمیم یافته برای برازش روی داده‌های مالی^۶ استفاده کردند. در بخش دوم مقاله، به معرفی توزیع لامبدای تعمیم یافته و ویژگی‌های این توزیع پرداخته شده و توابع چگالی و توزیع آن به ازای مقادیر مختلف پارامترها رسم شده‌اند. در بخش سوم، مفهوم تنش مقاومت توضیح داده شده و قابلیت اعتماد یک سیستم از دیدگاه تنش مقاومت^۷ مورد بررسی قرار می‌گیرد. در بخش چهارم، فرم ریاضی پارامتر تنش مقاومت در توزیع لامبدای تعمیم یافته تعیین و به ازای مقادیر مختلف پارامترهای توزیع لامبدای تعمیم یافته، پارامتر تنش مقاومت محاسبه شده است. در بخش پنجم، با یک مثال با داده‌های واقعی نتایج بدست آمده در توزیع لامبدای تعمیم یافته مورد بررسی قرار می‌گیرد و پارامتر تنش مقاومت با روش‌های عددی محاسبه می‌شود. بحث و نتیجه‌گیری در بخش ششم داده شده است.

۲ توزیع لامبدای تعمیم یافته و ویژگی‌های آن

در این بخش، توزیع لامبدای تعمیم یافته و ویژگی‌های آن مورد بررسی قرار گرفته است. تابع چگالی و تابع توزیع آن به ازای مقادیر مختلف پارامترها رسم شده‌اند. تابع توزیع تجمعی معکوس توزیع لامبدای تعمیم یافته به صورت

$$F^{-1}(u, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{u^{\lambda_2} - (1-u)^{\lambda_2}}{\lambda_2},$$

تعریف می‌شود، که در آن^۱ $0 \leq u < 1$ ، λ_1 پارامتر مکان^۸، λ_2 پارامتر مقیاس^۹، λ_3 و λ_4 پارامترهای شکل (به ترتیب متناظر با چولگی و کشیدگی) می‌باشند. اگر $\lambda_1 = 0$ و $\lambda_2 = \lambda_3 = \lambda_4 = \lambda$ باشند، آن‌گاه توزیع لامبدا (توزیع لامبدای

Financial data^۶
Stress-strength^۷
Location parameter^۸
Scale parameter^۹

توکی) بدست می‌آید. تابع چگالی متناظر آن عبارت است از:

$$F^{-1}(u, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \frac{\lambda_2}{\lambda_3 u^{\lambda_2-1} + \lambda_4 (1-u)^{\lambda_2-1}}. \quad (1)$$

تابع چگالی اخیر برای تمام مقادیر مختلف پارامترهای شکل، λ_2 و λ_4 ، سره نمی‌باشد. رامبرگ و اشمیسر (۱۹۷۴) نشان دادند که فقط در شش ناحیه از فضای پارامتری^{۱۰} مقدار تابع چگالی احتمال رابطه (۱) وجود دارد. مقادیر تابع چگالی و نواحی مربوطه به ترتیب در جدول (۱) و نمودار (۱) آورده شده‌اند.

جدول ۷: دامنه پارامترهای توزیع لامبدای تعمیم یافته و مقادیر مینیم و ماکسیم آن

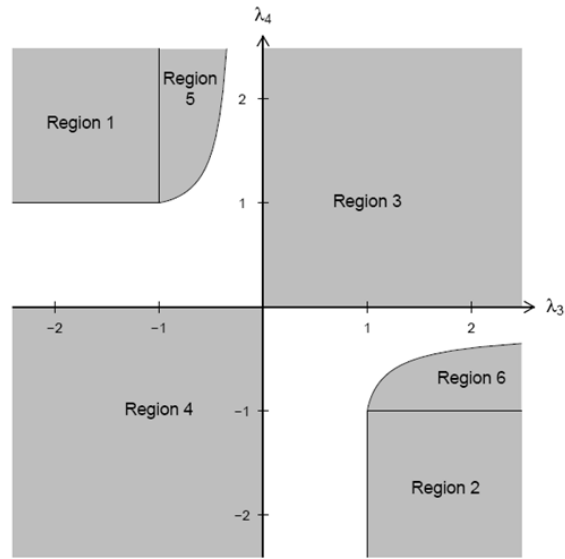
ناحیه	λ_1	λ_2	λ_3	λ_4	مینیم	ماکسیم
۱،۵	تمام مقادیر	$< \infty$	< -1	> 1	$-\infty$	$\lambda_1 + \frac{1}{\lambda_2}$
۲،۶	تمام مقادیر	$< \infty$	> 1	< -1	$\lambda_1 - \frac{1}{\lambda_2}$	∞
	تمام مقادیر	$> \infty$	$> \infty$	$> \infty$	$\lambda_1 - \frac{1}{\lambda_2}$	$\lambda_1 + \frac{1}{\lambda_2}$
۳	تمام مقادیر	$> \infty$	$= \infty$	$> \infty$	λ_1	$\lambda_1 + \frac{1}{\lambda_2}$
	تمام مقادیر	$> \infty$	$> \infty$	$= \infty$	$\lambda_1 - \frac{1}{\lambda_2}$	λ_1
	تمام مقادیر	$< \infty$	$< \infty$	$< \infty$	$-\infty$	∞
۴	تمام مقادیر	$< \infty$	$= \infty$	$< \infty$	λ_1	∞
	تمام مقادیر	$< \infty$	$< \infty$	$= \infty$	$-\infty$	λ_1

توزیع لامبدای تعمیم یافته در زمینه‌های مختلف علمی بکار می‌رود. به عنوان مثال، اوستورک و دال (۱۹۸۲) از این توزیع برای برآزش داده‌ها در علم هواشناسی استفاده کردند. دنگیز (۱۹۸۸) از توزیع لامبدای تعمیم یافته در شبیه‌سازی^{۱۱} سیستم‌های صف^{۱۲} استفاده کرد. برای مطالعه در مورد شبکه‌های عصبی با به‌کارگیری توزیع لامبدای تعمیم یافته به کاروانن و همکاران (۲۰۰۰) رجوع شود. از این توزیع برای قیمت‌گذاری‌های مالی و بازار اقتصادی جهت کنترل تورم استفاده می‌شود (کورادو، ۲۰۰۱). همچنین تارسیتانو (۲۰۰۴) به کاربرد توزیع لامبدای تعمیم یافته در مسایل مالی توجه کرد. بیگرل و همکاران (۲۰۰۶) از این توزیع همراه با آنالیز روش بوت‌استرپ برای پیش‌بینی طول عمر مواد و محاسبه بازه‌های اطمینان در سطوح مختلف معنی‌داری استفاده کردند. پارامترهای

^{۱۰}space Parameter

^{۱۱}Simulation

^{۱۲}Queue systems



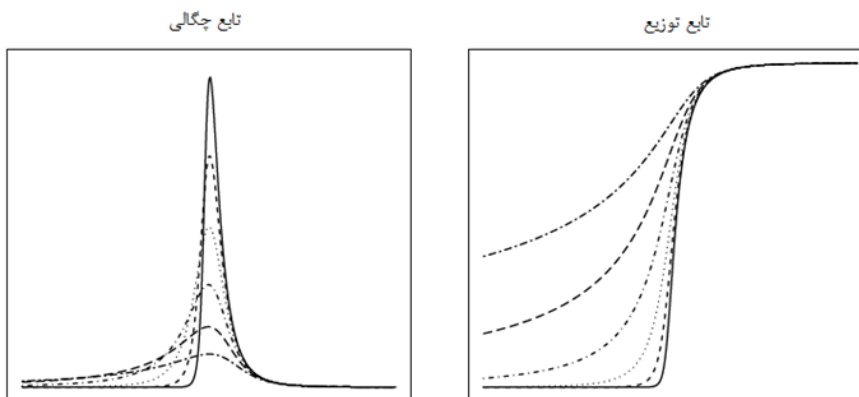
نمودار ۱: ناحیه پارامتری توزیع لامبدای تعمیم یافته در پارامترهای λ_3 و λ_4

تابع چگالی رابطه (۱,۲) که توسط رامبرگ و اشمیسر (۱۹۷۴) نشان داده شد، دارای شکل پیچیده و مشکلی می باشد. بنابراین فریمر و همکاران (۱۹۸۸) فرم جدیدی از چگالی توزیع لامبدای تعمیم یافته را برای از بین بردن محدودیت‌ها روی پارامترها به صورت

$$F^{-1}(u, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{1}{\lambda_2} \left[\frac{u^{\lambda_2} - 1}{\lambda_3} - \frac{(1-u)^{\lambda_4} - 1}{\lambda_4} \right]. \quad (2)$$

تعریف کردند. همانند قبل λ_3 و λ_1 به ترتیب پارامترهای مکان و پارامتر مقیاس و λ_4 و λ_2 پارامترهای شکل هستند. تابع چگالی رابطه (۲,۲) برای تمام مقادیر پارامترها خوش تعریف می باشد. نمودار تابع چگالی رابطه (۲,۲) و تابع توزیع آن با استفاده از نرم افزار *SPLUS* برای $\lambda_1 = 0$ و $\lambda_2 = -1$ و $\lambda_4 = -\frac{1}{4}$ و به ازای مقادیر مختلف λ_3 ، $-2, -1, -\frac{1}{2}, -\frac{1}{4}, -\frac{1}{8}, -\frac{1}{16}$ ، رسم و در نمودار ۲ نشان داده شده است.

بنا به رابطه (۱,۲)، تقارن و شیب چگالی توزیع لامبدای تعمیم یافته به پارامترهای λ_3 و λ_4 وابسته است، به طوری که



نمودار ۲: توابع چگالی و توزیع لامبدای تعمیم یافته به ازای مقادیر مختلف پارامترها

اگر این شاخص‌ها با هم برابر باشند، توزیع متقارن^{۱۳} خواهد بود. حتی اگر یکی از این شاخص‌ها افزایش یابد، باز هم توزیع متقارن خواهد بود. شیب توزیع به هر دوی این پارامترها وابسته است، در حقیقت افزایش هر دو شاخص با هم باعث تولید یک توزیع متقارن با دم‌های باریک می‌شود. بنابراین تشخیص این پارامترها برای اطلاع از تقارن و شیب توزیع لازم و اساسی خواهد بود.

۳ مفهوم تنش مقاومت

به‌طور کلی در یک سیستم حداقل دو عامل وجود دارد که یکی از آن‌ها بر دیگری تنش یا فشار وارد کرده و عامل دیگر مقاومت می‌کند. در این حالت پارامتر تنش مقاومت مطرح می‌شود. در سیستمی که به‌مولفه و اعضای آن فشار یا تنشی وارد می‌شود، سیستم در مقابل آن فشار مقاومت می‌کند. بر اساس این الگو هر چه تنش ایجاد شده بر سیستم بیشتر باشد، سیستم با احتمال شکست بیشتری همراه خواهد بود. به عبارت دیگر سیستم تا زمانی به کار و فعالیت خود ادامه می‌دهد که مقاومت سیستم از تنش وارده بر آن بیشتر باشد. پارامتر تنش مقاومت به‌صورت احتمال

$$P(X < Y) = \int_{-\infty}^{+\infty} P(X < Y | Y = y) f_Y(y) dy = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^x g(y) dy \right) f(x) dx. \quad (۳)$$

^{۱۳}Symmetric distribution

تعریف می‌شود، که در آن X متغیر تصادفی تنش و Y متغیر تصادفی مقاومت است و بر اساس آن می‌توان بقای یک سیستم را کنترل کرد. اصطلاح مدل تنش مقاومت اولین بار توسط چارج و هریس (۱۹۷۰) مطرح شد. پس از آن کارهای زیادی روی پارامتر تنش مقاومت بر اساس توزیع‌های مختلف و شرایط متفاوت حاکم بر متغیرهای تصادفی انجام شد. فرض کنید سیستمی دارای k مولفه است که هر یک دارای مقاومت‌های Y_1, Y_2, K, Y_k و فشار یکسان X باشند. یکی از مدل‌های معروف در این حالت مدل سیستم موازی است که این سیستم با موفقیت عمل می‌کند هرگاه حداقل یکی از مولفه‌ها در سیستم باقی باشد. در صورتی که سیستم سری باشد، بایستی برای ادامه کار تمام مولفه‌های آن کار کنند. بنابراین قابلیت اعتماد در سیستم‌های موازی و سری به ترتیب $P(X < \max(Y_1, Y_2, K, Y_k))$ و $P(X < \min(Y_1, Y_2, K, Y_k))$ می‌باشند. البته لازم به ذکر است که سیستم‌های دیگری نیز وجود دارند که قابلیت اعتماد این سیستم‌ها با احتمالات متفاوتی اندازه‌گیری می‌شوند. به عنوان مثال سیستم‌هایی که باید S مولفه در حال کار باشند یا سیستم‌هایی که شامل m زیر سیستم مستقل هستند. در برخی از سیستم‌ها به یک مولفه دو نوع تنش با اندازه‌های متفاوت وارد می‌شود که در این حالت قابلیت اعتماد سیستم از طریق مقدار احتمال $P(X < Y < Z)$ ، که Y متغیر تصادفی مقاومت و در آن X و Z متغیرهای تصادفی تنش می‌باشند، اندازه‌گیری می‌شود. به عنوان مثالی از این سیستم‌ها می‌توان به دستگاه‌هایی که باید بین دو دمای خاص کار کنند و یا سیستم بدن انسان که بین دو فشار خون مشخص در حال کار است، اشاره کرد. همچنین می‌توان سیستمی را در نظر گرفت که مولفه‌های تنش و مقاومت آن هر کدام شامل برداری از p متغیر تصادفی به صورت $\mathbf{X} = (X_1, X_2, K, X_p)$ و $\mathbf{Y} = (Y_1, Y_2, K, Y_p)$ باشند، که در این حالت نیز پارامتر تنش مقاومت از احتمال $P(\mathbf{X} < \mathbf{Y})$ محاسبه می‌شود. برای اطلاع بیشتر در مورد پارامتر تنش مقاومت و کاربرد آن می‌توان به کوتز و همکاران (۲۰۰۳) رجوع کرد. البته لازم به توضیح است که پارامتر تنش مقاومت توسط نویسندگان دیگری برای توزیع‌های آماری متفاوتی محاسبه شده است. برای مثال کوندو و گوپتا (۲۰۰۵) پارامتر تنش مقاومت را برای توزیع نمایی تعمیم یافته بدست آوردند. همچنین کوندو و گوپتا (۲۰۰۶) پارامتر تنش مقاومت را برای توزیع وایبل مورد بررسی قرار دادند. همچنین کوندو و گوپتا (۲۰۰۹) این کار را برای توزیع وایبل سه پارامتری انجام دادند. لازم به توضیح است تا آنجا که نویسنده مقاله اطلاع دارد محاسبه پارامتر تنش مقاومت در توزیع لامبدای تعمیم یافته اولین بار در مقاله حاضر مورد بررسی قرار گرفته است.

۴ محاسبه پارامتر تنش مقاومت در توزیع لامبدای تعمیم یافته

در این بخش به محاسبه پارامتر تنش مقاومت در توزیع لامبدای تعمیم یافته پرداخته شده است. البته با توجه به وجود پارامترهای توزیع لامبدای تعمیم یافته محاسبه پارامتر تنش مقاومت در آن به راحتی انجام نمی شود، بنابراین از روش های عددی یا شبیه سازی استفاده می شود. همان طور که گفته شد فقط در شش ناحیه از فضای پارامتری مقدار تابع چگالی احتمال رابطه (۱،۲) وجود دارد که در این بخش فقط به محاسبه پارامتر تنش مقاومت در ناحیه چهار پرداخته شده است. فرض کنید متغیرهای تصادفی X و Y به ترتیب از توزیع های $GLD(\lambda_1, \lambda_2; \lambda_4)$ و $GLD(\lambda'_1, \lambda'_2; \lambda'_4)$ انتخاب می شوند. آنگاه طبق رابطه (۱،۳)

$$1 - P(X < Y) = P(X > Y) = \int_{-\infty}^{+\infty} P(X > Y | Y = y) f_Y(y) dy = \int_{-\infty}^{+\infty} (1 - F_X(y)) f_Y(y) dy, \quad (۴)$$

خواهد بود. با تغییر متغیر $t = F_X(y)$ خواهیم داشت:

$$1 - P(X < Y) = \int_0^1 (1 - t) \frac{f_Y(F_X^{-1}(t))}{f_X(F_X^{-1}(t))} dt. \quad (۵)$$

با توجه به این که در ناحیه ۴ فضای پارامتری توزیع لامبدای تعمیم یافته λ_1 ، شامل تمام مقادیر است، بنابراین بدون اینکه از کلیت مساله کم شود $\lambda_1 = 0$ اختیار می شود. به راحتی با کمی محاسبه مقادیر $F_X^{-1}(t)$ و $f_X(F_X^{-1}(t))$ در رابطه (۲،۴) برای توزیع $GLD(\lambda_1, \lambda_2; \lambda_4)$ عبارتند از:

$$F_X^{-1}(t) = \frac{1 - (1 - t)^{\lambda_2}}{\lambda_2}, \quad (۶)$$

$$f_X(F_X^{-1}(t)) = \frac{\lambda_2}{\lambda_4 (1 - t)^{\lambda_2 - 1}}. \quad (۷)$$

همچنین با توجه به روابط (۳,۴) و (۴,۴) و توزیع $GLD = (\lambda'_1, \lambda'_2; \lambda'_3)$ مقدار $f_Y(F_X^{-1}(t))$ عبارت است از:

$$f_Y\left(\frac{1 - (1-t)^{\lambda'_2}}{\lambda'_2}\right) = \frac{\lambda'_2}{\lambda'_3(1-t)^{\lambda'_2-1}}. \quad (۸)$$

با تغییر متغیر $z = \frac{1 - (1-t)^{\lambda'_2}}{\lambda'_2}$ و جایگذاری در رابطه (۵,۴)، خواهیم داشت:

$$f_Y(z) = \frac{\lambda'_2}{\lambda'_3(1 - \lambda'_2 z)} \frac{\lambda'_2 - 1}{\lambda'_2}, z > 0. \quad (۹)$$

با تغییر متغیر $z = F_X^{-1}(t)$ در رابطه (۶,۴)، خواهیم داشت:

$$f_Y(F_X^{-1}(t)) = \frac{\lambda'_2}{\lambda'_3} \left[1 - \frac{\lambda'_2}{\lambda'_3}(1 - (1-t)^{\lambda'_2})\right] \frac{1 - \lambda'_2}{\lambda'_2}. \quad (۱۰)$$

با قرار دادن روابط (۴,۴) و (۷,۴) در رابطه (۲,۴)، خواهیم داشت:

$$1 - P(X < Y) = \frac{\lambda_2}{\lambda'_2} \frac{\lambda'_2}{\lambda'_3} \int_0^1 (1-t)^{\lambda_2} \left[1 - \frac{\lambda'_2}{\lambda'_3}(1 - (1-t)^{\lambda'_2})\right] \frac{1 - \lambda'_2}{\lambda'_2} dt. \quad (۱۱)$$

مقدار پارامتر تنش مقاومت، $P(X < Y)$ ، تحت پارامترهای مختلف از توزیع های X و Y با استفاده از نرم افزار *SPLUS* محاسبه و در جدول ۲ آورده شده‌اند.

جدول ۸: مقادیر عددی پارامتر تنش مقاومت در توزیع لامبدای تعمیم یافته تحت پارامترهای مختلف

λ_1	λ_2	λ'_1	λ'_2	$P(X < Y)$
-۶	-۴	-۳	-۲	۰/۲۸
-۵	-۳	-۲	-۱	۰/۵۱
-۴	-۳	-۲	-۱	۰/۷۵
-۴	-۳	-۱	-۱	۰/۴۶
-۴	-۲	-۲	-۱	۰/۳۲
-۳	-۳	-۲	-۲	۰/۴۱
-۳	-۳	-۱	-۱	۰/۳۸
-۲	-۱	-۲	-۴	۰/۸
-۱	-۲	-۳	-۱	۰/۲

۵ مثال کاربردی

در این بخش با یک مثال با داده‌های واقعی کاربرد مطالب گفته شده در توزیع لامبدای تعمیم یافته نشان داده شده و پارامتر تنش مقاومت در آن محاسبه می‌شود. کاریان و دودیچ (۱۹۹۹) در یک بررسی علمی روی ۲۱ انسان با تحصیلات فوق لیسانس و دکتری در رشته‌های مختلف دانشگاهی، میزان ساعات مطالعه آن‌ها را در یک روز، همراه با اسکن از مغز آن‌ها برای تعیین میزان مقدار ماده خاکستری در مغزشان را اندازه‌گیری کردند. (لازم به توصیه است که میزان ماده خاکستری در مغز انسان‌ها با هم‌دیگر متفاوت است، بنابراین می‌تواند به عنوان یک متغیر تصادفی مورد استفاده قرار گیرد.) داده‌های جمع‌آوری شده در جدول‌های ۳ و ۴ آورده شده‌اند.

جدول ۹: میزان ساعات مطالعه افراد مورد مطالعه

۷	۶/۵	۷/۵	۷	۶/۵	۶	۵/۵
۷/۵	۵	۴/۵	۸	۸/۵	۴/۵	۶/۵
۵	۴	۳/۵	۴/۵	۷/۵	۷	۵

متغیرهای تصادفی X و Y به ترتیب میزان ساعت مطالعه افراد در روز و مقدار ماده خاکستری در مغز آن‌ها می‌باشد. کاریان و همکاران (۱۹۹۶) نشان دادند که متغیر تصادفی X دارای توزیع لامبدای تعمیم یافته، $(\lambda_1 = 0.05478, \lambda_2 = 0.6378, \lambda'_1 = 0.0637, \lambda'_2 = 0.112/133)$ که

جدول ۱۰: مقدار ماده خاکستری در مغز افراد مورد مطالعه

۱۰۸/۷	۱۰۷	۱۱۰/۳	۱۱۰	۱۱۳/۶	۹۹/۲	۱۰۴/۵
۱۰۸/۱	۱۰۷/۲	۱۱۲	۱۱۵/۵	۱۰۸/۴	۱۱۳/۴	۱۰۱/۲
۹۸/۴	۱۰۰/۹	۱۰۷/۱	۱۰۷/۱	۱۰۸/۷	۱۰۲/۵	۱۰۳/۳

برای آن $\lambda_1 = 112/133$ ، $\lambda_2 = 0/0637$ ، $\lambda_3 = 0/6378$ ، $\lambda_4 = 0/5478$ و الجزار (۲۰۰۵) نشان دادند که متغیر تصادفی Y دارای توزیع لامبدای تعمیم یافته $GLD = (102/899, 0/0806, 0/1475, 0/8041)$ با $\lambda_1' = 102/899$ ، $\lambda_2' = 0/0806$ ، $\lambda_3' = 0/1475$ ، $\lambda_4' = 0/8041$ با کمی محاسبه به راحتی دیده می شود که مقدار پارامتر تنش مقاومت، $P(X < Y)$ ، مطابق رابطه (۸،۴) تحت پارامترهای توزیع های X و Y ، عدد $0/67$ می باشد.

۶ بحث و نتیجه گیری

در این مقاله توزیع لامبدای تعمیم یافته، ویژگی های آن و مفهوم تنش مقاومت بطور کامل توضیح داده شد و قابلیت اعتماد یک سیستم از دیدگاه تنش مقاومت مورد مطالعه قرار گرفت. همچنین فرم ریاضی پارامتر تنش مقاومت در توزیع لامبدای تعمیم یافته محاسبه و به ازای مقادیر مختلف پارامترهای توزیع لامبدای تعمیم یافته، پارامتر تنش مقاومت محاسبه شد. با یک مثال واقعی مربوط به میزان ساعات مطالعه روزانه ۲۱ انسان با تحصیلات بالا برای رشته های مختلف دانشگاهی و میزان مقدار ماده خاکستری در مغز آنها، با توجه به برآزش توزیع لامبدای تعمیم یافته روی متغیرهای مورد نظر، مقدار پارامتر تنش مقاومت، $P(X < Y)$ ، محاسبه شد.

مراجع

- [1] Aljazar, L. (2005), *Generalized Lambda Distribution and Estimation Parameters*, M. S. thesis, Islamic University of Gaza.

- [2] Beckwith N. B. and Dudewicz, E. J. (1996), *A Bivariate Generalized Lambda Distribution (GLD-2) using Plackett's Method of Construction*, Distribution, Examples and Applications, American Journal of Mathematical and Management Sciences, 16(3-4), 333-393.
- [3] Bergevin, R. J. (1993), *An Analysis of the Generalized Lambda Distribution*, Report number, A584262.
- [4] Bigerelle, M., Najjar, D. Fournier, B., Rupin, N. and Iost, A. (2006), *Application of lambda distribution and bootstrap analysis to the prediction of fatigue lifetime and confidence intervals*, International Journal of Fatigue, 28(3), 223-236.
- [5] Chalabi, Y., Scott, D. J. and Wuertz, D. (2012), *Flexible Distribution Modeling with the Generalized Lambda Distribution*, Munich Personal Repec Archive.
- [6] Corrado, C. J. (2001), *Option pricing on the generalized lambda distribution*, Journal of Futures Markets, 21(3), 213-236.
- [7] Dengiz, B. (1988), *The generalized lambda distribution in simulation of M/M/1 queue systems*, Journal of the Faculty of Engineering and Architecture of Gazi University, 3, 161-171.
- [8] Filliben, J. J. (1975), *The probability plot correlation test for normality*, Technometrics, 17(1), 111-117.
- [9] Freimer, M. Kollia, G. Mudholkar, G. and Lin, C. (1998), *A study of the generalized Tukey lambda*, Communications in Statistics-Theory and Methods, 17(10), 3547-3567.
- [10] Joiner, B. L and Rosenblatt, J. R. (1971), *Some properties of the range in samples Tukey's symmetric lambda distributions*, Journal of the American Statistical Association, 66(334), 394-399.
- [11] Karian, Z. and Dudevicz, E. (1999), *Modern Statistical Systems, and GPSS Simulation*, Second Edition, CRC Press LLC.
- [12] Karian, Z. and Dudevicz, E. (2000), *Fitting Statistical Distributions, The Generalized Lambda Distribution and Generalized Bootstrap Methods*, Chapman and Hall, New york.
- [13] Karvanen, J., Eriksson, J. and Koivunen, V. (2000), *Maximum likelihood estimation of ICA model for wide class of source distributions*, In Neural Networks for Signal Processing X, Processing of the 2000 IEEE Signal Processing Society Workshop, 1, 445-454.

- [14] King, R. and MacGillivray, H. (2007), *Fitting the generalized lambda distribution with location and scale-free shape functionals*, American journal of Mathematical and Management Sciences, 27(3-4), 441-460.
- [15] Kotz, S., Lumelskii, Y. and Pensky, M. (2003), *The Stress-Strength Model and its Generalization*, Theory and Application, New York: World scientific.
- [16] Ozturk, A. and Dale, R. (1982), *A study of fitting the generalized lambda distribution to solar-radiation data*, Journal of Applied Meteorology, 21(7), 995-1004.
- [17] Ramberg, J. S. and Schmeiser, B. W. (1974), *An approximate method for generating asymmetric random variables*, Commun, ACM, 17(2), 78-82.
- [18] Ramberg, J. S., Tadikamalla, P. R., Dudewicz, E. J. and Mytytha, E. F. (1979), *A probability distribution and its uses in fitting data*, Technometrics, 21(2), 201-214.
- [19] Riser, T. L. (1980), *Accounting for non-Normality and sampling Error in analysis of variance of construction data*, M. S. thesis, Department of civil Engineering, The Ohio State University, Columbus, Ohio.
- [20] Su, S. (2006), *Maximum Log Likelihood Estimation Using EM Algorithm and Partition Maximum Log Likelihood Estimation for Mixture of Generalized Lambda Distributions*, Working paper.
- [21] Su, S. (2007a), *Numerical maximum log likelihood estimation for generalized lambda distributions*, Computational Statistics and Data Analysis, 51, 3983-3998.
- [22] Su, S. (2007b), GLDEX, *Fitting Signal and Mixture of Generalized Lambda Distributions (RS and FMKL) Using Discretized and Maximum Likelihood Methods*, R package version 1.0.1, URL <http://CRAN.R-project.org/>

مروری بر پارامتر تنش-مقاومت در سیستم های منسجم تحت توزیع نرمال چوله

حسنعلی پور، پ^۱ رزمخواه، م^۲

گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده

در برخی از آزمایش ها بدنبال ارزیابی میزان تحمل مولفه مورد آزمایش در برابر فشار سایر متغیرها می باشیم، چنین حالتی را در مباحث قابلیت اعتماد یک مدل تنش-مقاومت گویند. در ساده ترین وضعیت یک مدل تنش-مقاومت اگر مقاومت یک مولفه از این مدل بیشتر از فشار وارد بر آن باشد، آن مولفه به فعالیت خود ادامه می دهد. قابلیت اعتماد یک مولفه از این سیستم به وسیله $R = P(X < Y)$ تعریف می شود که در آن Y مقاومت تصادفی آن مولفه و X استرس تصادفی وارد بر آن مولفه می باشد. در این مقاله فرض بر این است که تنش و مقاومت هر دو در قالب سیستم های منسجم ساماندهی شده اند بطوری که مولفه های هر یک دارای توزیع نرمال چوله هستند و علاقه مند بررسی پارامتر میزان قابلیت اعتماد $P(X_{r:n} < Y_{k:m})$ در حالت های خاص و به ازای مقادیر مختلف پارامترهای چوله هستیم.

کلمات کلیدی: توزیع نرمال چوله، آماره های ترتیبی، قابلیت اعتماد تنش-مقاومت.

^۱ hasanalipour@mail.um.ac.ir

^۲ razmkhah-m@um.ac.ir

۱ مقدمه

از معروف ترین سیستم ها در مهندسی قابلیت اعتماد می توان از سیستم هایی با ساختار سری، موازی و k از r نام برد.

سیستم های سری: ساده ترین نوع سیستم ها در مهندسی قابلیت اعتماد سیستم سری است. یک سیستم را سری گویند هرگاه فعال بودن سیستم مستلزم فعال بودن همه اجزای آن باشد. به عبارت دیگر، سیستم فعال است اگر و تنها اگر همه اجزای آن فعال باشند. اگر $X_1, \dots, X_n$ متناظر با طول عمر مولفه ها باشند، یعنی X_i طول عمر مولفه i

ام باشد، آنگاه طول عمر یک سیستم سری را با T_s نشان داده و عبارت است از $T_s = \min\{X_1, \dots, X_n\}$

سیستم های موازی: بر اساس تعریف، سیستم موازی سیستمی است که فعال بودن آن مستلزم فعال بودن حداقل یکی از اجزای آن باشد. به عبارت دیگر، سیستم هنگامی از کار می افتد که تمام اجزای آن غیر فعال باشد. طول عمر یک سیستم موازی در ارتباط با طول عمر مولفه یا ماکسیم مقدار است. اگر X_i طول عمر مولفه i ام باشد

آنگاه در یک سیستم موازی با n مولفه طول عمر را با T_p نشان داده و عبارت است از $T_p = \max\{X_1, \dots, X_n\}$.

سیستم های k از n : یک سیستم k از n تعمیمی از سیستم های سری و موازی است. یک سیستم شامل n جزء را k از n گوئیم هرگاه فعال بودن آن مستلزم فعال بودن حداقل k ($k \leq n$) جزء از n جزء آن باشد. با توجه به تعریف سیستم های سری و موازی، مشاهده می شود که سیستم سری یک سیستم n از n و سیستم موازی یک سیستم 1 از n است. اگر X_i نشان دهنده طول عمر مولفه i ام باشد آنگاه برای یک سیستم k از n ، طول عمر را با $T_{k,n}$ نشان می دهیم و عبارت است از $T_{k,n} = X_{n-k+1:n}$. منظور از نماد $X_{r:n}$ ، r امین آماره مرتب در نمونه ای به اندازه n است. تابع چگالی و توزیع احتمال $X_{i:n}$ در یک نمونه تصادفی از جامعه ای پیوسته با تابع توزیع تجمعی $F_X(x)$ و تابع چگالی $f_X(x)$ عبارات است از:

$$f_{X_{i:n}}(x) = \frac{n!}{(i-1)!(n-i)!} F^{i-1}(x) [1 - F(x)]^{n-i} f(x)$$

و

$$F_{X_{i:n}}(x) = \sum_{j=i}^n \frac{n!}{(j-1)!(n-j)!} F^j(x) [1 - F(x)]^{n-j}$$

امروزه نقش مهم آماره های مرتب آنچنان پررنگ شده که بسیاری از مباحث گوناگون آمار را به خود اختصاص داده است. آماره های مرتب در قسمت های مختلف توصیف و استنباط آماری دارای کاربرد است. از جمله می توان به نقش آماره های مرتب در قضایای مربوط به طول عمر، توصیف و تشخیص توزیعها، تشخیص داده های پرت، آزمون نیکویی برازش، تجزیه و تحلیل نمونه های سانسور شده، قابلیت اعتماد، کنترل کیفیت و برخی مسائل مشابه دیگر اشاره نمود. برای جزئیات بیشتر در مورد آماره های مرتب و کاربردهای آن می توان به کتاب های آرنولد و همکاران (۲۰۰۸) و یا دیوید و ناگاراچا (۲۰۰۳) مراجعه نمود.

مدل تنش-مقاومت چیست؟ در ساده ترین حالت میزان قابلیت اعتماد یک مولفه که آن را با متغیر تصادفی Y نشان می دهیم و تحت فشاری است که آن را با متغیر تصادفی X نشان می دهیم، میزان مقاومت آن واحد است که بر این فشار غلبه می کند. در این صورت اگر مولفه مورد نظر در مقابل این فشار مقاومت کند، یعنی $(Y > X)$ باشد، مولفه پیروز می شود. در غیر اینصورت اگر $(X > Y)$ باشد مولفه مورد نظر در مقابل فشار وارده شکست می خورد. پس میزان قابلیت اعتماد بر هر دو مبنای احتمال شکست و پیروزی تعریف می شود. مثلاً قابلیت اعتماد تعریف شده بر مبنای احتمال پیروزی عبارت است از $P(X < Y)$. در بسیاری از کارها در این زمینه فرض شده که متغیرهای تصادفی تنش-مقاومت هر دو از توزیع های مشابهی پیروی می کنند، مانند نرمال، نمایی، لگ نرمال، وایبل و غیره. اکثر مولفین فرض کرده اند که متغیرهای تصادفی تنش-مقاومت از هم مستقل باشند. با توجه به تعریف قابلیت اعتماد در یک مدل تنش-مقاومت، علاقه مندیم که در کاربردها مقدار قابلیت اعتماد تا حد امکان نزدیک به یک باشد تا طول عمر مفید وسیله افزایش یابد.

پژوهش های بسیاری از سال ۲۰۰۳ تاکنون روی پارامتر قابلیت اعتماد تنش-مقاومت تحت انواع مختلف توزیع های تنش-مقاومت با فرض استقلال و عدم استقلال مولفه های مدل و تحت انواع گوناگون داده ها انجام شده و در حال انجام است. یک مرور شفاف از عناوین مختلف روی مدل های تنش-مقاومت در ساراگوگو و کایا (۲۰۰۷)، کریشنامورتی و همکاران (۲۰۰۷)، کوندا و رکب (۲۰۰۹) و رضایی و همکاران (۲۰۱۰) آماده شده است.

توزیع نرمال چوله: توزیع نرمال چوله یکی از مهمترین توزیع های چوله است که با داشتن پارامتری، چولگی را تنظیم می کند. این توزیع نخستین بار توسط آزالینی (۱۹۸۵) ارائه شد. این خانواده از توزیع ها شامل توزیع نرمال است و دارای خواص متعددی است که با خواص توزیع نرمال همخوانی دارد و یا نزدیک به خواص توزیع نرمال است. این توزیع، به دلیل خواص مطلوب و شباهت آن به توزیع نرمال، در دهه های اخیر مورد توجه محققان قرار

گرفته است.

تعریف: متغیر تصادفی دارای توزیع نرمال چوله با پارامتر است و آن را با نماد نشان می دهند، هرگاه تابع چگالی احتمال آن بصورت زیر باشد.

$$\phi(x; \lambda) = \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{\lambda(x-\mu)}{\sigma}\right); \quad -\infty < x < \infty, \quad \lambda \in R, \quad (1)$$

که در آن $\phi(\cdot)$ و $\Phi(\cdot)$ به ترتیب تابع چگالی و تابع توزیع نرمال استاندارد می باشند. برخی ویژگیهای توزیع نرمال چوله بصورت زیر بیان می شود:

الف) هرگاه $\lambda = 0$ آنگاه چگالی نرمال چوله به چگالی نرمال استاندارد تبدیل می شود.

ب) اگر $X \sim SN(\lambda)$ باشد، آنگاه $-X \sim SN(-\lambda)$ است.

ج) اگر $\lambda \rightarrow \infty$ و $x > 0$ یا $\lambda \rightarrow -\infty$ و $x < 0$ ، چگالی نرمال چوله به چگالی نیم نرمال میل میکند. تابع توزیع نرمال چوله را می توان بصورت زیر نوشت.

$$\Phi(z; \lambda) = \Phi(z) - \frac{1}{\lambda} T(z; \lambda),$$

که در آن

$$T(x; \lambda) = \int_x^\infty \int_0^{\lambda t} \phi(u) \phi(t) du dt.$$

۲ مروری بر پارامتر تنش-مقاومت در سیستم های منسجم تحت توزیع نرمال چوله

همیشه پارامتر قابلیت اعتماد در یک سیستم به صورت $P(X < Y)$ نمی باشد. گاهی اوقات علاقه مند به میزان قابلیت اعتماد در سیستم هایی با پارامتر $P(X_{r:n} < Y_{k:m})$ هستیم که در آن $X_{r:n}$ و $Y_{k:m}$ به ترتیب مولفه های مرتب r ام و k ام در متغیرهای تنش-مقاومت می باشند. این پارامتر ما را قادر می سازد تا احتمال بزرگتر بودن $Y_{k:m}$ از $X_{r:n}$ را بدست آوریم و تحت مقاومت ها و تنش های مرتب شده قابلیت اعتماد یک سیستم را ارزیابی نماییم. در این بخش حالتی را در نظر می گیریم که متغیرهای تصادفی تنش-مقاومت هر دو از توزیع نرمال چوله با پارامترهای مختلف و مستقل از هم باشند.

توصیف مدل: فرض کنید X و Y متغیرهای تصادفی نرمال چوله به ترتیب با توابع چگالی احتمال

$$f_X(x; \lambda_1) = \varphi(x) \Phi(\lambda_1 x), \quad x \in R, \quad \lambda_1 \in R$$

و

$$f_Y(y; \lambda_2) = \varphi(y) \Phi(\lambda_2 y), \quad y \in R, \quad \lambda_2 \in R$$

باشند. در این مدل m مولفه تصادفی مقاومت تحت فشار n مولفه تصادفی تنش می باشد. از طرفی $X_1, \dots, X_n$ و $Y_1, \dots, Y_m$ از هم مستقل می باشند. تحت فرضیات فوق مقدار $P(X_{r:n} < Y_{k:m})$ را با $R_{r,k}$ نشان داده و بصورت زیر محاسبه می کنیم.

$$\begin{aligned} R_{r,k} &= P(X_{r:n} < Y_{k:m}) \\ &= \int_{-\infty}^{\infty} F_{X_{r:n}}(y) f_{Y_{k:m}}(y) dy \end{aligned}$$

که در آن $F_{X_{r:n}}(y)$ و $f_{Y_{k:m}}(y)$ به ترتیب توابع توزیع و چگالی احتمال r و k آماره مرتب از متغیر تصادفی X با اندازه نمونه n و متغیر تصادفی Y با اندازه نمونه m می باشند. در توزیع نرمال چوله برای مقادیر بزرگ مثبت (یا منفی) پارامتر λ ، آماره مرتب $X_{1:n}$ (یا $X_{n:n}$) پراطلاح ترین آماره مرتب است. بر اساس این آماره های پراطلاح حالت های مختلف مولفه های تصادفی تنش و مقاومت در یک سیستم تنش-مقاومت را در نظر می گیریم و یک عبارت ساده برای قابلیت اعتماد تنش-مقاومت ارائه می دهیم.

حالت اول: اگر $\lambda_1 > 0$ و $\lambda_2 > 0$ کمترین میزان مولفه تصادفی مقاومت در مقابل کمترین میزان مولفه تصادفی فشار در سیستم قرار می گیرد. بنابراین قابلیت اعتماد $R_{1,1} = P(X_{1:n} < Y_{1:m})$ را در نظر می گیریم.

$$\begin{aligned} R_{1,1} &= P(X_{1:n} < Y_{1:m}) \\ &= \int_{-\infty}^{\infty} F_{X_{1:n}}(y) f_{Y_{1:m}}(y) dy \\ &= \int_{-\infty}^{\infty} \left[1 - (1 - \Phi(y; \lambda_1))^n \right] m \phi(y; \lambda_2) (1 - \Phi(y; \lambda_2))^{m-1} dy \end{aligned}$$

که در آن $\phi(y; \lambda)$ و $\Phi(y; \lambda)$ تابع چگالی و تابع توزیع نرمال چوله می باشند.

حالت دوم: اگر $\lambda_1 > 0$ و $\lambda_2 < 0$ بیشترین میزان مولفه تصادفی مقاومت در مقابل کمترین میزان مولفه تصادفی

فشار در سیستم قرار می گیرد. بنابراین قابلیت اعتماد $R_{\lambda:m} = P(X_{\lambda:n} < Y_{m:m})$ را بدست می آوریم.

$$\begin{aligned} R_{\lambda:m} &= P(X_{\lambda:n} < Y_{m:m}) \\ &= \int_{-\infty}^{\infty} F_{X_{\lambda:n}}(y) f_{Y_{m:m}}(y) dy \\ &= \int_{-\infty}^{\infty} [1 - (1 - \Phi(y; \lambda_1))^n] m \phi(y; \lambda_2) [\Phi(y; \lambda_2)]^{m-1} dy \end{aligned}$$

حالت سوم: اگر $\lambda_1 < 0$ و $\lambda_2 < 0$ بیشترین میزان مولفه تصادفی مقاومت در مقابل بیشترین میزان مولفه تصادفی فشار قرار می گیرد. بنابراین $R_{n:m} = P(X_{n:n} < Y_{m:m})$ قابلیت اعتماد یک سیستم موازی را نشان می دهد.

$$\begin{aligned} R_{n:m} &= P(X_{n:n} < Y_{m:m}) \\ &= \int_{-\infty}^{\infty} F_{X_{n:n}}(y) f_{Y_{m:m}}(y) dy \\ &= \int_{-\infty}^{\infty} [\Phi(y; \lambda_1)]^n m \phi(y; \lambda_2) [\Phi(y; \lambda_2)]^{m-1} dy \end{aligned}$$

حالت چهارم: اگر $\lambda_1 < 0$ و $\lambda_2 > 0$ کمترین میزان مولفه تصادفی مقاومت در مقابل بیشترین میزان مولفه تصادفی تنش قرار می گیرد. در این صورت $R_{n:\lambda} = P(X_{n:n} < Y_{\lambda:m})$ قابلیت اعتماد یک سیستم سری را نشان می دهد و داریم:

$$\begin{aligned} R_{n:\lambda} &= P(X_{n:n} < Y_{\lambda:m}) \\ &= \int_{-\infty}^{\infty} F_{X_{n:n}}(y) f_{Y_{\lambda:m}}(y) dy \\ &= \int_{-\infty}^{\infty} [\Phi(y; \lambda_1)]^n m \phi(y; \lambda_2) [1 - \Phi(y; \lambda_2)]^{m-1} dy \end{aligned}$$

از آنجاییکه انتگرال های بدست آمده در حالت های چهار گانه ذکر شده پیچیده است، با استفاده از روشهای عددی برای مقادیر مختلف λ_1 و λ_2 عبارت $R_{r,k}$ را بدست می آوریم.

۳ محاسبات عددی

ابتدا پارامتر λ_2 را ثابت نگه می داریم و مقادیر قابلیت اعتماد را به ازای λ_1 های مختلف بدست می آوریم. این مقادیر در جدول ۱ و ۲ گزارش شده است. سپس با فرض λ_1 ثابت، مقادیر قابلیت اعتماد را به ازای λ_2 های مختلف محاسبه می کنیم. جدول ۳ و ۴ نشان دهنده این مقادیر بدست آمده است.

جدول ۱. مقادیر قابلیت اعتماد وقتی $\lambda_1 > 0$ و $\lambda_2 = 1, -1$ ثابت باشد

مقادیر قابلیت اعتماد	$\lambda_1 = 0.1$	۰.۵	۱	۲	۳	۵
$R_{1:1}$	۰.۷۸۳	۰.۶۳۴	۰.۴۱۶	۰.۲۹۵	۰.۲۰۹	۰.۱۰۰
$R_{1:m}$	۰.۹۴۶	۰.۸۶۹	۰.۷۰۴	۰.۵۲۸۷	۰.۴۵۸	۰.۳۰۰

جدول ۲. مقادیر قابلیت اعتماد وقتی $\lambda_1 < 0$ و $\lambda_2 = 1, -1$ ثابت باشد

مقادیر قابلیت اعتماد	$\lambda_1 = -0.1$	-۰.۵	-۱	-۲	-۳	-۵
$R_{n:m}$	۰.۲۱۶	۰.۳۶۵	۰.۵۸۳	۰.۹۰۴	۰.۹۹۰	۰.۹۹۹
$R_{n:1}$	۰.۰۵۳	۰.۱۳۰	۰.۲۹۵	۰.۷۱۲	۰.۹۴۱	۰.۹۹۹

جدول ۳. مقادیر قابلیت اعتماد وقتی $\lambda_1 > 0$ و $\lambda_2 = 1, -1$ ثابت باشد

مقادیر قابلیت اعتماد	$\lambda_2 = 0.1$	۰.۵	۱	۲	۳	۵
$R_{1:1}$	۰.۱۱۵	۰.۲۲۳	۰.۴۱۶	۰.۸۱۴	۰.۹۷۵	۰.۹۹۹
$R_{n:1}$	۰.۰۵۳	۰.۱۳۰	۰.۲۹۵	۰.۷۱۲	۰.۹۴۱	۰.۹۹۹

جدول ۴. مقادیر قابلیت اعتماد وقتی $\lambda_1 < 0$ و $\lambda_2 = 1, -1$ ثابت باشد

مقادیر قابلیت اعتماد	$\lambda_2 = -0.1$	-۰.۵	-۱	-۲	-۳	-۵
$R_{1:m}$	۰.۹۴۶	۰.۸۶۹	۰.۷۰۴	۰.۵۲۸۷	۰.۴۵۸	۰.۳۰۰
$R_{n:m}$	۰.۸۸۴	۰.۷۷۶	۰.۵۸۳	۰.۱۸۵	۰.۲۲۴	۰.۱۰۰

با توجه به جداول ۱ تا ۴ نتایج بصورت زیر گزارش می شود.

۱. با افزایش $\lambda_1 > 0$ (λ_2 ثابت)، $R_{1:1}$ کاهش می یابد و برای λ_1 های بزرگ مقدار قابلیت اعتماد صفر می شود ولی با افزایش $\lambda_2 > 0$ (λ_1 ثابت)، $R_{1:1}$ افزایش می یابد و برای λ_2 های بزرگ این مقدار به یک نزدیک می شود.
۲. با افزایش $\lambda_1 > 0$ (λ_2 ثابت)، $R_{1:m}$ کاهش می یابد و به ازای λ_1 های بزرگ این مقدار به صفر می رسد. در ضمن مقادیر $R_{1:m}$ برای حالتی که $\lambda_2 < 0$ (λ_1 ثابت) افزایش می یابد، یکسان بدست می آید.
۳. با افزایش $\lambda_1 < 0$ (λ_2 ثابت)، $R_{n:m}$ افزایش می یابد و برای λ_1 های بزرگ این مقدار به یک نزدیک می شود.

ولی با افزایش $\lambda_2 < 0$ (λ_1 ثابت)، $R_{n:m}$ کاهش می یابد و برای λ_2 های بزرگ این مقدار به صفر می رسد.
 ۴. با افزایش $\lambda_1 < 0$ (λ_2 ثابت)، $R_{n:1}$ افزایش می یابد و برای λ_1 های بزرگ این مقدار به یک می رسد. از طرفی این مقادیر بدست آمده دقیقا با مقادیر $R_{n:1}$ در حالتی که $\lambda_2 > 0$ باشد (λ_1 ثابت)، نیز مطابقت دارد.

۴ جمع بندی

بطور کلی می توان نتیجه گرفت که با افزایش λ_1 (λ_2 ثابت) در حالت های ۳ و ۴ و با افزایش λ_2 (λ_1 ثابت) در حالت های ۱ و ۴ مقدار قابلیت اعتماد به یک نزدیک می شود و در این حالت ها می توان گفت که طول عمر مفید وسیله افزایش می یابد. همچنین با افزایش λ_1 (λ_2 ثابت) در حالت های ۱ و ۲ و با افزایش λ_2 (λ_1 ثابت) در حالت های ۲ و ۳ مقدار قابلیت اعتماد به صفر می رسد به عبارتی می توان گفت که طول عمر مفید وسیله در این حالت ها کاهش می یابد.

مراجع

- [1] Azzalini, A. (1985), *A class of distributions which includes the normal ones*, Scandinavian Journal of Statistics. 12, 171-178.
- [2] Arnold, B.C., Balakrishnan, N. and Nagaraja, H.N. (2008), *A First Course in Order Statistics*, SIAM, Philadelphia.
- [3] David, H.A. and Nagaraja, H.N. (2003), *Order Statistics*, 3rd edition John Wiley, New York.
- [4] Eryilmaz, S. (2008), *Consecutive k-out-of n: G system in stress-strength setup*, Commun. Statist. Simul. Comput. 37, 579-589.
- [5] Eryilmaz, S. (2010), *On system reliability in stress-strength setup*, Statist. Probab. Letters. 80, 834-839.
- [6] Krishnamoorthy, K., Mukherjee, S. and Guo, H. (2007), *Inference on reliability in two parameter exponential stress-strength model*, Metrica. 65, 261-273.

- [7] Kunda, D. and Raqab, M.Z. (2009), *Estimation of $R = P(X < Y)$ for three-parameter weibull distribution* , Statist. Probab. Lettars. 70, 1839-1846.
- [8] Rezaei, S., Tahmasbi ,R. and Mahmoodi, M. (2010), *Estimation of $P(X < Y)$ for generalized pareto distribution*, Statist. Plam. Inference. 140, 480-494.
- [9] Saracoglu, B. and Kaya, M.F. (2007), *Maximum likelihood estimation and confidence intervals of system reliability for Gomperts distribution in stress-strength models.* , Selcuk Journal Application Mathematics . 8, 25-36.

برآورد قابلیت اطمینان تنش-مقاومت چند مؤلفه ای در توزیع رایلی

رحیمی ده چراغی،^۱ طاهری،^۲ س

^۱ فارغ التحصیل کارشناسی ارشد، دانشگاه حکیم سبزواری

^۲ فارغ التحصیل کارشناسی ارشد، دانشگاه پیام نور شیراز

چکیده

در یک سیستم چند مؤلفه ای، متغیرهای تصادفی مستقل و هم توزیع $x_1, x_2, \dots, x_k$ را k مؤلفه ی مقاومت و هر متغیر تصادفی تجربی Y را تنش در نظر می گیریم. این سیستم در صورتی به فعالیت خود ادامه می دهد که حداقل s خارج از k ($s < k$) باشد یعنی مقاومت ها از تنش بزرگتر باشند. قابلیت اعتماد این سیستم هنگامی که متغیرهای تصادفی تنش و مقاومت هر دو از توزیع رایلی تعمیم یافته با پارامترهای شکل مختلف بودند محاسبه شد. با استفاده از روش درستنمایی ماکزیمم برآورد قابلیت اعتماد بدست آمد و به صورت مجانبی بررسی شد. سپس از شبیه سازی مونت کارلو برای مقایسه ی برآورد قابلیت اعتماد در نمونه های کوچک و مجموعه داده های واقعی استفاده شد.

کلمات کلیدی: توزیع رایلی تعمیم یافته، برآورد قابلیت اعتماد، تنش-مقاومت، برآوردگر درستنمایی ماکزیمم، فاصله اطمینان.

^۱orahimi55@gmail.com

^۲taherisamira22@gmail.com

۱ مقدمه

توزیع دو پارامتری بر نوع ده^۳ توسط [10] و [9] معرفی و به عنوان توزیع رایلی تعمیم یافته نامگذاری شد. توزیع دو پارامتری بر نوع ده به عنوان توزیع رایلی تعمیم یافته (GRD) نامیده می شود. اگر $\alpha > 0$ و $\lambda > 0$ باشد تابع چگالی احتمال GRD دو پارامتری به صورت

$$f(x; \alpha, \lambda) = \alpha \lambda^\alpha x e^{-(\lambda x)^\alpha} (1 - e^{-(\lambda x)^\alpha})^{\alpha-1} \quad x > 0. \quad (1)$$

و تابع توزیع آن به صورت زیر است

$$F(x; \alpha, \lambda) = (1 - e^{-(\lambda x)^\alpha})^\alpha \quad x > 0. \quad (2)$$

که $\alpha > 0$ و $\lambda > 0$ به ترتیب پارامترهای شکل و مقیاس هستند. این توزیع به طور گسترده توسط [5] و [8] مورد مطالعه قرار گرفته است. GRD دو پارامتری با $GR(\alpha, \lambda)$ نشان داده می شود که در مدل های مقاومت و داده های جامع طول عمر کاربرد دارد. (مراجعه کنید به [10])

در این مقاله که مروری بر کارهای [7] می باشد به بررسی قابلیت اعتماد تنش مقاومت چند مؤلفه ای در توزیع رایلی تعمیم یافته می پردازیم که در آن X و Y دو متغیر تصادفی مستقل، هر دو از توزیع رایلی تعمیم یافته به ترتیب با پارامترهای شکل α و β و پارامتر مقیاس مشترک λ هستند. فرض کنید $y, x_1, x_2, \dots, x_k$ نمونه های تصادفی مستقلی باشند که در آن Y از توزیع پیوسته $G(y)$ و $x_1, x_2, \dots, x_k$ از توزیع پیوسته $F(x)$ پیروی کنند. قابلیت اعتماد مدل تنش - مقاومت که توسط [2] گسترش پیدا کرد به صورت زیر است

$$R_{s,k} = P(\text{حداقل } s \text{ از } x_1, x_2, \dots, x_k \text{ بزرگتر از } Y \text{ باشد}) \\ = \sum_{i=s}^k \binom{k}{i} \int_{-\infty}^{\infty} [1 - G(y)]^i [G(y)]^{k-i} dF(y), \quad (3)$$

زمانی که متغیرهای تصادفی مقاومت $x_1, x_2, \dots, x_k$ مستقل و هم توزیع، از توزیع $F(x)$ در معرض متغیر تصادفی تنش Y باشند، احتمال در رابطه (3) را قابلیت اعتماد مدل تنش - مقاومت چند مؤلفه ای می نامند. (مراجعه کنید به [2])

^۳Burr Type X

۲ برآوردگر درست‌نمایی ماکزیمم $R_{s,k}$

فرض کنید $X \sim GR(\alpha, \lambda)$ و $Y \sim GR(\beta, \lambda)$ با پارامترهای شکل نامعلوم α و β و پارامتر مقیاس مشترک λ هستند که در آن X و Y به صورت مستقل توزیع شده باشند. قابلیت اعتماد تنش مقاومت چند مؤلفه ای در توزیع ریلی تعمیم یافته با استفاده از رابطه ی (3) به صورت زیر محاسبه می شود.

$$\begin{aligned} R_{s,k} &= \sum_{i=s}^k \binom{k}{i} \int_0^\infty \left[1 - \left(1 - e^{-(\lambda y)^\gamma} \right)^\beta \right]^i \left[\left(1 - e^{-(\lambda y)^\gamma} \right)^\beta \right]^{k-i} \\ &\quad \times \gamma \alpha \lambda^\gamma y e^{-(\lambda y)^\gamma} \left(1 - e^{-(\lambda y)^\gamma} \right)^{\alpha-1} dy \\ &= \sum_{i=s}^k \binom{k}{i} \int_0^1 (1-t^\beta)^i (t^\beta)^{k-i} \alpha t^{\alpha-1} dt \\ &= \sum_{i=s}^k \binom{k}{i} v \int_0^1 z^{k-i+v-1} (1-z)^i dz \\ &= v \sum_{i=s}^k \binom{k}{i} \beta (k+v-i, i+1). \end{aligned}$$

در رابطه ی فوق از تغییر متغیرهای $t = 1 - e^{-(\lambda y)^\gamma}$ و سپس $z = t^\beta$ و جایگذاری $v = \frac{\alpha}{\beta}$ استفاده شد. پس از ساده کردن می نویسیم

$$R_{s,k} = v \sum_{i=s}^k \frac{k!}{(k-i)!} \left[\prod_{j=0}^i (k+v-j) \right]^{-1} \quad (۴)$$

اگر α و β نامعلوم باشند لازم است برای برآورد $R_{s,k}$ ، α و β را برآورد کنیم. می دانیم که روش ماکزیمم درست‌نمایی دارای خاصیت پایایی است. فرض کنید $x_1 < x_2 < \dots < x_n$ که $X \sim GRD(\alpha, \lambda)$ یک نمونه تصادفی مرتب شده از متغیرهای تصادفی مقاومت X به حجم n باشند و به طور مشابه فرض کنید $y_1 < y_2 < \dots < y_m$ که $Y \sim GRD(\beta, \lambda)$ یک نمونه تصادفی مرتب شده از متغیرهای تصادفی مقاومت Y به حجم m باشند. تابع درست‌نمایی این دو نمونه به صورت زیر است

$$\begin{aligned} L(\alpha, \beta, \lambda) &= (n+m) \ln \gamma + n \ln \alpha + m \ln \beta + \gamma (n+m) \ln \lambda \\ &\quad - \lambda^\gamma \left[\sum_{i=1}^n X_i^\gamma + \sum_{j=1}^m y_j^\gamma \right] + \sum_{i=1}^n \ln x_i + \sum_{j=1}^m \ln y_j \\ &\quad + (\alpha-1) \sum_{i=1}^n \ln \left(1 - e^{-(\lambda x_i)^\gamma} \right) + (\beta-1) \sum_{j=1}^m \ln \left(1 - e^{-(\lambda y_j)^\gamma} \right) \end{aligned} \quad (۵)$$

اگر MLE برای α و β و λ را به ترتیب با $\hat{\alpha}$ و $\hat{\beta}$ و $\hat{\lambda}$ نشان دهیم

$$\frac{\partial L}{\partial \alpha} = 0 \Rightarrow \frac{n}{\alpha} + \sum_{i=1}^n \ln \left(1 - e^{-(\lambda x_i)^r} \right) = 0 \quad (6)$$

$$\frac{\partial L}{\partial \beta} = 0 \Rightarrow \frac{m}{\beta} + \sum_{j=1}^m \ln \left(1 - e^{-(\lambda y_j)^r} \right) = 0 \quad (7)$$

$$\begin{aligned} \frac{\partial L}{\partial \lambda} = 0 \Rightarrow & \frac{r(m+n)}{\lambda} - r \lambda \left[\sum_{i=1}^n X_i^r - \sum_{j=1}^m y_j^r \right] + (\alpha - 1) \sum_{i=1}^n \frac{r \lambda x_i^r e^{-(\lambda x_i)^r}}{1 - e^{-(\lambda x_i)^r}} \\ & + (\beta - 1) \sum_{j=1}^m \frac{r \lambda y_j^r e^{-(\lambda y_j)^r}}{1 - e^{-(\lambda y_j)^r}} = 0 \end{aligned} \quad (8)$$

با توجه به روابط (6)، (7) و (8) داریم

$$\hat{\alpha} = \frac{-n}{\sum_{i=1}^n \ln \left(1 - e^{-(\hat{\lambda} x_i)^r} \right)} \quad (9)$$

$$\hat{\beta} = \frac{-m}{\sum_{j=1}^m \ln \left(1 - e^{-(\hat{\lambda} y_j)^r} \right)} \quad (10)$$

که در آن $\hat{\lambda}$ می تواند از حل معادله غیر خطی زیر به دست آید.

$$\begin{aligned} g(\lambda) = 0 \Rightarrow & \frac{m+n}{\lambda} - \frac{n \lambda \sum_{i=1}^n \frac{x_i^r e^{-(\lambda x_i)^r}}{1 - e^{-(\lambda x_i)^r}}}{\sum_{i=1}^n \ln \left(1 - e^{-(\lambda x_i)^r} \right)} - \frac{m \lambda \sum_{j=1}^m \frac{y_j^r e^{-(\lambda y_j)^r}}{1 - e^{-(\lambda y_j)^r}}}{\sum_{j=1}^m \ln \left(1 - e^{-(\lambda y_j)^r} \right)} \\ & - \sum_{i=1}^n \frac{\lambda x_i^r}{1 - e^{-(\lambda x_i)^r}} - \sum_{j=1}^m \frac{\lambda y_j^r}{1 - e^{-(\lambda y_j)^r}} = 0 \end{aligned}$$

بنابراین $\hat{\lambda}$ جواب تکراری ساده از معادله ی غیر خطی $g(\lambda) = 0$ است. وقتی $\hat{\lambda}$ معلوم باشد $\hat{\alpha}$ و $\hat{\beta}$ به ترتیب از رابطه

های (9) و (10) به دست می آیند بنابراین MLE برای $R_{s,k}$ وقتی که $\hat{v} = \frac{\hat{\alpha}}{\hat{\beta}}$ به صورت زیر می باشد

$$\hat{R}_{s,k} = \hat{v} \sum_{i=s}^k \frac{k!}{(k-i)!} \left[\prod_{j=0}^i (k + \hat{v} - j) \right]^{-1}. \quad (11)$$

برای محاسبه فاصله اطمینان مجانبی $R_{s,k}$ ابتدا واریانس مجانبی برآوردگر درستنمایی ماکزیمم توسط

$$Var(\hat{\alpha}) = \left[E \left(-\frac{\partial^2 L}{\partial \alpha^2} \right) \right]^{-1} = \frac{\alpha^r}{n} \quad \& \quad Var(\hat{\beta}) = \left[E \left(-\frac{\partial^2 L}{\partial \beta^2} \right) \right]^{-1} = \frac{\beta^r}{m} \quad (12)$$

به دست می آید که واریانس مجانب (AV) برای برآورد $R_{s,k}$ ، تابعی از دو آماره مستقل برای مثال α و β می باشد (مراجعه کنید به [6])

$$AV(\hat{R}_{s,k}) = Var(\hat{\alpha}) \left(\frac{\partial R_{s,k}}{\partial \alpha} \right)^2 + Var(\hat{\beta}) \left(\frac{\partial R_{s,k}}{\partial \beta} \right)^2 \quad (13)$$

با توجه به خاصیت بهینه ی مجانبی MLE ها [4] و برآوردگرهای نااریب خطی [3] معلوم می شود که MLE ها مجانباً به همان اندازه کارآمد هستند که واریانس مجانبی کران پایین کرامر راثوی آنها دارد همانطور که در رابطه (12) داده شده است. بنابراین در رابطه (13) واریانس مجانبی $\hat{R}_{s,k}$ به دست می آید. برای اجتناب از پیچیدگی مشتق $R_{s,k}$ آن را در حالتی در نظر می گیریم که $(s, k) = (1, 3)$ و $(s, k) = (2, 4)$ باشند.

$$\begin{aligned} \frac{\partial R_{1,3}}{\partial \alpha} &= \frac{-3}{\beta(3+\hat{v})^2} \quad \& \quad \frac{\partial R_{1,3}}{\partial \beta} = \frac{3\hat{v}}{\beta(3+\hat{v})^2} \\ \frac{\partial R_{2,4}}{\partial \alpha} &= \frac{-12(\gamma+2\hat{v})}{\beta[(3+\hat{v})(4+\hat{v})]^2} \quad \& \quad \frac{\partial R_{2,4}}{\partial \beta} = \frac{12\hat{v}(\gamma+2\hat{v})}{\beta[(3+\hat{v})(4+\hat{v})]^2} \end{aligned}$$

بنابراین

$$AV(\hat{R}_{1,3}) = \frac{9\hat{v}^2}{(3+\hat{v})^4} \left(\frac{1}{n} + \frac{1}{m} \right) \quad \& \quad AV(\hat{R}_{2,4}) = \frac{144\hat{v}^2(\gamma+2\hat{v})^2}{[(3+\hat{v})(4+\hat{v})]^4} \left(\frac{1}{n} + \frac{1}{m} \right)$$

اگر $n \rightarrow \infty$ و $m \rightarrow \infty$ میل کند آنگاه $N(0, 1)$ و $\frac{\hat{R}_{s,k} - R_{s,k}}{AV(\hat{R}_{s,k})} \xrightarrow{d} N(0, 1)$ و فاصله اطمینان مجانبی $100(1-\alpha)$ درصد برای $R_{s,k}$ به صورت زیر است.

$$\hat{R}_{s,k} \mp Z_{(1-\alpha/2)} \sqrt{AV(\hat{R}_{s,k})}$$

که در آن $Z_{(1-\alpha/2)}$ ، صدک $(1-\alpha/2)$ ام از توزیع نرمال استاندارد می باشد. فاصله اطمینان مجانبی $100(1-\alpha)$ درصد به ترتیب برای $R_{1,3}$ و $R_{2,4}$ به صورت

$$\hat{R}_{1,3} \mp Z_{(1-\alpha/2)} \frac{3\hat{v}}{(3+\hat{v})^2} \sqrt{\left(\frac{1}{n} + \frac{1}{m} \right)}$$

و

$$\hat{R}_{2,4} \mp Z_{(1-\alpha/2)} \frac{12\hat{v}(\gamma+2\hat{v})}{[(3+\hat{v})(4+\hat{v})]^2} \sqrt{\left(\frac{1}{n} + \frac{1}{m} \right)}$$

می باشند که در آن ها $\hat{v} = \frac{\hat{\alpha}}{\hat{\beta}}$ است.

جدول ۱۱: متوسط اربیی در شبیه سازی برآورد $R_{s,k}$

(s, k)	(m, n)	(α, β)								
		$(3, 5, 1, 5)$	$(3, 0, 1, 5)$	$(2, 5, 1, 5)$	$(2, 0, 1, 5)$	$(1, 5, 1, 5)$	$(1, 5, 2, 0)$	$(1, 5, 2, 5)$	$(1, 5, 3)$	$(1, 5, 3, 5)$
$(1, 3)$	$(10, 10)$	-۰,۰۲۰۳۶	-۰,۰۱۸۷۲	-۰,۰۱۶۵۰	-۰,۰۱۳۵۷	-۰,۰۰۹۸۴	-۰,۰۰۶۴۹	-۰,۰۰۴۲۹	-۰,۰۰۳۶۰	-۰,۰۰۲۴۹
	$(15, 15)$	-۰,۰۱۵۱۷	-۰,۰۱۴۱۲	-۰,۰۱۲۶۸	-۰,۰۱۰۷۵	-۰,۰۰۸۲۵	-۰,۰۰۵۹۲	-۰,۰۰۴۲۸	-۰,۰۰۳۱۱	-۰,۰۰۲۲۷
	$(20, 20)$	-۰,۰۰۸۶۰	-۰,۰۰۷۷۳	-۰,۰۰۶۶۹	-۰,۰۰۵۴۸	-۰,۰۰۳۶۷	-۰,۰۰۲۷۷	-۰,۰۰۱۷۹	-۰,۰۰۱۱۴	-۰,۰۰۱۰۱
	$(25, 25)$	-۰,۰۰۸۵۱	-۰,۰۰۷۶۶	-۰,۰۰۶۵۷	-۰,۰۰۵۲۱	-۰,۰۰۳۵۷	-۰,۰۰۲۱۵	-۰,۰۰۱۲۲	-۰,۰۰۰۶۰	-۰,۰۰۰۱۷
	$(30, 30)$	-۰,۰۰۶۷۹	-۰,۰۰۶۱۳	-۰,۰۰۵۲۸	-۰,۰۰۴۲۱	-۰,۰۰۲۹۰	-۰,۰۰۱۷۵	-۰,۰۰۰۹۸	-۰,۰۰۰۴۶	-۰,۰۰۰۱۰
	$(35, 35)$	-۰,۰۰۶۵۵	-۰,۰۰۶۱۰	-۰,۰۰۵۱۷	-۰,۰۰۴۱۳	-۰,۰۰۲۵۵	-۰,۰۰۱۴۷	-۰,۰۰۰۷۳	-۰,۰۰۰۲۱	-۰,۰۰۰۰۸
$(2, 4)$	$(10, 10)$	-۰,۰۱۱۱۳	-۰,۰۱۱۴۳	-۰,۰۱۱۲۴	-۰,۰۱۰۲۷	-۰,۰۰۸۱۹	-۰,۰۰۶۵۷	-۰,۰۰۵۳۹	-۰,۰۰۴۷۲	-۰,۰۰۳۴۸
	$(15, 15)$	-۰,۰۰۹۰۸	-۰,۰۰۹۴۵	-۰,۰۰۹۵۰	-۰,۰۰۹۰۳	-۰,۰۰۷۷۶	-۰,۰۰۶۰۱	-۰,۰۰۴۴۷	-۰,۰۰۳۲۴	-۰,۰۰۲۲۹
	$(20, 20)$	-۰,۰۰۶۰۱	-۰,۰۰۵۹۹	-۰,۰۰۵۷۱	-۰,۰۰۵۰۸	-۰,۰۰۴۰۰	-۰,۰۰۱۷۶	-۰,۰۰۱۲۰	-۰,۰۰۰۸۹	-۰,۰۰۰۱۳۸
	$(25, 25)$	-۰,۰۰۵۱۲	-۰,۰۰۵۰۵	-۰,۰۰۴۷۳	-۰,۰۰۴۰۵	-۰,۰۰۲۹۳	-۰,۰۰۱۶۸	-۰,۰۰۰۷۳	-۰,۰۰۰۵۴	-۰,۰۰۰۴۵
	$(30, 30)$	-۰,۰۰۴۲۰	-۰,۰۰۴۱۶	-۰,۰۰۳۹۱	-۰,۰۰۳۳۸	-۰,۰۰۲۴۷	-۰,۰۰۱۴۴	-۰,۰۰۰۶۳	-۰,۰۰۰۴۳	-۰,۰۰۰۴۰
	$(35, 35)$	-۰,۰۰۳۸۶	-۰,۰۰۳۹۰	-۰,۰۰۳۷۷	-۰,۰۰۲۳۷	-۰,۰۰۱۶۰	-۰,۰۰۱۲۵	-۰,۰۰۰۵۸	-۰,۰۰۰۱۹	-۰,۰۰۰۳۵

جدول ۱۲: متوسط MSE در شبیه سازی برآورد $R_{s,k}$

(s, k)	(m, n)	(α, β)								
		(۳,۵,۱,۵)	(۳,۰,۱,۵)	(۲,۵,۱,۵)	(۲,۰,۱,۵)	(۱,۵,۱,۵)	(۱,۵,۲,۰)	(۱,۵,۲,۵)	(۱,۵,۳,۰)	(۱,۵,۳,۵)
(۱,۳)	(۱۰,۱۰)	۰,۰۱۵۲۹	۰,۰۱۴۳۷	۰,۰۱۳۰۰	۰,۰۱۱۰۹	۰,۰۰۸۵۴	۰,۰۰۶۲۷	۰,۰۰۴۷۷	۰,۰۰۳۷۵	۰,۰۰۳۰۳
	(۱۵,۱۵)	۰,۰۱۰۵۲	۰,۰۰۹۸۸	۰,۰۰۸۹۴	۰,۰۰۷۶۴	۰,۰۰۵۹۰	۰,۰۰۴۳۶	۰,۰۰۳۳۴	۰,۰۰۲۶۴	۰,۰۰۲۱۴
	(۲۰,۲۰)	۰,۰۰۷۱۳	۰,۰۰۶۶۶	۰,۰۰۵۹۹	۰,۰۰۵۰۹	۰,۰۰۳۹۲	۰,۰۰۲۸۹	۰,۰۰۲۲۱	۰,۰۰۱۷۵	۰,۰۰۱۴۳
	(۲۵,۲۵)	۰,۰۰۵۹۲	۰,۰۰۵۵۱	۰,۰۰۴۹۴	۰,۰۰۴۱۸	۰,۰۰۳۲۲	۰,۰۰۲۳۶	۰,۰۰۱۸۱	۰,۰۰۱۴۴	۰,۰۰۱۱۷
	(۳۰,۳۰)	۰,۰۰۴۶۰	۰,۰۰۴۲۸	۰,۰۰۳۸۳	۰,۰۰۳۲۳	۰,۰۰۲۴۸	۰,۰۰۱۸۲	۰,۰۰۱۳۹	۰,۰۰۱۱۰	۰,۰۰۰۹۰
	(۳۵,۳۵)	۰,۰۰۴۰۲	۰,۰۰۳۷۴	۰,۰۰۳۳۷	۰,۰۰۲۸۶	۰,۰۰۲۲۰	۰,۰۰۱۶۲	۰,۰۰۱۲۵	۰,۰۰۰۹۹	۰,۰۰۰۸۱
(۲,۴)	(۱۰,۱۰)	۰,۰۱۸۰۱	۰,۰۱۸۷۷	۰,۰۱۹۰۰	۰,۰۱۸۳۱	۰,۰۱۶۱۱	۰,۰۱۳۲۰	۰,۰۱۰۷۷	۰,۰۰۸۸۹	۰,۰۰۷۴۴
	(۱۵,۱۵)	۰,۰۱۲۵۸	۰,۰۱۳۳۷	۰,۰۱۳۵۱	۰,۰۱۲۹۸	۰,۰۱۱۴۰	۰,۰۰۹۳۲	۰,۰۰۷۶۲	۰,۰۰۶۳۰	۰,۰۰۵۲۹
	(۲۰,۲۰)	۰,۰۰۹۰۶	۰,۰۰۹۳۶	۰,۰۰۹۳۸	۰,۰۰۸۵۹	۰,۰۰۷۸۱	۰,۰۰۶۳۵	۰,۰۰۵۱۸	۰,۰۰۴۲۹	۰,۰۰۳۶۱
	(۲۵,۲۵)	۰,۰۰۷۵۴	۰,۰۰۷۷۸	۰,۰۰۷۷۸	۰,۰۰۷۴۰	۰,۰۰۶۴۳	۰,۰۰۵۲۲	۰,۰۰۴۲۶	۰,۰۰۳۵۲	۰,۰۰۲۹۶
	(۳۰,۳۰)	۰,۰۰۵۹۴	۰,۰۰۶۱۲	۰,۰۰۶۱۰	۰,۰۰۵۷۸	۰,۰۰۵۰۰	۰,۰۰۴۰۴	۰,۰۰۳۲۸	۰,۰۰۲۷۱	۰,۰۰۲۲۸
	(۳۵,۳۵)	۰,۰۰۵۲۲	۰,۰۰۵۳۸	۰,۰۰۵۳۸	۰,۰۰۵۱۲	۰,۰۰۴۴۵	۰,۰۰۳۶۱	۰,۰۰۲۹۵	۰,۰۰۲۴۴	۰,۰۰۲۰۵

جدول ۱۳: متوسط طول اطمینان در شبیه سازی فاصله اطمینان ۹۵ درصد $R_{s,k}$

(s, k)	(m, n)	(α, β)								
		(۳,۵,۱,۵)	(۳,۰,۱,۵)	(۲,۵,۱,۵)	(۲,۰,۱,۵)	(۱,۵,۱,۵)	(۱,۵,۲,۰)	(۱,۵,۲,۵)	(۱,۵,۳,۰)	(۱,۵,۳,۵)
(۱,۳)	(۱۰,۱۰)	۰,۴۰۹۱	۰,۴۰۲۱	۰,۳۸۸۰	۰,۳۶۳۲	۰,۳۲۲۴	۰,۲۷۶۳	۰,۲۴۰۰	۰,۲۱۱۳	۰,۱۸۸۴
	(۱۵,۱۵)	۰,۳۳۹۹	۰,۳۳۳۴	۰,۳۲۱۰	۰,۲۹۹۹	۰,۲۶۵۸	۰,۲۲۷۹	۰,۱۹۸۱	۰,۱۷۴۷	۰,۱۵۵۹
	(۲۰,۲۰)	۰,۲۹۷۵	۰,۲۹۱۱	۰,۲۷۹۴	۰,۲۶۰۱	۰,۲۲۹۶	۰,۱۹۶۱	۰,۱۷۰۱	۰,۱۴۹۷	۰,۱۳۳۵
	(۲۵,۲۵)	۰,۲۶۷۵	۰,۲۶۱۷	۰,۲۵۱۲	۰,۲۳۳۸	۰,۲۰۶۳	۰,۱۷۶۲	۰,۱۵۲۹	۰,۱۳۴۶	۰,۱۲۰۱
	(۳۰,۳۰)	۰,۲۴۵۳	۰,۲۳۹۸	۰,۲۳۰۰	۰,۲۱۴۰	۰,۱۸۸۷	۰,۱۶۱۲	۰,۱۳۹۸	۰,۱۲۳۲	۰,۱۰۹۹
	(۳۵,۳۵)	۰,۲۲۷۷	۰,۲۲۲۶	۰,۲۱۳۵	۰,۱۹۸۶	۰,۱۷۵۳	۰,۱۴۹۸	۰,۱۳۰۱۱	۰,۱۱۴۶	۰,۱۰۲۳
(۲,۴)	(۱۰,۱۰)	۰,۴۵۶۹	۰,۴۷۱۹	۰,۴۸۰۲	۰,۴۷۶۰	۰,۴۴۹۸	۰,۴۰۵۵	۰,۳۶۳۷	۰,۳۲۷۴	۰,۲۹۶۶
	(۱۵,۱۵)	۰,۳۸۳۴	۰,۳۹۵۳	۰,۴۰۱۲	۰,۳۹۶۸	۰,۳۷۳۹	۰,۳۳۶۶	۰,۳۰۱۹	۰,۲۷۱۹	۰,۲۴۶۶
	(۲۰,۲۰)	۰,۳۳۹۶	۰,۳۴۹۲	۰,۳۵۳۳	۰,۳۴۷۹	۰,۳۲۶۰	۰,۲۹۲۰	۰,۲۶۱۰	۰,۲۳۴۶	۰,۲۱۲۳
	(۲۵,۲۵)	۰,۳۰۵۶	۰,۳۱۴۳	۰,۳۱۸۰	۰,۳۱۳۱	۰,۲۹۳۴	۰,۲۶۲۸	۰,۲۳۵۰	۰,۲۱۱۲	۰,۱۹۱۲
	(۳۰,۳۰)	۰,۲۸۱۳	۰,۲۸۹۱	۰,۲۹۲۳	۰,۲۸۷۵	۰,۲۶۹۱	۰,۲۴۰۹	۰,۲۱۵۳	۰,۱۹۳۵	۰,۱۷۵۲
	(۳۵,۳۵)	۰,۲۶۱۱	۰,۲۶۸۴	۰,۲۷۱۴	۰,۲۶۷۰	۰,۲۵۰۰	۰,۲۲۴۰۱	۰,۲۰۰۳	۰,۱۸۰۱	۰,۱۶۳۱

جدول ۱۴: متوسط احتمال پوشش در شبیه سازی فاصله اطمینان ۹۵ درصد $R_{s,k}$

(s, k)	(m, n)	(α, β)								
		(۳,۵,۱,۵)	(۳,۰,۱,۵)	(۲,۵,۱,۵)	(۲,۰,۱,۵)	(۱,۵,۱,۵)	(۱,۵,۲,۰)	(۱,۵,۲,۵)	(۱,۵,۳,۰)	(۱,۵,۳,۵)
(۱,۳)	(۱۰,۱۰)	۰,۹۰۹۰	۰,۹۱۴۰	۰,۹۱۹۳	۰,۹۲۸۰	۰,۹۳۱۷	۰,۹۳۰۳	۰,۹۲۶۷	۰,۹۲۶۰	۰,۹۲۶۷
	(۱۵,۱۵)	۰,۹۱۲۰	۰,۹۱۵۰	۰,۹۱۸۷	۰,۹۲۱۳	۰,۹۲۵۰	۰,۹۲۷۳	۰,۹۲۹۰	۰,۹۲۸۷	۰,۹۲۵۳
	(۲۰,۲۰)	۰,۹۳۰۳	۰,۹۳۴۷	۰,۹۳۷۰	۰,۹۳۹۰	۰,۹۳۹۰	۰,۹۳۷۷	۰,۹۳۸۰	۰,۹۳۴۷	۰,۹۳۰۳
	(۲۵,۲۵)	۰,۹۲۲۷	۰,۹۲۶۷	۰,۹۳۱۷	۰,۹۳۶۰	۰,۹۳۸۳	۰,۹۳۷۳	۰,۹۳۵۷	۰,۹۳۳۳	۰,۹۲۷۷
	(۳۰,۳۰)	۰,۹۳۵۳	۰,۹۴۰۳	۰,۹۴۲۳	۰,۹۴۶۳	۰,۹۴۹۰	۰,۹۴۶۳	۰,۹۴۳۳	۰,۹۴۰۰	۰,۹۳۹۰
	(۳۵,۳۵)	۰,۹۳۱۷	۰,۹۳۳۰	۰,۹۳۵۳	۰,۹۳۸۷	۰,۹۳۸۷	۰,۹۴۰۰	۰,۹۳۶۳	۰,۹۳۳۷	۰,۹۲۹۷
(۲,۴)	(۱۰,۱۰)	۰,۹۱۱۳	۰,۹۱۵۳	۰,۹۲۰۳	۰,۹۲۴۳	۰,۹۲۷۳	۰,۹۲۷۷	۰,۹۲۵۷	۰,۹۲۵۰	۰,۹۲۴۳
	(۱۵,۱۵)	۰,۹۱۰۳	۰,۹۱۶۰	۰,۹۱۹۰	۰,۹۱۹۳	۰,۹۲۳۳	۰,۹۲۶۳	۰,۹۲۹۷	۰,۹۲۶۰	۰,۹۲۲۳
	(۲۰,۲۰)	۰,۹۳۱۰	۰,۹۳۴۰	۰,۹۳۷۰	۰,۹۳۸۰	۰,۹۳۶۷	۰,۹۳۶۷	۰,۹۳۶۰	۰,۹۳۲۳	۰,۹۲۸۷
	(۲۵,۲۵)	۰,۹۲۳۷	۰,۹۲۸۷	۰,۹۳۱۳	۰,۹۳۵۳	۰,۹۳۷۰	۰,۹۳۵۰	۰,۹۳۲۳	۰,۹۳۰۷	۰,۹۲۵۷
	(۳۰,۳۰)	۰,۹۳۵۷	۰,۹۳۹۷	۰,۹۴۳۷	۰,۹۴۳۷	۰,۹۴۷۷	۰,۹۴۵۷	۰,۹۴۲۰	۰,۹۴۰۳	۰,۹۳۹۷
	(۳۵,۳۵)	۰,۹۲۹۷	۰,۹۳۱۷	۰,۹۳۶۷	۰,۹۳۶۰	۰,۹۴۰۷	۰,۹۳۹۳	۰,۹۳۶۷	۰,۹۳۲۷	۰,۹۳۰۰

۳ مطالعه ی شبیه سازی و تجزیه و تحلیل داده ها

۱.۳ مطالعه ی شبیه سازی

برای مقایسه ی کارایی $R_{s,k}$ بر اساس شبیه سازی مونت کارلو از نمونه هایی با اندازه های مختلف استفاده شد. ۳۰۰۰ نمونه تصادفی با حجم ۱۰(۵)۳۵ که هر کدام از جامعه ی تنش می باشد جامعه ی مقاومت برای (α, β) های (۳/۰, ۱/۰), (۲/۵, ۱/۰), (۲/۰, ۱/۰), (۱/۵, ۱/۰), (۱/۰, ۱/۰), (۱/۵, ۲/۰), (۱/۵, ۲/۵) و (۱/۵, ۳/۰) بر اساس خطوط [2] تولید شده است. برآوردگر ML برای α و β و سپس v جهت تعیین قابلیت اعتماد تنش-مقاومت چند مؤلفه ای برای $(s, k) = (۱, ۳), (۲, ۴)$ به دست آمد. متوسط اریبی و میانگین مربع خطا (MSE) برای برآورد قابلیت اعتماد با بیش از ۳۰۰۰ تکرار در جدول ۱ و جدول ۲ نشان داده شده است. متوسط طول اطمینان و احتمال پوشش شبیه سازی شده در فاصله اطمینان ۹۵ درصد برای $R_{s,k}$ در جدول ۳ و ۴ داده شده است. مقدار واقعی قابلیت اعتماد تنش-مقاومت چند مؤلفه ای ترکیبی از (α, β) برای $(s, k) = (۱, ۳)$ به صورت ۰/۵۶۳، ۰/۶۰۰، ۰/۶۴۳، ۰/۶۹۲، ۰/۷۵۰، ۰/۸۰۰، ۰/۸۳۳، ۰/۸۷۵، ۰/۸۵۷ و برای $(s, k) = (۲, ۴)$ به صورت ۰/۳۵۵، ۰/۴۰۰، ۰/۴۵۴، ۰/۵۱۹، ۰/۶۰۰، ۰/۶۷۴، ۰/۷۲۵، ۰/۷۶۲، ۰/۷۹۰ می باشد. بنابراین برای هر (s, k) مقدار واقعی قابلیت اعتماد تنش-مقاومت چند مؤلفه ای افزایش پیدا می کند زمانی که β افزایش یابد و α ثابت باشد. اگر چه با افزایش α و ثابت بودن β کاهش پیدا می کند. بنابراین مقدار واقعی قابلیت اعتماد

تنش-مقاومت چند مؤلفه ای افزایش می یابد وقتی V کاهش پیدا میکند و بالعکس. متوسط اریبی و متوسط MSE کاهش پیدا می کند وقتی که حجم نمونه برای هر (s, k) افزایش پیدا کند که ویژگی سازگاری MLE برای $R_{s,k}$ را تایید می کند. همچنین میزان اریبی در هر (s, k) منفی است. هرچند در میان پارامترها، اریبی مطلق و MSE با افزایش α و در یک مقدار ثابت از β در هر (s, k) افزایش پیدا می کند. اریبی مطلق و MSE با افزایش β برای یک مقدار ثابت از α در هر (s, k) کاهش پیدا می کند. متوسط طول فاصله اطمینان با افزایش حجم نمونه کاهش می یابد. در همه ی موارد احتمال پوشش نزدیک به مقدار اسمی اما در بیشتر ترکیبات کمی کمتر از ۰,۹۵ می باشد. بطور کلی عملکرد فاصله اطمینان در همه ی ترکیبات پارامترها مطلوب است در حالی که در میان پارامترهای مشاهده شده پدیده ی یکسانی برای متوسط طول و متوسط احتمال پوشش در مواردی از متوسط اریبی و MSE مشاهده می شود.

۲.۳ تجزیه و تحلیل داده ها

داده های مقاومت که اولین بار توسط [1] گزارش شد که نشان دهنده مقاومت اندازه گیری شده در GPA برای الیاف تک کربن و زنجیره ی الیاف ۱۰۰۰ کربنی هستند. الیاف تک کربنی تحت کشش در مقیاس طول ۲۰ میلی متر (مجموعه داده های شماره ۱) و ۱۰ میلی متر (مجموعه داده های شماره ۲) با حجم نمونه $n = ۶۹$ و $m = ۶۳$ مورد آزمایش قرار گرفتند.

مجموعه داده های شماره ۱:

1,312, 1,314, 1,479, 1,552, 1,700, 1,803, 1,861, 1,865, 1,944, 1,958, 1,966, 1,997, 2,006,
2,021, 2,027, 2,055, 2,063, 2,098, 2,140, 2,179, 2,224, 2,240, 2,253, 2,270, 2,272, 2,274,
2,301, 2,301, 2,359, 2,382, 2,382, 2,426, 2,434, 2,435, 2,478, 2,490, 2,511, 2,514, 2,535,
2,554, 2,566, 2,570, 2,586, 2,629, 2,633, 2,642, 2,648, 2,684, 2,697, 2,726, 2,770, 2,773,
2,800, 2,809, 2,818, 2,821, 2,848, 2,880, 2,809, 2,818, 2,821, 2,848, 2,880, 2,954, 3,012,
3,067, 3,084, 3,090, 3,096, 3,128, 3,233, 3,433, 3,585, 3,585.

مجموعه داده های شماره ۲:

1,901, 2,132, 2,203, 2,228, 2,257, 2,350, 2,361, 2,396, 2,397, 2,445, 2,454, 2,474, 2,518,
2,522, 2,525, 2,532, 2,575, 2,614, 2,616, 2,618, 2,624, 2,659, 2,675, 2,738, 2,740, 2,856,
2,917, 2,928, 2,937, 2,937, 2,977, 2,996, 3,030, 3,125, 3,139, 3,145, 3,220, 3,223, 3,235,
3,243, 3,264, 3,272, 3,294, 3,332, 3,346, 3,377, 3,408, 3,435, 3,493, 3,501, 3,537, 3,554,
3,562, 3,628, 3,852, 3,871, 3,886, 3,971, 4,024, 4,027, 4,225, 4,395, 5,020.

[10] و [9] مشاهده کردند که رایی تعمیم یافته برای داده های مقاومت به خوبی کار می کند. مجموعه داده های شماره ۱ و مجموعه داده های شماره ۲ به ترتیب با تفریق ۱ و ۱/۸ توسط [8] مورد تجزیه و تحلیل قرار گرفت با

این فرض که داده های تبدیل یافته مطابق با مقیاس طولی ۱۰ میلی متر و ۲۰ میلی متر به ترتیب از توزیع $GR(\alpha, \lambda)$ و $GR(\beta, \lambda)$ باشند. در این دو مجموعه از داده ها برآورد پارامترهای α ، β و λ محاسبه شدند که $\hat{\alpha} = ۲/۴۴۲۱$ ، $\hat{\beta} = ۱/۴۲۱۶$ و $\hat{\lambda} = ۰/۸۵۹۸$ می باشند همچنین اعتبار سنجی مدل برای هر یک از مجموعه داده ها با استفاده از آزمون کلموگراف اسمیرنف (K-S) مورد بررسی قرار گرفت. در هر دو مجموعه از داده های شماره ۱ و شماره ۲ فاصله های کلموگراف اسمیرنف ۰/۰۹ و ۰/۱۲ به ترتیب با p -مقدار متناظر ۰/۶۰۶۹ و ۰/۲۸۴۵ مشاهده شد که نشان می دهد مدل GR مدل مناسب و معقولی برای مجموعه داده های تبدیل یافته است. با توجه به برآورد α و β ، همچنین فاصله اطمینان ۹۵ درصد برای $R_{1,3}$ به صورت $(۰/۵۵۶۸۰, ۰/۷۱۴۹۶)$ و برای $R_{2,4}$ به صورت $(۰/۳۴۳۸۷, ۰/۵۴۵۸۱)$ می باشند وقتی که $(s, k) = (۱, ۳)$ و $(s, k) = (۲, ۴)$.
به دست می آید.

۴ نتیجه گیری

در این مقاله از دو مجموعه داده واقعی برای بررسی قابلیت اطمینان تنش مقاومت چند مؤلفه ای استفاده شد وقتی که متغیرهای تصادفی تنش و مقاومت هر دو از توزیع رایلی تعمیم یافته باشند فاصله اطمینان مجانبی برای قابلیت اعتماد تنش مقاومت چند مؤلفه ای با استفاده از روش ماکزیمم درستنمایی برآورد شد. نتایج شبیه سازی نشان می دهد که برای هر (s, k) با افزایش حجم نمونه، متوسط اریبی و متوسط MSE کاهش پیدا می کند. در میان پارامترها برای هر (s, k) اریبی مطلق و MSE افزایش (کاهش) پیدا می کند وقتی که α افزایش (β کاهش) پیدا کند. همچنین با افزایش حجم نمونه طول فاصله اطمینان کاهش می یابد و در تمامی مجموعه پارامترهای مطرح شده احتمال پوشش نزدیک به مقدار اسمی است.

مراجع

- [1] Badar, M. G. Priest, A. M. (1982), *Statistical aspects of fiber and bundle strength in hybrid composites*, Progress in Science and Engineering Composites, 1129-1136.
- [2] Bhattacharyya, G. K. Johnson, R. A. (1974), *Estimation of reliability in multicomponent stress – strength model*, JASA, 69: 966-970.

- [3] David, H. A., Nagaraja, H. N. (1981), *Order statistics*, New York: John Wiley and Sons.
- [4] Kendall, M. G. Stuart, A. (1979), *The Advanced Theory of Statistics*, Vol.London: Charles Griffin and Company Limited.
- [5] Kundu, D. Raqab, M. Z. (2005), *Generalized Rayleigh distribution: different methods of estimation*, Computational Statistics and Data Analysis, 49:187 - 200.
- [6] Rao, C. R. (1973), *Linear Statistical Inference and its Applications*, India:Wiley Eastern Limited.
- [7] Rao, G. S. (2014), *Estimation of reliability in multicomponent stress-strength based on generalized Rayleigh distribution*, Journal of Modern Applied Statistical Methods, 13-24.
- [8] Raqab, M. Z. Kundu, D. (2005), *Comparison of different estimators of $p(Y < X)$ for a scaled Burr type X distribution*, Communications in Statistics –Simulation and Computation, 34: 465 - 483.
- [9] Surles, J. G., Padgett, W. J. (1998), *Inference for $P(Y < X)$ in the Burr type X model*, Journal of Applied Statistical Science, 7, 225-238.
- [10] Surles, J. G., Padgett, W. J. (2001), *Inference for reliability and stress-strength for a scaled Burr Type X distribution. Lifetime Data Analysis*, 7, 187-200.



برآورد قابلیت اطمینان مدت سفر در شریان شهری مبتنی بر روش تابع مفصل

علیزاده، س^۱ زال زاده، س^۲

^{۱،۲} گروه آمار، دانشکده ریاضی، آمار و علوم کامپیوتر، دانشگاه سمنان

چکیده

در جوامع کنونی تسهیلات حمل و نقل از اهمیت بالایی برخوردار است. سرعت، زمان سفر و تاخیر شاخص‌های مهمی هستند که به عنوان نشانه عملکرد ترافیکی در نظر گرفته می‌شوند. پیش‌بینی زمان سفر نقش به سزایی در جلوگیری از به دام افتادن استفاده‌کننده‌ها در ترافیک سنگین و افزایش اعتماد آنها به تسهیلات ترافیکی دارد. در سیستم‌های پیشرفته کنترل ترافیک روش‌های برآورد قابلیت اعتماد زمان سفر TTR بیان می‌کند که مدت پیمودن مسیر از توزیع خاصی پیروی می‌کند و در مطالعات غالباً همبستگی میان قطعات مسیر در نظر گرفته نمی‌شود. توزیع زمان سفر که در اصطلاح با TTD نشان داده می‌شود مبنای کمیت سازی برای TTR است. در این مطالعه به روشی برای ساخت TTD پرداخته می‌شود [۴] و فرض می‌کنیم که زمان سفرها در قطعات وابسته است. برای توصیف این وابستگی‌ها از تابع مفصل استفاده می‌گردد و سپس تابع مفصلی که بهترین برازش را بر اساس پارامترهای برآوردی داشته باشد تعیین می‌گردد. سپس روشی برای مدل سازی مدت سفر معرفی می‌گردد و درخاتمه مدل بیان شده با روش مطالعاتی با داده‌های AVL آزمون می‌گردد.

کلمات کلیدی: شریان شهری، توزیع مدت سفر، تابع مفصل.

^۱ Sanaz Alizadeh72@yahoo.com

^۲ SZalazadeh@Semnan.ac.ir

۱ پیش‌گفتار

زمان سفر در تعریفی کوتاه و ساده عبارت است از مدت زمان لازم برای جابه‌جایی از نقطه‌ای تا نقطه‌ی دیگر. روشن است که زمان سفر در یک جاده به مشخصات فیزیکی مانند طول، تعداد خطوط و شیب، شرایط محیطی و آب و هوایی و وضعیت ترافیک عبوری از آن وابسته است. (برای مطالعه بیشتر به [1] مراجعه شود)

قابلیت اطمینان مدت سفر یک معیار مهم برای ارزیابی بازدهی عملیات تجهیزات راه، بررسی سایر گزینه‌های استراتژی مدیریت ترافیک و ارائه‌ی مسیره‌ی به موقع و قابل اطمینان به مسافران محسوب می‌شود. به منظور مشخص سازی کامل TTR ^۳ لازم است توزیع مدت سفر به عنوان یک اولویت تعیین گردد. به علت همبستگی بین قطعات مجاور مدت سفر حاصل در شریان‌ها اغلب توزیعات گوناگونی را در سطوح متفاوت راه‌بندان نشان می‌دهند.

آخرین مطالعات بر روی TTD ^۴ مسیر تلاش فراوانی را صرف شناسایی بهترین مدل آماری برای برازش دهی مشاهدات مدت سفر کرده‌اند اما اساساً به این نتیجه رسیده‌اند که TTD مسیر در شرایط روانی جریان تردد و شرایط راه‌بندان متفاوت است. با توجه به اینکه امکان دارد مسیرهای مختلفی بین یک جفت مبدأ-مقصد وجود داشته باشد و هر مسیر خودش شامل چندین بخش باشد عملاً برآورد TTD مسیر براساس هریک از TTD های قطعی واقع بینانه‌تر خواهد بود بنابراین انتخاب مسیر با انتخاب بهترین مسیر از میان تمام مسیرهای قابل استفاده طبق معیارهای قابلیت اطمینان صورت می‌گیرد. مطالعه‌ی حاضر به این منظور مدل تابع مفصل را جهت برآورد TTD مسیر با توجه به مجموعه‌ی قطعات و با در نظر گرفتن همبستگی مکانی بین مدت سفرهای قطعی معرفی کرده است.

۲ تئوری تابع مفصل

توابع مفصل کاربرد وسیعی در مدلسازی توزیع چند متغیره در میان متغیرهای تصادفی با توزیع‌های حاشیه‌ای از پیش تعیین شده دارد. رویکرد تابع مفصل یک روش انعطاف پذیر برای تشریح وابستگی غیر خطی در میان مدت سفرهای قطعی متوالی به تفکیک از توزیع حاشیه‌ای آن‌ها ارائه می‌کند (برای مطالعه توابع مفصل و ویژگیهای آن به [2] مراجعه شود). روش تابع مفصل به دلیل سادگی نسبت به روش‌های آماری قدیمی‌تر مانند روش همگشت برتری

^۳Travel Time Reliability

^۴Travel Time Distribution

دارد. با استفاده از مثالی که در پایان ارائه می‌گردد این برتری قابل رویت است. براساس قضیه‌ی اسکولار برای تابع توزیع n متغیره $F(x_1, x_2, \dots, x_n)$ با تابع توزیع حاشیه‌ای $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$ یک تابع مفصل مخصوص وجود دارد که رابطه‌ی زیر را برآورد می‌کند:

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))$$

۱.۲ خانواده توابع مفصل دو متغیره

۱.۱.۲ تابع مفصل گاوسی

توابع مفصل گاوسی به عنوان پرکاربردترین تابع مفصل بیضوی و متقارن است، این تابع مختص به توزیع نرمال دو متغیره بیضوی است.

این تابع مفصل به صورت زیر تعریف می‌شود:

$$C(u_1, u_2; \theta) = \frac{\partial^2}{\partial u_1 \partial u_2} C(u_1, u_2; \theta) \\ = \frac{1}{\sqrt{1-\theta^2}} \exp\left(\frac{2\theta\Phi^{-1}(u_1)\Phi^{-1}(u_2) - \theta^2(\Phi^{-1}(u_1))^2 + \Phi^{-1}(u_2)^2}{2(1-\theta^2)}\right)$$

۲.۱.۲ تابع مفصل فارلی-گامبل-مورگسترن FGM

تابع توزیع تجمعی توام این تابع مفصل دو متغیره با رابطه‌ی زیر به دست می‌آید:

$$c_\theta(u_1, u_2) = u_1 u_2 (1 + \theta(1 - u_1)(1 - u_2))$$

۳.۱.۲ تابع مفصل ارشمیدسی

این نوع از توابع مفصل پرکاربرد هستند و با روشی کاملاً متفاوت با توابع مفصل نرمال و FGM ^۵ با استفاده از تابع مولد φ ساخته می‌شود.

$$c(u_1, u_2, \dots, u_n) = \Phi^{[-1]} \{ \varphi(u_1) + \varphi(u_2) + \dots + \varphi(u_n) \}$$

^۵Farlie-Gumbel-Morgenstern

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t), & 0 \leq t \leq \varphi(0), \\ 0, & \varphi(0) < t < \infty \end{cases}$$

توابع مولد متفاوت انواع مختلفی از توابع ارشمیدسی را به وجود می‌آورند. هر نوع از توابع مفصل ارشمیدسی شکل خاص خودش را برای وابستگی دارد. به عنوان مثال توابع مفصل گامبل، کلایتون فقط وابستگی مثبت رامجاز می‌داند. تابع مفصل فرانک هر دو وابستگی مثبت و منفی را مجاز می‌داند. این در حالی است که تابع FGM فقط وابستگی ضعیف را مشخص می‌کند.

۴.۱.۲ تابع مفصل AMH

اگر تابع مولد $\varphi(t) = \ln \frac{1-\theta(1-t)}{t}$ باشد θ در دامنه $1 - \theta$ تا 1 باشد تابع مفصل، AMH است و به صورت زیر تعریف می‌شود:

$$c_{\theta}(u_1, u_2) = \frac{u_1 u_2}{1 - \theta(1 - u_1)(1 - u_2)}$$

تای کندال به صورت $\tau = \frac{2\theta-2}{3\theta} - \frac{2(1-\theta)^2}{3\theta^2}$ خواهد بود به طوریکه τ در دامنه $0/182 -$ و $0/333$ قرار می‌گیرد.

۵.۱.۲ تابع مفصل فرانک

اگر تابع مولد $\varphi(t) = \ln \frac{\theta - \theta t}{\theta - \theta - 1}$ و $\theta \neq 0$ باشد تابع مفصل را فرانک می‌نامیم که توسط فرانک پیشنهاد شده است و به صورت زیر بیان می‌شود:

$$c_{\theta}(u_1, u_2) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right)$$

۲.۲ تابع مفصل سه متغیره گاوسی

تابع مفصل سه متغیره گاوسی در حالت سه متغیره میتواند مانند حالت دو متغیره به صورت زیر بیان شود:

^۶ Ali-Mikhail-Hap

$$C_p(u) = \Phi_p(\Phi^{-1}(u_1), \Phi^{-1}(u_2), \Phi^{-1}(u_3)) \\ = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \int_{-\infty}^{\Phi^{-1}(u_3)} \frac{1}{(\sqrt{2\pi})^3 |P|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} w^T P^{-1} w\right) dw$$

علاوه بر این نوع از توابع، توابع مفصل سه متغیره ارشمیدسی هم برای مدل سازی توزیع چند متغیره به کار می روند. اما این نوع از توابع مفصل فقط برای مدل سازی وابستگی مثبت به کار می رود. بنابراین با توجه به رابطه احتمالی بین مدت سفرهای قطاعی (مدت زمان سفر در هر قطعه از مسیر) در این مطالعه فقط تابع مفصل سه متغیره ی گاوسی را لحاظ می کنیم.

۳.۲ برآورد توزیع حاشیه ای در توابع مفصل

همانطور که در بخش های قبلی گفتیم امکان دارد مدت سفر قطاعی به دلیل حالت های ناهمگن تردد از توزیع های چندنمایی پیروی کند. پس مدل مرکب گاوسی GMM ^۷ به عنوان یک برآوردگر پارامتری و معرف توزیع چندنمایی عملاً برای برازش دادن TTD قطاعی ارجحیت دارد. GMM متناهی با k مؤلفه به صورت زیر نشان داده می شود:

$$p(y|u_1, u_2, \dots, u_k; s_1, s_2, \dots, s_k; \pi_1, \pi_2, \dots, \pi_k) = \sum_{j=1}^k \pi_j N(u_j, s_j^{-1})$$

در این معادله u_j میانگین ها را نشان می دهد، s_j دقت (واریانس معکوس) را نشان می دهد، π_j نسبت های مرکب را نشان می دهد که باید مثبت و جمعاً برابر با یک باشند و N یک گاوسی (نرمال شده با میانگین و واریانس مشخص) می باشد.

بعلاوه برآوردگر هموار کنند هسته برای کمک به تفسیر منحنی های تراکم اضافه می شود. تراکم در x به روش هموارکنندگی هسته به صورت زیر به دست می آید:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right)$$

^۷Gaussian Mixture Model

در اینجا k تابع هسته، h پارامتر هموارکنندگی با نام پهنای باند و x_i داده‌های مشاهده شده هستند. در این مطالعه کرنل گاوسی انتخاب شده و پهنای باند H طبق قواعد بهینه بودن در کار بومن و آزالینی^۸ انتخاب شده است. در انتخاب توزیع حاشیه‌ای فرض می‌کنیم مدت سفر قطاعی از یک توزیع بخصوص پیروی می‌کند. بهترین مدل برازش شده برای توزیع مدت سفرهای قطاعی با توجه به AIC ^۹ و آزمون $(K - S)$ ^{۱۰} انتخاب می‌شود.

۳ مدل‌سازی قابلیت اطمینان مدت سفر

در اصطلاحات حمل و نقل یک قطعه منفرد چنان تعریف می‌شود که مجموعه‌ای از این قطعات به طور متوالی می‌توانند بخشی از راه را بوجود آورند. بدین ترتیب برای به دست آوردن برآورد مدت سفر در مسیری با چندین قطعه منفرد رابطه‌ای به صورت زیر تعریف می‌شود:

$$R_x = P(T_x < \lambda t_x^s)$$

R_x نشان دهنده TTR مسیر x و t_x مدت زمان سفر در مسیر x و t_x^s مدت سفر برای مسیر x در جریان تردد به صورت روان است. λ یک عدد ثابت نشانگر استاندارد برای مدت سفر مورد انتظار می‌باشد. نکاتی که در برآورد TTR حائز اهمیت است محاسبه چواگی و عرض TTD است [3]. زیرا به هر میزان قدر مطلق چولگی بزرگ باشد احتمال مدت سفر نیز بیشتر خواهد بود. به همین علت به منظور کمیت دادن به چولگی برای TTD مسیر، رابطه زیر تعریف می‌شود:

$$s = \frac{E(x = \mu)^2}{\sigma^2}$$

در این معادله μ میانگین متغیر X است و σ انحراف استاندارد X را نشان می‌دهد و $E(T)$ مقدار مورد انتظار کمیت t را نشان می‌دهد. عرض TTD به صورت زیر تعریف می‌شود:

$$\lambda^{var} = \frac{T_{\delta^0} - T_{\lambda^0}}{T_{\delta^0}}$$

^۸Bowman and Azzalini

^۹Akaike Information Criterion

^{۱۰}Kolmogrov-Smirnov

در این رابطه هرچه عرض توزیع بیشتر باشد قابلیت اطمینان کمتر است. میتوانیم برای توصیف کردن TTD با توجه به وابستگی بین مدت سفر قطعات تابع مفصل مورد نظر را تعیین کنیم. در این میان با توجه به آزمون نیکویی برازش تابع مفصل مناسب برای توصیف این وابستگی انتخاب می شود. به طور مثال در این مطالعه فرآیند برآورد TTD مسیر براساس تابع مفصل FGM را انتخاب می کنیم.

تابع توزیع تجمعی مدت سفر در مسیر را می توان به صورت زیر در نظر گرفت:

$$f_{T_1 T_2 \dots T_n} = c \left(F_{T_1}(t_1), F_{T_2}(t_2), \dots, F_{T_n}(t_n) \right) \prod_{i=1}^n f_{T_i}(t_i)$$

$$= [\lambda + \theta(F_T(t) - \lambda)] \prod_{i=1}^n f_{T_i}(t_i)$$

در این معادله $F_T = F_{T_1}(t), F_{T_2}(t), \dots, F_{T_n}(t); F_{T_i}(t_i)$ و تابع توزیع تجمعی قطعه i ام و $f_{T_i}(t_i)$ هم تابع توزیع احتمال قطعه i می باشد. بعد از انتخاب TTD با ترکیب این معادله با معادله $TTD, R_x = p(T_x < \lambda t_x)$ مسیر را محاسبه می کنیم:

$$R_x = P(T_x < \lambda t_x) = F_z(\lambda t_x)$$

$$F_z(\lambda t_x) = p(T_1 + T_2 + \dots + T_n \leq \lambda t_x) = \int \int_{z < \lambda t_x} \dots \int f(t_1, t_2, \dots, t_n) dt_1 dt_2 \dots dt_n$$

در نهایت TTR مسیر از طریق رابطه ی زیر برآورد می شود:

$$(*)R_x = \int \int_{z < \lambda t_x} [\lambda + \theta(F_T(t) - \lambda)] \prod_{i=1}^n f_{T_i}(t_i) dt_1 dt_2 \dots dt_n \quad (1)$$

۴ مطالعه موردی

به منظور تفهیم موارد ذکر شده حرکت در یک شریان در شانگهای چین را مورد بررسی قرار می دهیم. آنچه در ادامه ارائه می گردد شرح حال سایت مطالعاتی و فرآیند تهیه ی داده ها است سپس توابع مفصل براساس TTD های قطاعی

برای برآورد TTD کل مسیر برآورد می‌گردد و TTD های مسیر که توسط توابع مفصل برآورد شده بودند مورد مقایسه قرار گرفته و در خاتمه TTR مسیر برآورد می‌شود.

۵ سایت مطالعاتی

سایت مطالعاتی مسیر ۱,۸ کیلومتری یک شریان شهری به نام خیابان چانگشو در شهر شانگهای چین است که سرعت مجاز در آن ۴۰ کیلومتر در ساعت است. مقطع مطالعاتی از هفت تقاطع دارای چراغ علائم تشکیل شده است. داده‌های مدت سفر توسط تشخیص اتوماتیک پلاک خودرو و شناسایی اتوماتیک وسیله نقلیه AVL ^{۱۱} جمع‌آوری شده‌اند. جمعا مدت سفر ۴۲۵۷ خودرو به دست آمد.

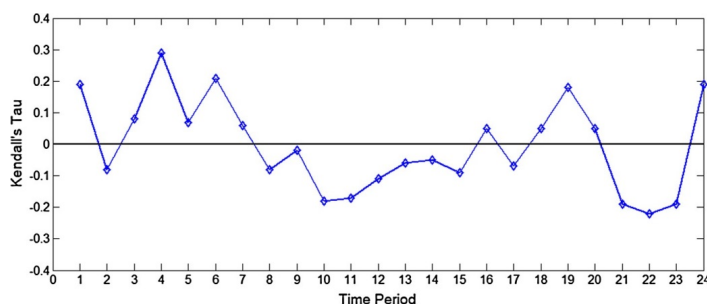
با توجه به ساعت روز، راه‌بندان و بروز هر اتفاقی که تغییر معنیداری در مدت سفر در مسیر ایجاد کند امکان دارد ساختار وابستگی بین مدت سفر در قطعات نیز بسته به شرایط تردد تغییر کند. ابتدا وابستگی ساعتی در میان مدت سفر را ارزیابی میکنیم که مبتنی بر داده‌های AVL در سطح خودرو به خودرو می‌باشد. برای هر ساعت از ضریب همبستگی خطی پیرسون (ρ_p)، تای کندال τ ، و رو اسپیرمن ρ_s ، برای سنجش وابستگی استفاده شده است. از نتایج مشخص شد فاصله اطمینان برای رو اسپیرمن در مقایسه با تائو کندال قابلیت اطمینان و تفسیرپذیری کمتری دارد در حالیکه ضریب همبستگی خطی پیرسون در قدر مطلق کلاً با اغراق بیان شده است. پس ما از تائو کندال که قابلیت اطمینان زیادی دارد برای آنالیز وابستگی بین مدت سفر در قطعات دارد استفاده می‌کنیم. چند تابع مفصل برای مشخص سازی این همبستگی ضعیف وجود دارد. براساس آمار، به دلیل وجود انشعابات و تصادفی بودن سفر مردم تعداد نمونه‌ها فقط ۲۰ درصد از تردد در قطعات منفرد را به خود اختصاص داده است. مقدار تای کندال $|\tau| < 0.1$ یا $|\tau| \approx 0.1$ برای دوره‌های اوج تردد نشان می‌دهد که قطعات مجاور در طول این دوره‌ها به سختی به یکدیگر وابسته هستند. در اصل امکان دارد سطح همبستگی بین مدت سفر در قطعات تحت تأثیر پارامترها و متغیرهای متعددی مثل طول قطعه، حجم تردد، علائم راهنمایی و رانندگی قرار بگیرد.

^{۱۱} Automatic Vehicle Location

۶ انتخاب تابع مفصل بهینه

منحنی تای کندال در دوره‌ای ۲۴ ساعته را در شکل زیر نشان داده‌ایم. نتایج نشان می‌دهند که ساختار وابستگی بین مدت سفرهای قطعی در کل در طول خیابان چانگشو همبستگی ضعیفی را نشان می‌دهند. تای کندال برای این همبستگی برابر با ۰/۱۶۴ - به دست آمد که نشان می‌دهد وابستگی بین مدت سفر در قطعات منفی و ضعیف است. بنابراین توابع مفصل مورد بررسی شامل FGM ، AHM ، فرانک و گاوسی می‌باشند. پس از برآورد پارامتر هر مفصل لازم است که بهترین مفصل برازش شده انتخاب گردد. با توجه به نتایجی که برای مقادیر AIC ، لگاریتم درستنمایی، P -مقدار و آماره CVM برای آزمون نیکویی برازش مفصل در جدول زیر نشان داده شده است مفصل فرانک که دارای کمترین مقدار AIC ، و P -مقدار ۰,۰۱۱ و بزرگترین مقدار لگاریتم درستنمایی است می‌تواند به عنوان بهترین انتخاب در نظر گرفته شود.

شکل ۱



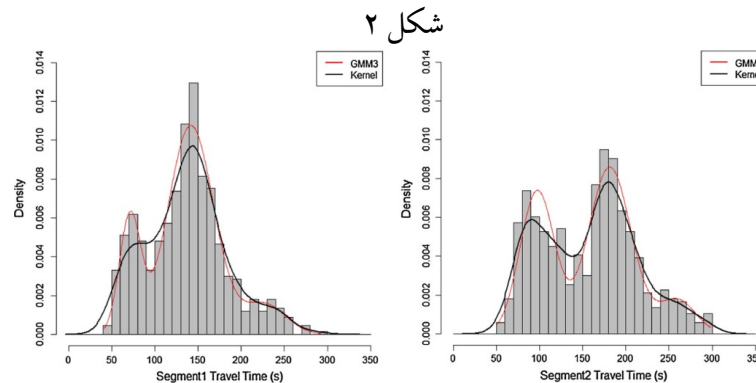
جدول ۱۵

	Goodness-of-fit statistics	AMH	FGM	Frank	Gaussian
Segment ۱ and Segment ۲	CvM	۰/۳۷۷	۰/۲۰۹	۰/۲۵۱	۰/۲۱۱
	p-value	۰/۱۷۵	۰/۲۸۷۲	۰/۱۰۹۴	۰/۲۷۲۲
	Log-likelihood	-۶۹۶۹,۰۰۹	-۶۹۶۴,۱۲	-۶۹۶۳,۴۴۱	-۶۹۷۳,۷۳
	AIC	۱۳۹۵۸,۰۲	۱۳۹۴۸,۲۴	۱۳۹۴۶,۸۸	۱۳۹۶۷,۴۶

۷ برآورد TTD

۱.۷ بررسی TTD قطاعی

در رابطه با رویکرد تابع مفصل باید توزیع حاشیه‌ای مدت سفرهای قطاعی را به طور جداگانه مشخص کنیم، براساس نتایج به دست آمده از برآورد هموارسازی هسته‌ای و نمودارهای مستطیلی (شکل مربوطه) توزیع مدت سفر از یک توزیع چندنمایی با سه نقطه‌ای اوج پیروی می‌کنند. از این رو GMM^3 برای برازش دهی مدت سفر در هر دو قطعه‌ای ۱ و ۲ انتخاب شد. نتایج برازش دهی توسط منحنی‌های قرمز رنگ در شکل زیر مشخص شده‌اند.



۲.۷ بررسی TTD مسیر

برآورد TTD مسیر مبنای آماده سازی مقدماتی برای آنالیز TTR است، در کل دو رویکرد برای TTD مسیر وجود دارد. یکی شامل شناسایی بهترین مدل آماری برای برازش دهی مشاهدات مدت سفر و دیگری مدل سازی مشاهدات است. TTD مسیر از یک توزیع چند نمایی پیروی می‌کند. TTD در سطح قطعه پیچیده‌تر از TTD در سطح مسیر است. یک پدیده‌ی توزیع چند نمایی برای مدت سفرهای قطاعی وجود دارد زیرا حساسیت مدت سفر به تقاطع‌ها و تأخیرات ناشی از توقف در فواصل کوتاه‌تر بیشتر است اما مدت سفرها تقریباً یک توزیع تک نمایی را برای کل مسیر با فاصله‌ی طولانی‌تر نشان دادند آزمون‌های $K-S$ برای TTDهای برآورد شده توسط کاپولای FGM، مدل پیچش و GMM^2 اجرا شدند (در جدول زیر ارائه شده‌اند). نتایج برتری تابع مفصل FGM بر روی برازش دهی توزیع و پیچش را تأیید کرده و برتری آماری رویکرد تابع مفصل را به اثبات رسانده‌اند.

جدول ۱۶

Model	test K-S		
	h	p-value	k
FGM(GMM ^۳)	۰	۰.۸۱۴۱	۰.۰۳۴۶
Convolution	۱	۰.۰۰۷۹	۰.۰۶۴۳
GMM ^۲	۰	۰.۴۲۳۸	۰.۰۴۸۲

۳.۷ ارزیابی TTR مسیر

بعد از برآورد TTR TTD، مسیر با استفاده از معادله‌ی (*) محاسبه می‌شود. براساس حداکثر سرعت مجاز در مسیر (۳۵ مایل در ساعت) مدت سفر با جریان آزاد و روان برای کل مسیر ۳۱ ثانیه برآورد شد. نتیجه‌ی برآوردهای TTR مسیر در جدول زیر ارائه شده است. با توجه به نتایج مدل همگشت مقدار TTR مسیر را بالاتر از حد انتظار در نظر گرفت در حالیکه نتیجه‌ی برآورد مدل تابع مفصل تطابق خوبی با مشاهدات تجربی دارد.

۸ دست‌آوردهای پژوهش

مقاله حاضر یک رویکرد مبتنی بر تابع مفصل را برای مشخص کردن ساختار وابستگی بین قطعات یک مسیر معین در شریان‌های شهری معرفی کرد. TTD ‌های مسیر که توسط مدل تابع مفصل برآورد شدند با TTD ‌های برآوردی توسط مدل پیچش در داده‌های AVL مورد مقایسه قرار گرفتند. در پایان TTR مسیر براساس رویکردهای مختلف

جدول ۱۷

k	TTR Path		
	Empirical	۲D-Copula	۲D-Convolution
۱.۵	۰.۳۱۶۹	۰.۳۵۳۳	۰.۳۰۶
۱.۶	۰.۳۶۰۷	۰.۳۹۸۵	۰.۳۴۱۵
۱.۷	۰.۴۰۴۴	۰.۴۵۱۸	۰.۳۸۸
۱.۸	۰.۴۳۹۹	۰.۵۰۴۹	۰.۴۵۶۳
۱.۹	۰.۵۲۴۶	۰.۵۶۱۷	۰.۵۳۲۸
۲.۰	۰.۵۸۴۷	۰.۶۰۸	۰.۵۹۵۶
۲.۱	۰.۶۲۳	۰.۶۵۴۱	۰.۶۳۹۳
۲.۲	۰.۶۸۰۳	۰.۶۹۱۵	۰.۶۸۸۵
۲.۳	۰.۷۴۸۶	۰.۷۲۷۲	۰.۷۳۵
۲.۴	۰.۷۸۱۴	۰.۷۵۲	۰.۷۶۷۸
۲.۵	۰.۸۰۰۵	۰.۷۷۲۵	۰.۷۸۶۹

ارزیابی شده و یافته‌های اصلی به شرح زیر خلاصه شده‌اند:

- مدت سفر قطاعی از نظر مکانی در سایتهای مطالعاتی وابستگی ضعیفی دارند.
- در رابطه با برآورد TTR مسیر مدل تابع مفصل با توزیع حاشیه‌های مدت سفر قطاعی عملکرد خوبی را نشان داد. همچنین برتری آماری مدل تابع مفصل بر مدل پیچش و مدل برازش‌دهی توزیع به اثبات رسید.
- در رابطه با ارزیابی TTR مسیر به هنگام چشمپوشی از تأثیر همبستگی قطعات مشخص شد TTR مسیر توسط مدل پیچش در دامنه‌های خاص به میزان قابل توجهی نادیده گرفته می‌شود. در مقایسه، TTR مسیر و چولگی و عرض برآورد شده با مدل تابع مفصل در هر دو مطالعه موردی به مشاهدات تجربی نزدیکتر بودند.

- [1] Eman and AL-Deek, 2006;Uno et al., 2009; Susilawati et al., 2013.
- [2] Nelsen, R. B. (2006). An Introduction to copulas. Springer series in statistics (2nd ed.,).New York: Springer.
- [3] Van Lint, J.W.C., Van Zuylen, H.J., Tu, H., 2008. Travel time unreliability on freeways: why measures based on variance tell only half the story. Transp. Res. Part A: Policy Pract. 42 (1), 258–277.
- [4] Min Chen, Guizhen Yu, Peng Chen , Yunpeng Wang, 2. A copula-based approach for estimating the travel timereliability of urban arteria.



مشخص سازی های توزیع هندسی از طریق باقیمانده طول عمر

گودرزی، ف^۱

گروه آمار، دانشکده علوم ریاضی، دانشگاه کاشان

چکیده

در این مقاله، برخی مشخص سازی های توزیع هندسی بر اساس دو نوع باقیمانده طول عمر ارائه می شود. ابتدا ما توزیع هندسی را با استفاده از روابط معین گشتاورهای طول عمر بریده شده از زیر، یعنی $\max\{X - t, 0\}$ ، که در آن X یک متغیر تصادفی صحیح مقدار مثبت است، مشخص سازی می کنیم. سپس با استفاده از روابط معین گشتاورهای باقیمانده عمر در زمان t ی یک فرایند تجدید، توزیع مشترک زمانهای بین ورود مشخص سازی می شود.

کلمات کلیدی: مشخص سازی، توزیع نمایی، توزیع هندسی، باقیمانده طول عمر.

^۱f-goodarzi@um.ac.ir

۱ پیش‌گفتار

مشخص است که تنها توزیع‌های نمایی و هندسی دارای خاصیت فاقد حافظه بودن

$$P(X > t + x | X > t) = P(X > x), \quad x, t > 0, \quad (1)$$

هستند. در طی دهه‌های گذشته، مطالعات متعددی روی مشخص‌سازی توزیع نمایی (از میان تمامی توزیع‌های مطلقاً پیوسته نامنفی) تحت برخی شرایط ضعیف‌تر از (1) روی داده است. برای مثال، دالاس [1]، برای تعیین اینکه X به طور نمایی توزیع می‌شود، از

$$E[(X - t)^n | X > t] = c, \quad t > 0,$$

که در آن $n \in \mathcal{N}^+$ و c یک ثابت است، استفاده نمود. در اینجا $\mathcal{N}^+$ و $\mathcal{N}$ به ترتیب نشان دهنده‌ی اعداد صحیح مثبت و نامنفی هستند. نتایج مرتبط را می‌توان در کتاب رائو و شانباگ [8] یافت. لین [5] روابط معین گشتاورهایی از نوع دیگر باقیمانده عمر $(X - t)_+ = \max[X - t, 0]$ را برای توسیع مشخص‌سازی توزیع نمایی، به کار برد. در زبان علوم بیمه، n امین گشتاور $E[(X - t)_+^n]$ تبدیل مازاد زیان درجه n ام مخاطره‌ی X نامیده شده و یک اندازه استاندارد خطر X است، برای مثال، دنویت و همکاران [2] را ببینید. چندین مشخص‌سازی مشابه برای توزیع هندسی وجود دارد. برای مثال، وقتی X یک متغیر تصادفی صحیح مقدار مثبت است و $c > 1$ یک ثابت است، هر دو

$$E[X - t | X > t] = E(X), \quad t \in \mathcal{N}, \quad (2)$$

و

$$E[(X - s)_+]E[(X - t)_+] = cE[(X - s - t)_+], \quad s, t \in \mathcal{N}, \quad (3)$$

می‌توانند X را مشخص‌سازی کنند. مشخص‌سازی‌های مرتبط توزیع هندسی بر اساس روابط خاص در میان تابع جرم احتمال، تابع توزیع، تابع نرخ خطر و میانگین باقیمانده عمر را می‌توان در سودهیش و نیر [9] و نیر و سانکاران [6] یافت.

از طرف دیگر، همچنین چندین مشخص‌سازی توزیع‌ها در فرایندهای تجدید بر اساس خاصیت فاقد حافظه بودن وجود دارد. فرض کنید γ_t باقیمانده عمر در زمان t فرایند تجدید $\{A(t); t \geq 0\}$ باشد که دارای $F(0) = 0$

به عنوان تابع توزیع مشترک زمانهای بین ورود است. سینلار و جاگرز (۱۹۷۳)، گوپتا و گوپتا (۱۹۸۶) ثابت کردند که، برای G ملایم و یک ثابت c

$$E[G(\gamma_t)] = c, \quad t > 0, \quad (4)$$

می تواند دلالت کند که F تابع توزیع نمایی است.

متعاقباً، رائو و شانباگ (۱۹۹۴ صفحه ۲۱۳) نشان داد که (4) با G کلی تر می تواند مشخص سازی کن که F تابع توزیع نمایی یا هندسی است. علاوه بر این، هوانگ و همکاران (۱۹۹۳) و هوانگ و لی (۱۹۹۳) به ترتیب، به کار بردند

$$E[G(\gamma_t)|A(t) = n] = c, \quad t > 0.$$

و

$$Var[\gamma_t] = c, \quad t > 0.$$

برای مشخص سازی اینکه F تابع توزیع نمایی یا هندسی است. در این جا، توزیع هندسی را با تابع جرم احتمال

$$P(X = x) = p(1-p)^{x-1}, \quad x \in \mathcal{N}^+ \quad (5)$$

نشان می دهیم که در آن $0 < p < 1$.

۲ مشخص سازی ها بر اساس گشتاورهای عمر بریده شده

در سر تا سر این مقاله فرض می کنیم X یک متغیر تصادفی صحیح مقدار مثبت و $\phi(\theta)$ تبدیل لاپلاس استیلتیس $P(X = x)$ باشد، یعنی

$$\phi(\theta) = \sum_{x=1}^{\infty} e^{-\theta x} P(X = x), \quad \theta > 0.$$

برای هر $n \in \mathcal{N}$ ، فرض کنید $\phi^{(n)}(\theta)$ ، n امین مشتق $\phi(\theta)$ باشد. همچنین برای راحتی نشان می دهیم $\phi^{(1)}(\theta) = \phi'(\theta)$ ، $\phi^{(2)}(\theta) = \phi''(\theta)$ و $\phi^{(3)}(\theta) = \phi'''(\theta)$.

از طرف دیگر، برای $q(t)$ ، به ازای هر $t \in \mathcal{N}$ ، $L(q(t))$ را به عنوان تبدیل لاپلاس استیلتیس $q(t)$ در نظر می گیریم،

یعنی

$$L(q(t)) = \sum_{t=0}^{\infty} e^{-\theta t} q(t), \quad \theta \geq 0.$$

حال، برای $n \in \mathcal{N}$ ، تعریف می‌کنیم

$$q_1(t; n) = \sum_{x=t+1}^{\infty} x^n P(X=x), \quad t \in \mathcal{N} \text{ هر برای}, \quad (6)$$

$$q_2(t; n) = t \sum_{x=t+1}^{\infty} x^n P(X=x), \quad t \in \mathcal{N} \text{ هر برای}, \quad (7)$$

$$q_3(t; n) = t^2 \sum_{x=t+1}^{\infty} x^n P(X=x), \quad t \in \mathcal{N} \text{ هر برای}, \quad (8)$$

$$q_4(t; n) = t^3 \sum_{x=t+1}^{\infty} x^n P(X=x), \quad t \in \mathcal{N} \text{ هر برای}. \quad (9)$$

در زیر با بیان لم‌هایی، تبدیلات لاپلاس استیلیس (6) - (9) را به ترتیب محاسبه می‌کنیم.

لم 1

$$L(q_1(t; n)) = \frac{\mu_n - (-1)^n \phi^{(n)}(\theta)}{1 - e^{-\theta}}, \quad \theta \geq 0. \quad (10)$$

لم 2

$$L(q_2(t; n)) = \frac{e^{-\theta} \mu_n - (-1)^n e^{-\theta} \phi^{(n)}(\theta) + (-1)^n (1 - e^{-\theta}) \phi^{(n+1)}(\theta)}{(1 - e^{-\theta})^2}, \quad \theta \geq 0. \quad (11)$$

لم 3

$$L(q_3(t; n)) = \frac{1}{1 - e^{-\theta}} \left(\mu_n - (-1)^n \phi^{(n+2)}(\theta) - 2(-1)^n \phi^{(n+1)}(\theta) - (-1)^n \phi^{(n)}(\theta) \right. \\ \left. + 2L(q_2(t; n)) - L(q_1(t; n)) \right), \quad \theta \geq 0. \quad (12)$$

اگر X دارای توزیع $Geo(p)$ باشد آنگاه می‌توان به دست آورد

$$E[(X - t)_+] = \frac{1}{p} (1 - p)^t, \quad t \in \mathcal{N} \text{ هر برای}, \quad (13)$$

$$E[(X - t)_+]^2 = \frac{2 - p}{p^2} (1 - p)^t, \quad t \in \mathcal{N} \text{ هر برای}. \quad (14)$$

پریا و همکاران [7]، با انگیزه گرفتن از (2) و (3) ثابت کردند که به شرط $t \in \mathcal{N}$ ، توزیع X را می‌توان به گونه‌ای تعیین کرد که

$$P\{X \geq t\} = E((X - t + 1)_+) - E((X - t)_+), \quad t \in \mathcal{N}^+, \quad (15)$$

و این منجر می‌شود که (13) در حقیقت یک مشخص‌سازی از توزیع هندسی باشد. با وجود این هنوز یک فرض اندکی ضعیف‌تر از (13) برای مشخص‌سازی توزیع هندسی در زیر به کار برده می‌شود. فرض کنید

$$E[(X - t)_+] = c_1 c_2^t, \quad t \in \mathcal{N} \text{ هر برای}, \quad (16)$$

که در آن c_1 و c_2 ثابت هستند. آنگاه $c_1 = \mu_1 > 1$ و $c_2 = 1 - \frac{1}{\mu_1} > 0$ دارای توزیع $Geo(\frac{1}{c_1})$ است. مجدداً تلاش می‌کنیم توزیع هندسی را با استفاده از یک فرض ضعیف‌تر از (14) مشخص‌سازی کنیم. هنوز به یک فرض اضافی برای اطمینان از این مشخص‌سازی، نیازمندیم.

قضیه فرض کنید

$$E[(X-t)_+]^2 = c_1 c_2^t, \quad t \in \mathcal{N} \text{ هر برای}, \quad (17)$$

که در آن c_1 و c_2 ثابت هستند. همچنین فرض کنید

$$\mu_2 = 2\mu_1^2 - \mu_1. \quad (18)$$

آنگاه $c_1 = \mu_2 > 1 + \frac{(1-2\mu_1)}{\mu_2} > 0$ و $c_2 = 1 + \frac{(1+c_1-c_1c_2)}{\mu_2}$ دارای توزیع $Geo(\frac{1}{c_1})$ است.

برهان. از (17)، داریم $c_1 = \mu_2 > 1$ و

$$\sum_{x=t+1}^{\infty} (x-t)^2 P(X=x) = \mu_2 c_2^t \quad t \in \mathcal{N} \text{ هر برای}. \quad (19)$$

با گرفتن تبدیلات لاپلاس استیلتیس از هر دو طرف (19) نسبت به t و استفاده از لم‌های 1 تا 3 نتیجه می‌گیریم

$$\begin{aligned} & \frac{\mu_2 - \phi''(\theta)}{1 - e^{-\theta}} - \frac{2(\mu_1 e^{-\theta} - (1 - e^{-\theta})\phi''(\theta) + e^{-\theta}\phi'(\theta))}{(1 - e^{-\theta})^2} \\ & + \frac{(e^{-\theta} + e^{-2\theta})(1 - \phi(\theta))}{(1 - e^{-\theta})^3} + \frac{2e^{-\theta}\phi'(\theta)}{(1 - e^{-\theta})^2} - \frac{\phi''}{(1 - e^{-\theta})} \\ & = \frac{\mu_2}{1 - c_2 e^{-\theta}}, \quad \theta \geq 0. \end{aligned} \quad (20)$$

همچنین پس از برخی ساده‌سازی‌ها داریم

$$\phi(\theta) = \frac{(1 + \mu_2 - 2\mu_1 - c_2 \mu_2) + (\mu_2 - c_2 - 2c_2 \mu_1 - c_2 \mu_2)e^{-2\theta}}{(1 - c_2 e^{-\theta})(1 + e^{-\theta})} \quad (21)$$

$$+ \frac{(1 - 2\mu_2 - c_2 + 2\mu_1 + 2c_2 \mu_1 + 2c_2 \mu_2)e^{-\theta}}{(1 - c_2 e^{-\theta})(1 + e^{-\theta})}, \quad \theta \geq 0. \quad (22)$$

چون $\lim_{\theta \rightarrow \infty} \phi(\theta) = P(X \leq \infty) = 0$ ، داریم

$$c_2 = 1 + \frac{1}{\mu_2} - \frac{2\mu_1}{\mu_2} > 0. \quad (23)$$

این همراه با $c_1 = \mu_2$ و (18) می‌دهد

$$\phi(\theta) = \frac{(1/\mu_1)e^{-\theta}}{1 - (1/\mu_1)e^{-\theta}}, \quad \theta \geq 0, \quad (24)$$

□

و ادعا بلافاصله نتیجه می‌شود.

فرض اضافی (18) یک رابطه طبیعی برقرار شده توسط اولین دو گشتاور توزیع هندسی است. این فرض به عنوان یک شرط لازم برای مشخص سازی تابع مشخصه توزیع هندسی استفاده می شود. بعلاوه، سودهیش و نیر [9] تعمیمی از (18)، یعنی

$$Var[X|X > t] = (E(X - t|X > t))^2 - E[X - t|X > t], \quad t \in \mathcal{N} \text{ هر برای} \quad (25)$$

را برای مشخص سازی توزیع هندسی به کار بردند.

تذکر ۱.۲. می توان (18) را به عنوان محدودیتی از c_1 و c_2 نگریست. به طور دقیقتر، به شرط $c_1 = \mu_2$ و (23)، می توان دریافت که (18) و

$$(1 - c_2)(c_1 - c_1 c_2 + 1) = 2 \quad (26)$$

معادل هستند. در نتیجه، در قضیه 2، همراه با مراحل اثبات، ادعا هنوز نتیجه می شود وقتی (18) توسط (26)، جایگزین می شود. توجه کنید که برای متغیر تصادفی X از توزیع $Geo(p)$ با $(c_1, c_2) = (\frac{1-p}{p}, 1-p)$ ، (26) برقرار است. لیکن، وقتی شرط اضافی (18) حذف می شود، تنها به دست می آوریم

$$\phi(\theta) = \frac{2(\mu_2 - 2\mu_1^2 + \mu_1)(1 - e^{-\theta})e^{-\theta}}{\mu_2(1 - c_1 e^{-\theta})(1 + e^{-\theta})} + \frac{(2\mu_1 - 1)e^{-\theta}}{\mu_2(1 - c_2 e^{-\theta})}, \quad \theta \geq 0, \quad (27)$$

که این شبیه به تبدیل لاپلاس استیلتیس آمیخته توزیع هندسی و توزیع نامعلوم دیگر است.

تذکر ۲.۲. طبیعی است که، توزیع هندسی را توسط این فرض کلی مشخص سازی کنیم که، برای $n \in \mathcal{N}$

$$E((X - t)_+^n) = c_1 c_2^t, \quad t \in \mathcal{N} \text{ هر برای}, \quad (28)$$

که در آن c_1 و c_2 ثابت هستند. توجه کنید که برای X با توزیع $Geo(p)$ داریم

$$E((X - t)_+^n) = E(X^n)(1 - p)^t, \quad t \in \mathcal{N} \text{ هر برای}. \quad (29)$$

شبیه به اثبات قضیه 2، لازم است تبدیلات لاپلاس استیلتیس

$$t^{n-i} \sum_{x=t+1}^{\infty} x^i P(X = x), \quad t \in \mathcal{N} \text{ هر برای}, \quad (30)$$

را برای $0 \leq i \leq n$ را یافته و سپس با استفاده از (28) یک معادله دیفرانسیل n ام $\phi(\theta)$ به دست آوریم. می توان حدس زد که معادله دیفرانسیل برای $n \geq 3$ پیچیده تر است و حل آن توسط یک معادله اضافی مشابه با (18) یا (26)، ساده نیست.

پریا و همکاران [7] گشتاور فاکتوریل صعودی r ام X تعریف شده توسط

$$E((X-t)_+^{[r]}) = \sum_{x=t+1}^{\infty} (x-t) \dots (x-t+r-1) P(X=x) \quad t \in \mathcal{N}, r \in \mathcal{N}^+, \quad (31)$$

را در نظر گرفتند و نشان دادند که، برای $r \geq 2$ ، تابع $E((X-t)_+^{[r]})$ ، و نیز $E((X-t)_+)$ توزیع X را تعیین می کند. به ویژه، می توان (31) را با $r=2$ به صورت $E((X-t)_+^{[2]}) = E((X-t)_+^2) + E((X-t)_+)$ برای $t \in \mathcal{N}$ بیان کرد. از این، همراه با (13) و (14)، می توان دید که

$$E((X-t)_+^2) + E((X-t)_+) = \frac{2(1-p)^t}{p^2}, \quad t \in \mathcal{N} \text{ برای هر } ,$$

مشخص سازی از توزیع $Geo(p)$ است. علاوه بر این، برای $r \in \mathcal{N}$ ، پریا و همکاران [7] توزیع هندسی را با استفاده از $E((X-t-1)_+^{[r]})$ متناسب با $E((X-t)_+^{[r]})$ ، به ازای هر $t \in \mathcal{N}$ مشخص سازی کردند.

با انگیزه گرفتن از این نتایج، مشاهده می کنیم که $E((X-t)_+^2) = (2/p-1)E((X-t)_+)$ به ازای هر $t \in \mathcal{N}$ ، برای X که به صورت $Geo(p)$ توزیع می شود و بالعکس دارای قضیه ذیل است. فرض کنید، به ازای هر $t \in \mathcal{N}$ ،

$$E((X-t)_+^2) = cE((X-t)_+), \quad (32)$$

که در آن c یک ثابت است. آنگاه $c = \frac{\mu_2}{\mu_1} > 1$ و X دارای توزیع $Geo(\frac{2}{1+c})$ است.

برهان. از (32)، داریم $c = \frac{\mu_2}{\mu_1}$ و به ازای هر $t \in \mathcal{N}$ ،

$$\sum_{x=t+1}^{\infty} (x-t)^2 P(X=x) = c \sum_{x=t+1}^{\infty} (x-t) P(X=x). \quad (33)$$

حال با گرفتن تبدیلات لاپلاس استیلیس از هر دو طرف (33) نسبت به t و استفاده از لم های 1 تا 3 خواهیم داشت

$$\begin{aligned} & \frac{\mu_2 - \phi''(\theta)}{1 - e^{-\theta}} - \frac{2(\mu_1 e^{-\theta} + e^{-\theta} \phi'(\theta) - (1 - e^{-\theta}) \phi''(\theta))}{(1 - e^{-\theta})^2} \\ & + \frac{(e^{-\theta} + e^{-2\theta})(1 - \phi(\theta))}{(1 - e^{-\theta})^2} + \frac{2e^{-\theta} \phi'(\theta)}{(1 - e^{-\theta})^2} - \frac{\phi''(\theta)}{1 - e^{-\theta}} \\ & = c \frac{\mu_1 + \phi'(\theta)}{1 - e^{-\theta}} - c \frac{e^{-\theta} + (1 - e^{-\theta}) \phi'(\theta) - e^{-\theta} \phi(\theta)}{(1 - e^{-\theta})^2}, \quad \theta \geq 0. \end{aligned} \quad (34)$$

پس از برخی ساده سازی ها، به دست می آوریم که

$$\phi(\theta) = \frac{\mu_1 - 2\mu_1^2 + \mu_2 + (\mu_1 + 2\mu_1^2 - \mu_2)e^{-\theta}}{\mu_1(1 - e^{-\theta}) + \mu_2(1 - e^{-\theta})}, \quad \theta \geq 0. \quad (35)$$

از طرف دیگر، چون

$$\lim_{\theta \rightarrow \infty} \phi(\theta) = P(X \leq \infty) = 1,$$

داریم

$$\mu_2 = 2\mu_1^2 - \mu_1,$$

و این به نوبه خود نتیجه می دهد

$$\phi(\theta) = \frac{(1/\mu_1)e^{-\theta}}{1 - (1 - 1/\mu_1)e^{-\theta}}, \quad \theta \geq 0,$$

□

و ادعا بلافاصله نتیجه می شود.

در حقیقت وقتی X به صورت $Geo(p)$ توزیع می شود و $n \in \mathcal{N}^+$ ، از (30)، به ازای هر $t \in \mathcal{N}$ داریم

$$E((X - t)_+^{n+1}) = a_n E((X - t)_+^n), \quad (36)$$

که در آن $a_n = \frac{E(X^{n+1})}{E(X^n)}$ تنها به p بستگی دارد. توجه کنید که a_n در (36) پیچیده تر می شود همانطور که n افزایش می یابد (برای مثال مشاهده کنید که

$$a_1 = \frac{2}{p} - 1, \quad a_2 = \frac{6 - 6p + p^2}{p(2 - p)} \quad \text{و} \quad a_3 = \frac{24 - 36p + 14p^2 - p^3}{p(6 - 6p + p^2)}, \quad (37)$$

که از چهار گشتاور نخست توزیع $Geo(p)$ نتیجه می شود). جالب است توزیع هندسی را بر اساس (36) همانند در قضیه 3 مشخص سازی کنیم. لیکن، همانطور که در توجه 2.2 ذکر شد و نظر به (37) به نظر می رسد که حالت $n = 2$ می تواند قابل حل باشد. نتیجه ذیل به روشی بسیار شبیه به قضیه 2 ثابت می شود.

قضیه 4 فرض کنید، به ازای هر $t \in \mathcal{N}$

$$E((X - t)_+^3) = c E((X - t)_+^2), \quad (38)$$

که در آن c یک ثابت است. همچنین فرض کنید

$$\mu_2 = 2\mu_1^2 - \mu_1.$$

آنگاه $c = \frac{\mu_2}{\mu_1} > 1$ و X دارای توزیع $Geo((c + 3 - \sqrt{c^2 + 3})/(c + 1))$ است.

۳ مشخص سازی های فرایند تجدید هندسی

فرض کنید $\{X_i; i \geq 1\}$ یک دنباله از متغیرهای تصادفی مستقل و هم توزیع با یک تابع توزیع مشترک F باشد. همچنین فرض کنید $S_0 = 0$, $S_n = \sum_{i=1}^n X_i$, $n \geq 1$ و برای $t \geq 0$, $A(t) = \sup\{n | S_n \leq t\}$. مشخص است که $\{A(t); t \geq 0\}$ یک فرایند تجدید است. $\gamma_t = S_{A(t)+1} - t$ را باقیمانده عمر در t نشان می دهیم. برای F مطلقاً پیوسته با $F(0) = 0$, هیونگ و چانگ [3] نشان داد که برای $t \geq 0$, $E[\gamma_t] = E[\gamma_1]$, $Var[\gamma_t] = Var[\gamma_1]$ می تواند X_1 را مشخص سازی کند که به طور نمایی توزیع می شود.

با این انگیزه، حالت F تعریف شده روی $\mathcal{N}^+$ را مطالعه می کنیم. برای راحتی در نمادگذاری فرض می شود $\{X_i; i \geq 1\}$ هم توزیع با X هستند. همانطور که آن بعداً مورد استفاده قرار می گیرد، ما لم ذیل از نتیجه یک هوانگ و همکاران [4] را ارائه می دهیم. فرض کنید $h(x)$ یک تابع نامنفی با $h(0) = 0$ باشد. فرض کنید به ازای هر $t \in \mathcal{N}$ فرض کنید به ازای هر

$$\sum_{x=t+1}^{\infty} h(x-t)P(X=x) = \theta(1-F(t)),$$

که در آن $\theta > h(1)$ یک ثابت است. آنگاه $P(X=t) = p^{t-1}(1-p)$, که در آن $0 < p < 1$ توسط $h(i) + (1-p)^i = \theta$ برای هر $i \geq 1$ تعریف می شود. وقتی X_1 به صورت $Geo(p)$ توزیع می شود، $Var(\gamma_t) = (2-p)E[\gamma_1]$ ، برای هر $t \in \mathcal{N}$ برقرار است، که معادل است با $E(\gamma_t^2) = (2-p)E(\gamma_t)$ ، برای هر $t \in \mathcal{N}$. بالعکس، ما نتیجه ذیل را داریم. قضیه 5 فرض کنید برای هر $t \in \mathcal{N}$,

$$E[\gamma_t^2] = cE(\gamma_t) \quad (39)$$

که در آن c یک ثابت است. همچنین فرض کنید

$$E(\gamma_1) = \mu_1 \quad (40)$$

آنگاه $2 < c = 2 - \frac{1}{\mu_1} < 2$ و X_1 به صورت $Geo(2-c)$ توزیع می شود.

برهان. فرض کنید $g(t) = E(\gamma_t)$.

می‌توان به دست آورد که برای هر $t \in \mathcal{N}$

$$\begin{aligned} g(t) &= E(E(S_{A(t)+1} - t) | X_1) \\ &= \sum_{x=1}^t E(S_{A(t)+1} - (t-x))P(X_1 = x) + \sum_{x=t+1}^{\infty} (x-t)P(X_1 = x) \\ &= \sum_{x=1}^t g(t-x)P(X_1 = x) + \sum_{x=t+1}^{\infty} (x-t)P(X_1 = x), \end{aligned} \quad (41)$$

که در آن $\sum_{x=1}^{\infty} = 0$ تعریف می‌شود. با گرفتن تبدیلات لاپلاس از هر دو طرف در (41)، نسبت به t و پس از برخی محاسبات، می‌دهد

$$(L(g(t)) + \frac{e^{-\theta}}{(1-e^{-\theta})^2})(1-\phi(t)) = \frac{\mu_1}{1-e^{-\theta}}, \quad \theta \geq 0. \quad (42)$$

از طرف دیگر، از (39)، برای هر $t \in \mathcal{N}$ به دست می‌آوریم

$$\begin{aligned} cg^{\vee}(t) &= E(E((S_{A(t)+1} - t)^{\vee} | X_1)) \\ &= \sum_{x=1}^t E((S_{A(t)+1} - (t-x))^{\vee})P(X_1 = x) + \sum_{x=t+1}^{\infty} (x-t)^{\vee}P(X_1 = x) \\ &= \sum_{x=1}^t cg^{\vee}(t-x)P(X_1 = x) + \sum_{x=t+1}^{\infty} (x-t)^{\vee}P(X_1 = x). \end{aligned} \quad (43)$$

مجدداً با گرفتن تبدیلات لاپلاس از هر دو طرف در (43) نسبت به t و پس از برخی ساده‌سازی‌ها خواهیم داشت،

$$\left(cL(g^{\vee}(t)) - \frac{e^{-\theta} + e^{-2\theta}}{(1-e^{-\theta})^2} \right) (1-\phi(\theta)) = \frac{\mu_2}{1-e^{-\theta}} - \frac{2\mu_1 e^{-\theta}}{(1-e^{-\theta})^2}, \quad \theta \geq 0. \quad (44)$$

نظر به (41) و (44)، داریم

$$cL(g^{\vee}(t)) = \left(\frac{\mu_2}{\mu_1} - \frac{2e^{-\theta}}{1-e^{-\theta}} \right) L(g(t)) + \left(1 + \frac{\mu_2}{\mu_1} \right) \frac{e^{-\theta}}{(1-e^{-\theta})^2}, \quad \theta \geq 0. \quad (45)$$

از (45) و قضیه یکتایی، نتیجه می‌دهد، به ازای $t \in \mathcal{N}$

$$g^{\vee}(t) = \frac{1}{c} \frac{\mu_2}{\mu_1} g(t) - \frac{2}{c} \sum_{x=1}^t g(x) + \frac{1}{c} \left(1 + \frac{\mu_2}{\mu_1} \right) t.$$

این به نوبه خود نتیجه می‌دهد

$$g^{\vee}(t+1) - g^{\vee}(t) = \frac{1}{c} \frac{\mu_2}{\mu_1} (g(t+1) - g(t)) - \frac{2}{c} g(t+1) + \frac{1}{c} \left(1 + \frac{\mu_2}{\mu_1} \right). \quad (46)$$

به وضوح، جایگزین کردن $t = 0$ در (39) خواهد داد

$$c = \frac{\mu_2}{\mu_1} > 1. \quad (47)$$

شرط (۴۰)، همراه با جایگزین کردن $t = ۰$ در (۴۶)، منجر می‌شود به

$$c = ۲ - \frac{۱}{\mu_1} < ۲. \quad (۴۸)$$

از (۴۶) تا (۴۸) داریم

$$g'(t+1) + \left(\frac{۲\mu_1}{۲\mu_1-1} - \mu_1 \right) g(t+1) - g'(t) + \mu_1 g(t) - \frac{۲\mu_1^2}{۲\mu_1-1} = ۰. \quad (۴۹)$$

فرض کنید برای یک $k \geq ۱$ ، $g(k) = \mu_1$ ، از (۴۹)، جوابهای

$$g(k+1) = \mu_1, \quad (۵۰)$$

و

$$g(k+1) = -\frac{۲\mu_1}{۲\mu_1-1}, \quad (۵۱)$$

را نتیجه می‌دهد. چون به ازای هر $t \in \mathcal{N}$ ، $g(t) > ۰$ ، مشاهده می‌شود که (۵۱) مستثنی می‌شود. از این رو، با استفاده از استقرای ریاضی نتیجه می‌شود که به ازای هر $t \in \mathcal{N}$ ، $g(t) = \mu_1$ ، در نتیجه، برای هر $t \in \mathcal{N}$ (۴۱) می‌دهد

$$\sum_{x=t+1}^{\infty} (x-t)P(X_1=x) = P(X_1>t)\mu_1$$

و با استفاده از لم ۱ ادعا نتیجه می‌شود.

□

مراجع

- [1] Dallas, A.C. (1979), *On the exponential law*, *Metrika* 26(1), 105–108.
- [2] Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2006), *Actuarial theory for dependent risks: measures, orders and models*, John Wiley & Sons, New York.
- [3] Huang, W.J., Chang, W.C. (2000), *On a study of the exponential and poisson characteristics of the poisson process*, *Metrika* 50(3), 247–254.
- [4] Huang, W.J., Li, S.H., Su, J.C. (1993), *Some characterizations of the poisson process and geometric renewal process*, *J Appl Probab* 30, 121–130.

- [5] Lin, G.D. (2003), *Characterizations of the exponential distribution via the residual lifetime*, Sankhya 65, 249–258.
- [6] Nair, N.U., Sankaran, P.G. (2013), *Characterizations of discrete distributions using reliability concepts in reversed time*, Stat Probab Lett 83(9), 1939–1945.
- [7] Priya, P., Sankaran, P.G., Nair, N.U. (2000), *Factorial partial moments and their properties*, J Indian Stat Assoc 38, 45–55.
- [8] Rao, C.R., Shanbhag, D.N. (1994), *Choquet-Deny type functional equations with applications to stochastic models*, Wiley, New York.
- [9] Sudheesh, K.K., Nair, N.U. (2010), *Characterization of discrete distributions by conditional variance*, Metron 68(1), 77–85.

کاربرد ترتیب‌های تصادفی در بیمه: چالش‌ها و نیازها

محمودوند، ر^۱

گروه آمار، دانشکده علوم پایه، دانشگاه بوعلی سینا

چکیده

بیمه به عنوان یکی از راهکارهای مدیریت ریسک، نیازمند استفاده از علوم مختلفی است. علم اکچوئریال بر پایه روش‌های احتمالی، ریاضی و اقتصادی محاسبات فنی بیمه را انجام می‌دهد. در این مطالعه پس از مروری کوتاه بر اهداف اصلی اکچوئریال در حوزه بیمه به بیان نقش نظریه قابلیت اعتماد در این زمینه می‌پردازیم. به طور خاص با تمرکز بر مدل‌های ریسک فردی و جمعی، در خصوص بکارگیری ترتیب‌های تصادفی مطالبی را ارائه می‌کنیم. در پایان با مرور چند مقاله اخیر به طرح چند پرسش در این زمینه می‌پردازیم.

کلمات کلیدی: اکچوئریال، توزیع زیان، نظریه قابلیت اعتماد، ترتیب‌های تصادفی.

^۱ r.mahmoudvand@basu.ac.ir

۱ مقدمه

گسترش خدمات و فناوری‌ها هرچند رفاه فراوانی برای بشر به همراه دارد، باعث به وجود آمدن و تنوع ریسک‌ها نیز خواهد شد. به عنوان مثال با وجود آسایشی که اتومبیل در چرخه زندگی بشر ایجاد کرده است، خسارت‌های جانی و مالی فراوانی هم به دنبال دارد. بنابراین به صورت شخصی و اجتماعی ناگزیر از برخورد با ریسک در زندگی روزمره هستیم. کنترل ریسک و تامین مالی آن به عنوان دو راهکار متداول در این زمینه مورد توجه است. به منظور انتخاب روش صحیح برای برخورد با ریسک، باید ارزیابی صحیحی از آن داشته باشیم. از حدود یک قرن گذشته تا به امروز شیوه‌های مختلفی برای این کار معرفی و به کار گرفته شده است که عمدتاً بر پایه روش‌های آماری استوار هستند. قاسمی و محمودوند (۲۰۱۳) مرور نسبتاً جامعی به این روش‌ها ارائه کرده‌اند که می‌تواند برای بکارگیری در عمل مفید باشد.

بیمه به عنوان یکی از راهکارهای تامین مالی ریسک، در شرایطی که دارنده ریسک قادر به تامین آن نباشد، پیشنهاد می‌شود. در مطالعه بیمه چند مولفه به صورت مستقیم درگیر هستند: بیمه‌گذار (فردی حقیقی یا حقوقی که قرارداد بیمه با وی تنظیم می‌شود)، ذینفع یا ذینفعان بیمه، موضوع بیمه (ریسک)، بیمه‌گر (شرکت بیمه) و حق بیمه (مبلغی که بیمه‌گذار به بیمه‌گر می‌پردازد). در اکثر قراردادهای بیمه‌ای انفرادی، بیمه‌گذار و ذینفع یکی است اما در حالت کلی ممکن است چنین نباشد. قابل ذکر است که محاسبات فنی مرتبط با ریسک و بیمه در علوم اکچوئریال و مالی مورد بررسی قرار می‌گیرد و فردی که چنین توانایی دارد اکچوئری نامیده می‌شود. در چنین فضایی برخی از مهم‌ترین کارکردهای اکچوئری‌ها عبارت است از:

- بررسی بیمه‌پذیری ریسک‌ها

- تعیین حق بیمه مناسب برای پوشش ریسک‌ها

- بررسی توانگری مالی بیمه‌گر

در هر سه مورد اکچوئری بایستی توزیع خسارت‌های مربوط به ریسک را مطالعه کند. اما تنوع ریسک‌ها و ماهیت متفاوت آنها باعث شده است تا اکچوئری‌ها گستره بسیار وسیعی از توزیع‌های احتمالی را مطالعه و بررسی کنند. به عنوان مثال در بیمه‌های زندگی، متغیر اساسی طول عمر است و لذا توزیع‌هایی که برای طول عمر مطرح شده است می‌بایست مورد استفاده قرار بگیرد. در بیمه‌های غیرزندگی با توجه به آنکه معمولاً زمان قراردادهای کوتاه است، توجه

به توزیع طول عمر اهمیت کمتری دارد. در عوض توزیع مبالغ خسارت‌ها و تعداد خسارت‌ها اهمیت بیشتری پیدا می‌کند.

در نظریه قابلیت اعتماد توزیع‌های مختلفی برای مطالعه سیستم‌های طول عمر وسایل یا به طور کلی مجموعه‌ای از مولفه‌ها معرفی و به کار گرفته شده است. این توزیع‌ها معمولاً بر حسب کمیت‌هایی مانند نرخ‌های خطر و توزیع شرطی طول عمر باقیمانده دسته‌بندی می‌شوند. به عنوان مثال توزیع‌های با نرخ خطر افزایشی (IFR)، با نرخ خطر کاهشی (DFR)، توزیع‌های با میانگین طول عمر باقیمانده افزایشی (IMRL) و توزیع‌های با میانگین طول عمر باقیمانده کاهشی (DMRL) از جمله رده توزیع‌های متداولی است که بسیار مورد بررسی قرار گرفته‌اند. البته باید اذعان کنیم که با معرفی نظریه قابلیت اعتماد، به لحاظ تاریخی تلاش‌های نظری بسیاری برای توسعه کاربردهای این شاخه در قرن بیستم و دو دهه اخیر صورت گرفته است. فهرست منابع موجود به اندازه‌ای طولانی است که ذکر آن بسیار خسته‌کننده و در این مقاله مقدور نیست. با این وجود اوشاکف (۲۰۱۲) فهرست جامعی به تفکیک بخش‌های مختلفی از کاربردهای نظریه قابلیت اعتماد ارائه کرده است که خواننده علاقمند می‌تواند به آنها رجوع کند. به منظور تعیین جایگاه قابلیت اعتماد در حوزه بیمه، در این مقاله ابتدا به معرفی مهم‌ترین مدل‌های ارزیابی ریسک که در بیمه رایج هستند می‌پردازیم. سپس به برخی از پژوهش‌هایی که محققین حوزه قابلیت اعتماد برای مقاصد بیمه‌ای در سال‌های اخیر انجام داده‌اند می‌پردازیم.

۲ روش‌های رایج برای ارزیابی ریسک به شیوه کمی در بیمه

ابتدا متذکر می‌شویم که منظور از ارزیابی ریسک، تحلیل و ارزشیابی ریسک است. از طرفی بخش فنی ارزیابی ریسک همان تحلیل ریسک است که خود شامل شناسایی، توصیف و برآورد است. در حوزه روش‌های کمی دو مدل مناسب برای تحلیل ریسک شامل مدل‌های ریسک فردی و مدل‌های ریسک جمعی است. در ادامه به اختصار به معرفی این دو مدل پرداخته می‌شود.

۱.۲ مدل ریسک فردی

فرض کنید پرتفوی تحت بررسی شامل n ریسک است و متغیر تصادفی X_i نشان‌دهنده کل مجموع زیان‌های مربوط به ریسک i ام در بازه $[0, t]$ باشد. تحت چنین شرایطی کل زیان مربوط به این پرتفوی از رابطه زیر به دست

می آید:

$$S(t) = X_1 + X_2 + \dots + X_n$$

۲.۲ مدل ریسک جمعی

در این مدل مجموع شدت زیان‌ها به صورت زیر است:

$$S(t) = Y_1 + Y_2 + \dots + Y_{N(t)}$$

که در آن Y_i نشان‌دهنده‌ی خسارت i ام در پرتفوی است و زمانی که $N(t) = 0$ باشد بطور طبیعی مقدار زیان برابر صفر تعریف می‌شود؛ یعنی $S(t) = 0$. قابل ذکر است که اگر متغیر تصادفی $N(t)$ از توزیع پواسون تبعیت کند، توزیع متغیر تصادفی $S(t)$ را پواسون مرکب می‌نامند.

۳.۲ ترکیب دو مدل جمعی و فردی

اجازه دهید این موضوع را با مثالی شرح دهیم. در یک مجموعه مانند پالایشگاه ریسک‌های مختلفی وجود دارد که از حیث ماهیت رخداد و نیز تبعات هر ریسک با همدیگر متفاوت هستند. به عنوان مثال ریسک انفجار مخازن با ریسک خطای انسانی دو موضوع متفاوت هستند و می‌توانند منجر به خسارت‌های کاملاً متفاوتی نیز شوند. برای مدل‌بندی کل زیان پالایشگاه با فرض اینکه K گروه ریسک متفاوت در پالایشگاه وجود دارد، می‌توانیم از ترکیب دو مدل ریسک فردی و جمعی میزان کل زیان را به صورت زیر بنویسیم:

$$S(t) = S_1(t) + \dots + S_K(t)$$

که در آن منظور از $S_j(t)$ برای $j = 1, \dots, K$ همان مجموع زیان‌های مربوط به گروه ریسک j ام است که از مدل ریسک جمعی حاصل می‌شود؛ یعنی اینکه:

$$S_j(t) = Y_{1,j} + \dots + Y_{N_j(t),j} = \sum_{i=1}^{N_j(t)} Y_{i,j} ; \quad j = 1, \dots, K$$

که در این برابری $Y_{i,j}$ برای $i = 1, \dots, N_j(t)$ نشان‌دهنده‌ی مقدار خسارت ناشی از رخداد i امین حادثه در گروه ریسک j ام است و $N_j(t)$ تعداد حوادثی است که در گروه ریسک j ام رخ می‌دهد.

در ادبیات نظریه قابلیت اعتماد از رویکرد متفاوتی برای معرفی مدل‌های ریسک فردی استفاده می‌کنند که در ادامه به آن اشاره شده است. با نمادگذاری بالا و تعریف نماد $p_j = P(N_j(t) > 0)$ برای $j = 1, \dots, K$ داریم:

$$S(t) = \sum_{j=1}^K I_{p_j} S_j(t) \quad (1)$$

که در این رابطه متغیر نشانگر I_{p_j} زمانی برابر ۱ است که $N_j(t) > 0$.

برای یافتن توزیع $S(t)$ در صورتی که فرض استقلال برقرار باشد می‌توان از روش پیش‌فرض استفاده کرد. اما فرض استقلال بین انواع زیان‌های یک پرتفوی در نمونه‌های بزرگ بیمه‌ای، فرض واقعی‌ای نیست. بنابراین در صورت امکان باید همبستگی بین ریسک‌ها را در نظر بگیریم. یک راه کار برای این منظور استفاده از مفصل است که در پژوهش‌های مختلفی در نظر گرفته شده است (به عنوان مثال اسماعیلی و کلاپلبرگ، ۲۰۱۰ را ببینید). با تعیین نوع مدل ریسک، تعریف نمادها و یافتن توزیع زیان در پرتفوی می‌توانیم برای مقاصد بیمه‌ای مانند تعیین حق بیمه و بررسی توانگری مالی شرکت‌ها اقدام کنیم. در عمل صنایع بزرگی مانند پالایشگاه می‌توانند با توجه به این رویکرد ریسک‌های خود را ارزیابی کرده و پس از تحلیل آنها در مورد امکان واگذاری آنها به بیمه‌گران اقدام کنند. هر چند به طور طبیعی بیمه‌گران ممکن است مجدداً ارزیابی ریسک را انجام داده و بر این اساس نسبت به اعلام نرخ به پالایشگاه اعلام نظر و مذاکره کنند. به واقع بحث بالا نشان می‌دهد که هر دو موضوع بیمه‌پذیری ریسک‌ها و محاسبه حق بیمه قراردادهای بیمه‌ای مبتنی بر توزیع زیان پرتفوی خواهد بود.

۳ اولویت‌بندی ریسک‌ها

بر پایه استانداردهای مدیریت ریسک، ارزشیابی ریسک آخرین مرحله در فرآیند ارزیابی ریسک است. منظور از ارزشیابی ریسک به زبان ساده، اولویت‌بندی و تعیین ریسک‌های مهم است. دقت داریم که به عنوان یک سازمان بزرگ (مثلاً همان پالایشگاه را در نظر بگیرید) یا به عنوان یک فرد پس از تحلیل ریسک‌هایی که با آنها روبرو هستیم می‌توانیم تصمیم بگیریم که برخی ریسک‌ها را نزد خود نگهداریم و برخی را به بیمه‌گر منتقل کنیم. به طور شهودی ریسکی را به بیمه‌گر منتقل می‌کنیم که از توان ما خارج باشد. علاوه بر این ریسک‌هایی که نزد بیمه‌گر هستند با توجه به عوامل تشکیل دهنده ریسک ممکن است تفاوت‌های بسیار زیادی با یکدیگر داشته باشند. به عنوان مثال در بیمه بدنه اتومبیل، نوع و قیمت اتومبیل تأثیر زیادی در خسارت‌های احتمالی دارد. اما ملاک اولویت‌بندی ریسک‌ها با توجه به ماهیت تصادفی آنها چگونه است؟ این یکی از سوال‌هایی است که اکچوئری‌ها در ارزیابی

ریسک باید برای آن پاسخ داشته باشند. برای این منظور ابتدا باید برای هر ریسک، نمایه‌های ریسک را مشخص کرد. مثلاً در رشته‌های بیمه غیر زندگی (مانند اتومبیل، باربری، مسئولیت، آتش‌سوزی و ...) نمایه‌های ریسک رایج شامل تعداد و مبلغ خسارت‌های ناشی از رخداد حوادث مرتبط با آنهاست. پس از تعیین نمایه ریسک می‌توانیم از رویکردهای آماری برای مقایسه ریسک‌ها استفاده کنیم. اطلاعات جدول ۱ بر اساس ۷۱۷۶۰ بیمه‌نامه بدنه اتومبیل مربوط به یکی از شرکت‌های بیمه ایرانی در سال ۱۳۹۰ به دست آمده است. در فایل اطلاعاتی مربوط به این پرتفوی برای هر بیمه‌گذار تعداد خسارت‌ها و مجموع کل مبلغ خسارت پرداختی مشخص شده است. یعنی به تک تک خسارت‌های پرداخت شده دسترسی نداریم. بنابراین مقادیر ارائه شده در سطر سوم این جدول، بر اساس مجموع خسارت‌های مثبت بر تعداد خسارت‌ها محاسبه شده است. با توجه به این اطلاعات آیا می‌توانیم ریسک خسارت‌های بیمه بدنه اتومبیل را با توجه به نوع سواری با یکدیگر مقایسه کنیم؟ در پاسخ به این سوال به نظر می‌رسد که همین شاخص‌های آماری ساده که در جدول ۱ ارائه شده است راهگشا باشند. بر این اساس ملاحظه می‌کنیم که سواری‌های گران‌قیمت‌تر به طور متوسط خسارت بیشتری نسبت به سواری‌های با قیمت پایین‌تر دریافت کرده‌اند. هر چند متوسط تعداد خسارت‌های پرداختی عکس این نتیجه را نشان می‌دهد.

جدول ۱۸: توصیف میزان خسارت پرداختی بیمه بدنه اتومبیل در یکی از شرکت‌های بیمه ایرانی

نام شاخص ریسک	کمتر از ۴ سیلندر	۴ سیلندر معمولی	سایر سواری‌های با قیمت بالاتر
متوسط تعداد خسارت‌ها	۰,۱۰	۰,۰۹	۰,۰۶
انحراف معیار تعداد خسارت‌ها	۰,۳۰	۰,۳۱	۰,۲۶
متوسط مبلغ خسارت (تومان)	۱۸۳۸۷۵	۴۲۶۱۲۶	۸۱۶۰۷۵

اخیراً راهکار دیگری برای مقایسه ریسک‌ها مبتنی بر ترتیب‌های تصادفی ارائه شده است. در چنین رویکردی به جای مقایسه دو یا چند شاخص مانند مثال بالا، مقایسه کلی‌تری صورت می‌گیرد. در بخش بعدی این موضوع را مورد بررسی قرار می‌دهیم.

۴ کاربرد ترتیب‌های تصادفی در بیمه

مروری کلی بر مطالعه‌های صورت گرفته در خصوص کاربرد ترتیب‌های تصادفی در بیمه نشان می‌دهد که میزان و نوع ناهمگنی پرتفوی تحت بررسی یکی از پارامترهای اصلی در تولید نتایج نظری است. برای مدل ارائه شده در

رابطه 1 همگنی پرتفوی را می‌توان به دو طریق در نظر گرفت:

• اگر $S_j(t)$ ها هم‌توزیع باشند شرط همگنی برابری p_j هاست.

• اگر p_j ها برابر باشند شرط همگنی هم‌توزیعی $S_j(t)$ هاست.

بنابراین تغییر p_j ها در شرایطی که پذیره هم‌توزیعی $S_j(t)$ ها برقرار باشد میزان ناهمگنی را نشان می‌دهد. همچنین تفاوت بین توزیع $S_j(t)$ ها در شرایطی که p_j ها برابر باشند میزان ناهمگنی را نشان می‌دهد. ترکیب این دو حالت نیز به شرط آنکه اثر یکدیگر را خنثی نکنند، میزان ناهمگنی پرتفوی را تغییر می‌دهد. این موضوع از زوایای مختلف نظری در دهه اخیر مورد مطالعه قرار گرفته است. تقریباً در اکثر این پژوهش‌ها سه گام زیر طی شده است:

الف- زیر فضایی از فضای پارامتر ساخته می‌شود.

ب- یک ترتیب ریاضی (مانند بیشاندن) روی اعضای این فضا در نظر گرفته می‌شود

ج- دو عضو دلخواه از زیرفضا که در ترتیب قسمت ب صدق کنند را در نظر گرفته و مبنای اثبات یک ترتیب تصادفی روی کمیت $S(t)$ قرار می‌دهند.

به عنوان نمونه‌هایی از چنین مطالعه‌هایی دنوئیت و فراستینگ (۲۰۰۶)، برمالزن و همکاران (۲۰۱۵) و بالاکریشن و همکاران (۲۰۱۷) را ببینید. به منظور بررسی قابلیت‌های کاربردی نتایج این مطالعه‌ها در حوزه بیمه باید به چند سوال اساسی پاسخ داد. نخستین سوال مربوط به انگیزه ایجاد زیرفضاهای تعریف شده است. به عنوان مثال در مطالعه برمالزن و همکاران (۲۰۱۵) و بالاکریشن و همکاران (۲۰۱۷) از مجموعه زیر استفاده شده است:

$$\mathcal{U}_n = \left\{ (x, y) = \begin{pmatrix} x_1, \dots, x_n \\ y_1, \dots, y_n \end{pmatrix} ; x_i, y_j > 0, (x_i - x_j)(y_i - y_j) > 0, i, j = 1, \dots, n \right\}$$

آنها اشاره می‌کنند که این مجموعه نسبت به جایگشت ناورد است و لذا قابلیت‌های نظری بهتری فراهم می‌کند. با این وجود هیچ گونه توجیهی در مورد نیاز به چنین ویژگی در پرتفوی بیمه‌ای ارائه نمی‌شود. سوال دوم مربوط به نحوه تعیین (برآورد) پارامترهای مدل است. در واقع چارچوب نظری مطرح شده در این مطالعه‌ها موضوع اصلی را که قابلیت پیاده‌سازی نتایج بر اساس یک مجموعه داده است به فراموشی سپرده است و مدام در تلاش است با تغییر مولفه‌های نظری موضوع، نتایج نظری بیشتری بدست آورد. در اکثر مقاله‌های چاپ شده در این زمینه

مثال‌های ساختگی بسیار ساده‌ای مطرح شده است که کمک زیادی به علاقمندان حوزه بیمه برای استفاده از این روش‌ها نمی‌کند. یکی دیگر از سوال‌های اساسی که در این زمینه مطرح است در خصوص ابهام در عمل مقایسه دو پرتفوی است. به نظر می‌رسد هدف از به کارگیری ترتیب‌های تصادفی در این زمینه مقایسه دو پرتفوی مجزا نیست؛ بلکه هدف یافتن شرایط بهینه‌ای برای یک پرتفوی با در نظر گرفتن یک فضای پارامتر و مجموعه‌ای از فرض‌های توزیعی است. با وجود اینکه این موضوع در مطالعه دنوئیت و فراستینگ (۲۰۰۶) تا حدودی در نظر گرفته شده است در کارهای بعدی قدری کم‌رنگ‌تر شده است. در این خصوص کران‌های بالایی و پایینی که بر مبنای حالت همگن برای پرتفوی ناهمگن ساخته شده است، بسیار مفید است. در نهایت سوال آخر مربوط به نوع مدل ریسک در نظر گرفته شده در مطالعه‌های مربوطه است. در مثال عددی بالا ۷۱۷۶۰ بیمه‌گذار وجود دارد. آیا می‌بایست برای بررسی ترتیب تصادفی بردار p_j ها یک بردار با ۷۱۷۶۰ مقدار باشد؟ یا اینکه می‌توان سه مجموع مرکب $S_1(t)$ ، $S_2(t)$ و $S_3(t)$ را به ترتیب برای سه نوع سواری در نظر گرفت و سپس از مدل ۱ استفاده کرد. این موضوع در مقاله‌های مرتبط با این موضوع به درستی مطرح نشده است و حتی در برمالزن و همکاران (۲۰۱۵) صحبت از تصادفی بودن تعداد ادعاهای خسارت و در نظر گرفتن متغیر تصادفی K در مدل ۱ به میان آمده است که به نظر می‌رسد نیاز به بازنگری دارد.

۵ نتیجه‌گیری و مطالعات آتی

مطالعه حاضر مربوط به مسائلی است که محققان حوزه نظریه قابلیت اعتماد به هنگام کاربرد نتایج خود در حوزه بیمه می‌بایست مورد توجه قرار دهند. با توجه به مرور اهداف حوزه کارهای محاسباتی بیمه، ملاحظه کردیم که نظریه قابلیت اعتماد جایگاه بسیار خوبی در کنار اکچوئری‌ها دارد. لیکن نمونه‌هایی از پژوهش‌های انجام گرفته در سال‌های اخیر، نشان می‌دهد که توجه به جنبه بیمه‌ای آنها نیازمند بازنگری دقیق‌تری می‌باشد.

مراجع

- [1] Balakrishnan, N., Zhang, Y., and Zhao, P. (2017). *Ordering the largest claim amounts and ranges from two sets of heterogeneous portfolios*, Scandinavian Actuarial Journal, <http://dx.doi.org/10.1080/03461238.2017.1278717>.

- [2] Barmalzan, G., Najafabadi, A. T. P. and Balakrishnan, N. (2015). *Stochastic comparison of aggregate claim amounts between two heterogeneous portfolios and its applications*. Insurance: Mathematics and Economics 61, 235–241.
- [3] Denuit, M., Frostig, E. (2006). *Heterogeneity and the need for capital in the individual model*. Scandinavian Actuarial Journal, 1, 42-66.
- [4] Esmaceli, H., Klüppelberg, C., (2010). *Parameter estimation of a bivariate compound Poisson process*. Insurance Mathematics and Economics, 47, 224–233.
- [5] Ghasemi, S., Mahmoudvand, R., (2013). *A New Insight into the Failure Mode and Effect Analysis and its Application in Insurance*, International Journal of Energy and Statistics, 1(4), 352–362.
- [6] Ushakov, I., (2012). *Reliability Theory: history & current state in bibliographies*, Reliability: Theory & Applications, 7(1), 1–24.

بررسی تأثیر ساختار وابستگی و توزیع‌های حاشیه‌ای در محاسبه قابلیت اعتماد مدل‌های تنش-مقاومت

محمودی‌فرد، ف^۱ دست‌برآورده، ع^۲ میرحسینی، س. م^۳

گروه آمار، دانشکده علوم ریاضی، دانشگاه یزد

چکیده

در این مقاله تأثیر ساختار وابستگی و توزیع‌های حاشیه‌ای را در محاسبه قابلیت اعتماد مدل‌های تنش-مقاومت، $R = P(X > Y)$ ، مورد مطالعه قرار می‌دهیم. به‌طور خاص شرایطی را که منجر می‌شود R ، بزرگتر مساوی یا کوچکتر از $\frac{1}{2}$ باشد بررسی می‌نماییم. برای این منظور از رویکرد مفصل مبنا بهره می‌بریم.

کلمات کلیدی: تابع مفصل، قابلیت اعتماد، مدل تنش-مقاومت، وابستگی.

^۱ mahmodif94@gmail.com

^۲ dastbaravarde@yazd.ac.ir

^۳ mmirhoseini@yazd.ac.ir

۱ پیش‌گفتار

مسئله قابلیت اعتماد مدل تنش-مقاومت یکی از مسائل مورد علاقه پژوهشگران است. مسئله قابلیت اعتماد در سیستم‌هایی که مقاومت X در معرض تنش Y قرار دارد مطرح می‌شود. در حالتی که $X \leq Y$ ، سیستم دچار شکست شده یا عملکرد سیستم قابل قبول نیست. اگر X و Y متغیر تصادفی باشند، قابلیت اعتماد سیستم را می‌توان به صورت $R = P(X > Y)$ بیان کرد که P نماد احتمال است. [2]

کاربردهای اولیه مدل تنش-مقاومت در علوم مهندسی است. اما کاربرد آن محدود به این علوم نیست. در حقیقت در زمینه‌های مختلفی از جمله کنترل کیفیت، ژنتیک، روانشناسی، تعلیم و تربیت و بهداشت و درمان کاربرد دارد.

برخی از پژوهش‌های انجام شده در این زمینه، R را وقتی متغیرهای X و Y مستقل با توزیع‌های حاشیه‌ای مشترک از جمله نمایی تعمیم‌یافته [11]، گاما [6]، وایبل [7]، پاراتو تعمیم‌یافته [2] هستند محاسبه و برآورد کرده‌اند. با وجود اینکه در برخی مسائل کاربردی ممکن است تنش و مقاومت مستقل باشند، اما در بسیاری از مسائل کاربردی تنش و مقاومت وابسته‌اند و برای محاسبه R نیازمند توزیع توأم هستیم. برخی از پژوهش‌ها برای تعدادی توزیع‌های توأم خاص مانند نرمال دو متغیره [3]، پاراتو دو متغیره [5]، بتا دو متغیره [8]، گاما دو متغیره [9]، لگ نرمال دو متغیره [4]، R را محاسبه و برآورد کرده‌اند؛ این پژوهش‌ها متغیرهای تنش و مقاومت را تنها در فرمت خاصی از وابستگی در نظر گرفته و در آنها هر دو توزیع حاشیه‌ای از یک خانواده هستند.

با رویکرد مفصل مبنا می‌توان این محدودیت‌ها را برطرف کرد. از یک سو با استفاده از تابع مفصل می‌توان توابع توزیع حاشیه‌ای مختلف را در نظر گرفت و لازم نیست که این توابع از یک خانواده باشند. از سوی دیگر می‌توان ساختارهای وابستگی مختلف را مورد مطالعه قرار داد. به‌طور کلی، با این رویکرد می‌توان تأثیر ساختارهای وابستگی و توزیع‌های حاشیه‌ای مختلف را به تفکیک مورد مطالعه قرار داد.

اسکلار در سال ۱۹۵۹ نشان داد که رابطه‌ای خاص بین تابع توزیع توأم و توابع توزیع حاشیه‌ای متغیرهای تصادفی وجود دارد. فرض کنید X و Y دو متغیر تصادفی با تابع توزیع توأم H و توابع توزیع حاشیه‌ای به‌ترتیب F و G باشند آنگاه، تابع حقیقی مقدار $[0, 1] \times [0, 1] \rightarrow [0, 1]$ وجود دارد به‌قسمی که برای هر دو مقدار حقیقی x و y

$$H(x, y) = C(F(x), G(y)).$$

تابع C را مفصل می‌نامند که دارای ویژگی‌های زیر است:

۱. به ازای کلیه u و v های متعلق به I ,

$$C(u, \circ) = C(\circ, v) = \circ,$$

$$C(u, \mathbf{1}) = u, \quad C(\mathbf{1}, v) = v,$$

۲. اگر $u_1 < u_2$ و $v_1 < v_2$ آنگاه

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

در صورتی که X و Y مطلقاً پیوسته بوده و C دارای مشتقات جزئی باشد، تابع چگالی توأم X و Y ، h ، را می توان به صورت زیر نوشت:

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

که در آن f و g به ترتیب تابع چگالی حاشیه ای X و Y ، و $c(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}$ هستند. فرض کنید C یک تابع مفصل پیوسته مطلق باشد. در این صورت

$$\begin{aligned} \frac{\partial C(u, v)}{\partial u} &= P(V \leq v \mid U = u), \\ \frac{\partial C(u, v)}{\partial v} &= P(U \leq u \mid V = v). \end{aligned}$$

فرض کنید C یک تابع مفصل دوبعدی باشد. مفصل C را متقارن گوئیم، هرگاه به ازای کلیه u و v های متعلق به I ، $C(u, v) = C(v, u)$.

فرض کنید X و Y متغیرهای تصادفی پیوسته با تابع مفصل C باشند. زوج (X, Y) هم توزیع با (Y, X) است اگر و تنها اگر X و Y هم توزیع و مفصل C متقارن باشد. به عبارت دیگر،

$$(X, Y) \stackrel{d}{=} (Y, X) \iff X \stackrel{d}{=} Y, \quad C(u, v) = C(v, u) \forall u, v \in I. \quad (1)$$

در حالتی که $(X, Y) \stackrel{d}{=} (Y, X)$ ، X و Y را تبادله پذیر گوئیم.

(مفصل بقا) برای هر مفصل C ، تابع $\hat{C}(u, v) = u + v - \mathbf{1} + C(\mathbf{1} - u, \mathbf{1} - v)$ نیز یک مفصل است که آن را مفصل بقای C می نامند. فرض کنید C یک مفصل باشد به طوری که $C = \hat{C}$ ، در این حالت مفصل C را متقارن

شعاعی گوئیم. (تقارن شعاعی) زوج (X, Y) در نقطه (a, b) دارای تقارن شعاعی است اگر زوج $(X - a, Y - b)$ با $(a - X, b - Y)$ هم توزیع باشد. زوج (X, Y) با ساختار وابستگی C در نقطه (a, b) دارای تقارن شعاعی است اگر و تنها اگر $X - a \stackrel{d}{=} a - X$ ، $Y - b \stackrel{d}{=} b - Y$ و $C = \hat{C}$. (وابستگی رگرسیونی منفی) زوج متغیر تصادفی (X, Y) را وابسته‌ی رگرسیونی منفی گوئند هرگاه $P(Y \leq y | X = x)$ به ازای تمام y ها بر حسب x صعودی باشد. زوج متغیر تصادفی (X, Y) با ساختار وابستگی C وابسته‌ی رگرسیونی منفی است اگر و تنها اگر $\frac{\partial C(u, v)}{\partial u}$ به ازای تمام v ها بر حسب u صعودی باشد.

در ادامه مقاله، در بخش 2 محاسبه پارامتر مدل تنش-مقاومت با رویکرد تابع مفصل ارائه می‌شود. در بخش 3 قضایایی در مورد نقش توزیع‌های حاشیه‌ای و وابستگی بر مقدار R بیان می‌شود. در نهایت، جمع‌بندی نتایج در بخش 4 ارائه می‌شود.

۲ محاسبه R تحت ساختار وابستگی C

فرض کنید X و Y متغیرهای تصادفی پیوسته با ساختار وابستگی C و توابع توزیع حاشیه‌ای به ترتیب F ، G و توابع چگالی حاشیه‌ای به ترتیب f و g باشند، آنگاه R به صورت زیر محاسبه می‌شود:

$$\begin{aligned} R = P(X > Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^x h(x, y) dy dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^x c(F(x), G(y)) g(y) f(x) dy dx, \end{aligned}$$

یا به طور معادل،

$$R = P(X > Y) = \int_0^1 \int_0^{G(F^{-1}(u))} c(u, v) dv du, \quad (2)$$

از طرف دیگر، با توجه به

$$G_{Y|X=x}(y) = \frac{\partial C(u, v)}{\partial u} \Big|_{u=F(x), v=G(y)}, \quad R = \int_{\mathbb{R}} G_{Y|X=x}(x) f(x) dx,$$

داریم:

$$R = E \left(\frac{\partial C(u, v)}{\partial u} \Big|_{u=F(X), v=G(X)} \right). \quad (3)$$

اگر در مدل تنش- مقاومت متغیرهای تنش و مقاومت هم توزیع باشند، R با استفاده از رابطه (2) به صورت زیر محاسبه می شود.

$$R = P(X > Y) = \int_0^1 \int_0^u c(u, v) dv du = E \left(\frac{\partial C(u, v)}{\partial u} \Big|_{u=v=W} \right),$$

که در آن W متغیر تصادفی با توزیع یکنواخت روی بازه $[0, 1]$ است.

۳ برخی نتایج در مورد ساختار وابستگی و نقش توزیع های حاشیه ای بر

مقدار R

از فواید رویکرد مفصل مبنا در محاسبه R این است که می توان نقش و میزان تأثیر توزیع های حاشیه ای و ساختار وابستگی را به تفکیک مورد مطالعه قرار داد. در این بخش برخی نتایج در مورد نقش و میزان تأثیر ساختار وابستگی و توزیع های حاشیه ای در مقدار R ارائه شده است. فرض کنید X و Y متغیرهای تصادفی پیوسته و هم توزیع با تابع چگالی مفصل c باشند. برای هر $(u, v) \in I^2$ که $u > v$

۱. اگر $c(u, v) = c(v, u)$ آنگاه، $R = \frac{1}{2}$ ؛

۲. اگر $c(u, v) > c(v, u)$ آنگاه، $R > \frac{1}{2}$ ؛

۳. اگر $c(u, v) < c(v, u)$ آنگاه، $R < \frac{1}{2}$.

برهان. ۱. با توجه به نکته 1 ،

$$\begin{aligned} P(X > Y) &= \int_0^1 \int_0^u c(u, v) dv du \\ &= \int_0^1 \int_0^u c(v, u) dv du \\ &= \int_0^1 \int_v^1 c(v, u) du dv \\ &= P(X < Y). \end{aligned} \quad (4)$$

تساوی (۴) بنابر فرض قضیه برقرار است.

۲. اثبات موارد (2) و (3) به صورت مشابه است. تنها کافی در رابطه (۴) تساوی با نامساوی های متناظر جایگزین شود.

□

در این قضیه با توجه به هم توزیع بودن حاشیه ها ساختار وابستگی نقش تعیین کننده ای در مقدار R دارد؛ به طوریکه، ساختار وابستگی که احتمال بیشتری به ناحیه $u > v$ اختصاص دهد مقدار بیشتر R را موجب می شود. فرض کنید X و Y متغیرهای تصادفی پیوسته با توزیع های حاشیه ای به ترتیب F و G ، و مفصل متقارن C باشند.

۱. اگر $F = G$ آنگاه، $R = \frac{1}{4}$ ؛

۲. اگر $F > G$ آنگاه، $R \leq \frac{1}{4}$ ؛

۳. اگر $F < G$ آنگاه، $R \geq \frac{1}{4}$.

برهان. ۱. از اینکه C مفصلی متقارن است و $F = G$ نتیجه می شود که بردارهای تصادفی (X, Y) و (Y, X) هم توزیع هستند (مراجعه شود به قضیه ۲,۷,۴ مرجع ([2])). بنابراین $P(X > Y) = P(X < Y) = \frac{1}{4}$.

۲. از اینکه $\frac{\partial C(u, v)}{\partial u} = P(V \leq v | U = u)$ و $F > G$ ، برای هر مقدار x ، داریم:

$$\frac{\partial C(u, v)}{\partial u} \Big|_{u=F(x), v=G(x)} \leq \frac{\partial C(u, v)}{\partial u} \Big|_{u=F(x), v=F(x)}.$$

بنابراین

$$\begin{aligned} R &= \int_{-\infty}^{\infty} \frac{\partial C(u, v)}{\partial u} \Big|_{u=F(x), v=G(x)} f(x) dx \\ &\leq \int_{-\infty}^{\infty} \frac{\partial C(u, v)}{\partial u} \Big|_{u=F(x), v=F(x)} f(x) dx \\ &= \frac{1}{4}; \end{aligned}$$

تساوی آخر بنا به قسمت (1) برقرار است.

۳. به صورت مشابه قابل اثبات است.

□

مفصل های ارشمیدسی در قضیه 2، صدق می کنند. فرض کنید X و Y متغیرهای تصادفی پیوسته با توابع توزیع حاشیه ای به ترتیب F و G ، و دارای مفصل متقارن شعاعی C باشند.

۱. اگر $F = G$ آنگاه، $R = \frac{1}{4}$ ؛

۲. اگر $F < G$ آنگاه، $R \geq \frac{1}{4}$ ؛

۳. اگر $F > G$ آنگاه، $R \leq \frac{1}{4}$.

برهان. با توجه به اینکه $C = \hat{C}$ ، داریم:

$$c(u, v) = c(1-u, 1-v). \quad (5)$$

۱. قابلیت اعتماد با استفاده از رابطه (1)، به صورت زیر محاسبه می شود.

$$R = P(X > Y) = \int_0^1 \int_0^u c(u, v) dv du \quad (6)$$

$$= \int_0^1 \int_0^u c(1-u, 1-v) dv du \quad (7)$$

$$= \int_0^1 \int_s^1 c(s, t) dt ds \quad (8)$$

$$= P(Y > X); \quad (9)$$

بنابراین، $R = \frac{1}{4}$. تساوی (6)، بنا به $F = G$ ، تساوی (7)، با توجه به رابطه‌ی (5)، و تساوی (8)، با تغییر متغیرهای $1-u = s$ و $1-v = t$ برقرار است.

۲.

$$R = P(X > Y) = \int_0^1 \int_0^{G(F^{-1}(u))} c(u, v) dv du$$

$$\geq \int_0^1 \int_0^u c(u, v) dv du \quad (10)$$

$$= \int_0^1 \int_u^1 c(u, v) dv du \quad (11)$$

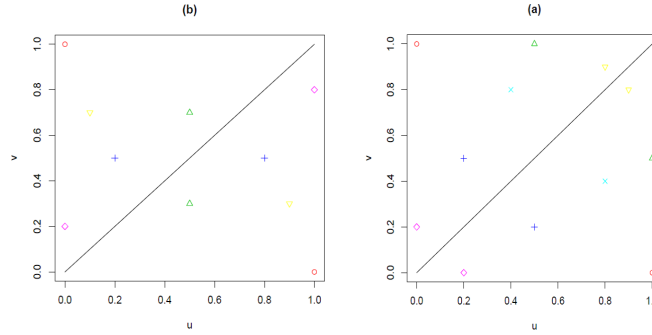
$$= \frac{1}{4};$$

نامساوی (10)، با قرار دادن $x = F^{-1}(u)$ در رابطه $F(x) \leq G(x)$ به دست می آید و تساوی (11) بنا به رابطه (5) برقرار است.

۳. به صورت مشابه قابل اثبات است.

□

قضیه‌های 2 و 4 بیان می کنند اگر مفصل دارای نوعی تقارن باشد آنگاه، توزیع‌های حاشیه‌ای نقش تعیین کننده‌ای در مقدار R دارند. در مفصل متقارن چگالی مفصل حول صفحه $u = v$ متقارن است و در مفصل متقارن شعاعی چگالی مفصل حول خط $u = \frac{1}{4}$ و $v = \frac{1}{4}$ متقارن است. نمودارهای (1) تعبیر هندسی از تقارن و تقارن شعاعی



شکل ۱: نمایش نقاطی که دارای چگالی مفصل یکسان در یک مفصل (a) متقارن (b) متقارن شعاعی دارند.

مفصل را نشان می‌دهد. فرض کنید X و Y دارای توزیع نمایی با میانگین λ_1 و λ_2 و همچنین مفصل FGM با پارامتر وابستگی θ باشند، قابلیت اعتماد آن به صورت زیر محاسبه می‌شود.

بنا به رابطه (2)،

$$\begin{aligned} R &= E \left(\frac{\partial C(u, v)}{\partial u} \middle| u = F(X), v = G(X) \right) \\ &= E(G(X)) + \theta E[G(X) \bar{G}(X) (1 - 2F(X))] \\ &= \frac{\lambda_2}{\lambda_1 + \lambda_2} + \frac{\theta \lambda_1 \lambda_2 (\lambda_1 - \lambda_2)}{(\lambda_1 + \lambda_2)(\lambda_1 + 2\lambda_2)(2\lambda_1 + \lambda_2)}, \end{aligned}$$

مفصل FGM دارای ساختار وابستگی تقارن شعاعی است، بنابراین قضیه 2 بر آن صادق است. در نمودار (2) قابلیت اعتماد نسبت به $\frac{\lambda_1}{\lambda_2}$ را برای θ های ۱، ۰، -۱ رسم شده است. همانطور که از نمودار (2) مشاهده می‌شود برای $\lambda_1 < \lambda_2$ ($X >_{st} Y$) در نتیجه بنا به قضیه $\frac{1}{2} < R$ ، برای $\lambda_1 = \lambda_2$ ($X \stackrel{d}{=} Y$)، $R = \frac{1}{2}$ و برای $\lambda_1 > \lambda_2$ ($X <_{st} Y$)، $R < \frac{1}{2}$ است و همچنین، با افزایش θ ، R روند صعودی دارد و به سمت یک میل می‌کند.

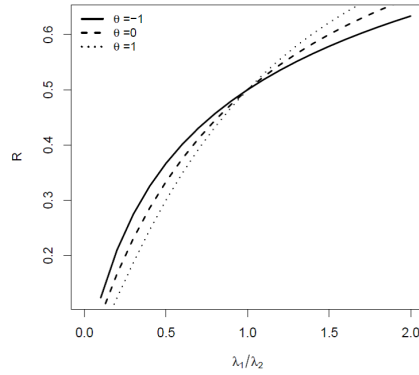
قضیه

فرض کنید X و Y متغیرهای تصادفی پیوسته و دارای مفصل C باشند. اگر زوج (X, Y) حول (a, b) متقارن شعاعی باشند؛ در این صورت،

۱. اگر $a = b$ آنگاه، $R = \frac{1}{2}$ ؛

۲. اگر $a > b$ آنگاه، $R \geq \frac{1}{2}$ ؛

۳. اگر $a < b$ آنگاه، $R \leq \frac{1}{2}$.



شکل ۲: نمودار R برای مفصل FGM با حاشیه‌های نمایی با میانگین‌های λ_1 و λ_2

برهان. با توجه به اینکه زوج (X, Y) متقارن شعاعی حول (a, b) هستند داریم:

۱. با توجه به گزاره، $(X - a, Y - a) \stackrel{d}{=} (a - X, a - Y)$. بنابراین،

$$P(X > Y) = P(X - a > Y - a) = P(a - X > a - Y) = P(X < Y).$$

۲. با توجه به گزاره، دارای تقارن شعاعی حول (a, b) است. بنابراین،

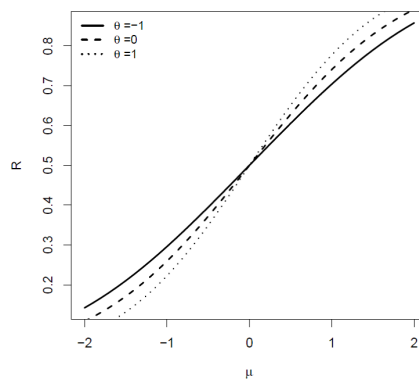
$$\begin{aligned} P(X > Y) &= P(X - a > Y - a) \\ &\geq P(X - a > Y - b) \\ &= P(a - X > Y - b) \\ &\geq P(a - X > a - Y) \\ &= P(X < Y), \end{aligned}$$

بنابراین، $P(X > Y) \geq \frac{1}{3}$.

۳. به صورت مشابه قابل اثبات است.

□

فرض کنید X و Y دارای توابع توزیع حاشیه‌ای به ترتیب $N(\mu, 1)$ و $Lap(0, 1)$ و مفصل FGM باشند. توزیع‌های X و Y حول صفر متقارن هستند بنابراین، بنا به قضیه ۴ مورد (3)، $R = \frac{1}{3}$ ، که درستی آن را می‌توان در نمودار (3)، مشاهده کرد.



شکل ۳: نمودار قابلیت اعتماد برای مفصل FGM

فرض کنید X و Y متغیر تصادفی پیوسته با ساختار وابستگی C باشند.

۱. اگر $F < G$ آنگاه،

$$R > E \left(\frac{\partial C(u, v)}{\partial u} \Big|_{u=v=W} \right); \quad (12)$$

۲. اگر $F > G$ آنگاه،

$$R < E \left(\frac{\partial C(u, v)}{\partial u} \Big|_{u=v=W} \right). \quad (13)$$

که در آن امید ریاضی نسبت به متغیر تصادفی W با توزیع یکنواخت روی بازه $(0, 1)$ است.

برهان. ۱. طبق رابطه (2)، داریم:

$$\begin{aligned} R &= \int_{-\infty}^{\infty} \frac{\partial C(u, v)}{\partial u} \Big|_{u=F(x), v=G(x)} f(x) dx \\ &> \int_{-\infty}^{\infty} \frac{\partial C(u, v)}{\partial u} \Big|_{u=v=F(x)} f(x) dx \\ &= E \left(\frac{\partial C(u, v)}{\partial u} \Big|_{u=v=F(X)} \right), \end{aligned} \quad (14)$$

نامساوی (14)، بنابر گزاره 1، غیر نزولی بودن تابع توزیع و همچنین بنابر فرض اینکه $F < G$ برقرار است.

۲. به صورت مشابه قابل اثبات است.

□

قضیه 8 ، کرانی برای R با توجه به رابطهای توزیعهای حاشیه‌ای ارائه داده است. برای مفصل‌های مختلف از جمله مفصل ارشمیدسی کران $\frac{1}{2}$ و برای مفصل‌های غیر ارشمیدسی $E\left(\frac{\partial C(u,v)}{\partial u}\bigg|_{u=v=F(X)}\right)$ محاسبه می‌شود. فرض کنید (X, Y) دارای رگرسیون منفی باشد.

۱. اگر $F < G$ آنگاه،

$$E\left(\frac{\partial C(u,v)}{\partial u}\bigg|_{u=v=F(X)}\right) < R < E\left(\frac{\partial C(u,v)}{\partial u}\bigg|_{u=v=G(X)}\right); \quad (15)$$

۲. اگر $F > G$ آنگاه،

$$E\left(\frac{\partial C(u,v)}{\partial u}\bigg|_{u=v=G(X)}\right) < R < E\left(\frac{\partial C(u,v)}{\partial u}\bigg|_{u=v=F(X)}\right). \quad (16)$$

برهان. ۱. نامساوی $R > E\left(\frac{\partial C(u,v)}{\partial u}\bigg|_{u=v=F(X)}\right)$ بنابر قضیه 8 ، برقرار است؛ از سوی دیگر، بنا به 10 و اینکه $F < G$ داریم:

$$\frac{\partial C(u,v)}{\partial u}\bigg|_{u=F(x), v=G(X)} > \frac{\partial C(u,v)}{\partial u}\bigg|_{u=v=G(X)}. \quad (17)$$

بنابراین،

$$R > E\left(\frac{\partial C(u,v)}{\partial u}\bigg|_{u=v=G(X)}\right). \quad (18)$$

۲. به صورت مشابه قابل اثبات است.

□

با توجه به قضیه 8 ، می‌توان نقش توزیعهای حاشیه‌ای را در مقدار R ، زمانی که X و Y دارای وابستگی منفی باشند، مشاهده کرد؛ به‌طوری‌که توزیعهای حاشیه‌ای محدوده تغییر R را تعیین می‌کنند.

۴ دست‌آوردهای پژوهش

در این مقاله پارامتر مدل تنش-مقاومت را با رویکرد تابع مفصل مورد بررسی قرار دادیم. پارامتر مدل تنش-مقاومت تحت تاثیر دو عامل ساختار وابستگی تنش-مقاومت و توزیعهای حاشیه‌ای آن دو است. نتایجی در رابطه با تاثیر این دو عامل بر مقدار پارامتر مدل تنش-مقاومت ارائه شد.

- [1] Domma, F. and Giordano, S. (2013), *A copula-based approach to account for dependence in stress-strength models*, Statistical Papers, 54(3), 807–826.
- [2] Guo, H. and Krishnamoorthy, K. (2004). *New approximate inferential methods for the reliability parameter in a stress-strength model: The normal case*. Communications in Statistics-Theory and Methods, 33(7), 1715-1731.
- [3] Gupta, R.C. and Subramanian, S. (1998). *Estimation of reliability in a bivariate normal distribution with equal coefficients of variation*, Communications in Statistics - Simulation and Computation, 27(3), 675–698.
- [4] Gupta, R.C., Ghitany, M.E and Al-Mutairi, D.K. (2013). *Estimation of reliability from a bivariate log-normal data*, Journal of Statistical Computation and Simulation, 83(6), 1068–1081.
- [5] Hanagal, D.D. (1997). *Note on estimation of reliability under bivariate pareto stress-strength model*, Statistical Papers, 38, 453–459.
- [6] Huang, K., Mi, J. and Wang, Z. (2012). *Inference about reliability parameter with gamma strength and stress*, Journal of tatistical Planning and Inference, 142(4), 848–854.
- [7] Krishnamoorthy, K., and Lin, Y. (2010). *Confidence limits for stress-strength reliability involving Weibull models*, Journal of Statistical Planning and Inference, 140, 1754–1764.
- [8] Nadarajah, S. (2005). *Reliability for some bivariate gamma distributions*, Mathematical Problems in Engineering, 2005(1), 151–163.
- [9] Nadarajah, S. (2005). *Reliability for some bivariate beta distributions*, Mathematical Problems in Engineering, 2005(2), 101–11
- [10] Nelsen, R. B. (2007), *An introduction to copulas*, Springer Science and Business Media.
- [11] Raqab, M. Z., Madi, M. T. and Kundu, D. (2008). *Estimation of $P(Y<X)$ from the three parameter generalized exponential distribution*, Communications in Statistics Theory and Methoods, 37, 2854–2864.

- [12] Rezaei, S., Tahmasbi, R. and Mahmoodi, M. (2010). *Estimation of $P(Y < X)$ for generalized pareto distribution*, Journal of Statistical Planning and Inference, 140, 480–494.



مقایسه دو برآوردگر ناپارامتری قابلیت اعتماد

مهرآور، ف^۱ مهدی زاده، م^۲

گروه آمار، دانشگاه حکیم سبزواری

چکیده

در این مقاله، برآورد ناپارامتری قابلیت اعتماد بر اساس تابع توزیع تجربی و برآورد هسته‌ای تابع چگالی مطالعه می‌شود. به کمک خواص U -آماره‌ها، واریانس و توزیع حدی دو برآوردگر را به دست می‌آوریم. سپس با استفاده از شبیه‌سازی، رفتار این برآوردگرها برای حجم نمونه متناهی مقایسه می‌شود.

کلمات کلیدی: احتمال تنش-مقاومت؛ برآورد هسته‌ای تابع چگالی؛ تابع توزیع تجربی.

^۱ farzanehmehravar@yahoo.com

^۲ mahdizadeh.m@live.com

۱ مقدمه

استنباط آماری در مورد پارامتر $\theta = P(X > Y)$ وقتی X و Y دو متغیر تصادفی مستقل هستند، موضوعی مهم در قابلیت اعتماد^۳ بوده که مورد توجه محققان زیادی قرار گرفته است. اگر X میزان مقاومت یک سیستم، و Y میزان تنش وارد شده به آن باشد، سیستم در صورتی به کار خود ادامه می‌دهد که مقاومت آن از تنش بیشتر باشد. بنابراین، پارامتر θ احتمال کار کردن سیستم را نشان می‌دهد. علاوه بر مبحث قابلیت اعتماد، برآورد این پارامتر دارای کاربردهای دیگری در پزشکی، روانشناسی، علوم تربیتی و ... می‌باشد. به عنوان مثال، اگر X و Y به ترتیب طول عمر بیماران باشد که از داروی جدید و دارونما^۴ استفاده می‌کنند، آنگاه پارامتر θ را می‌توان به عنوان معیار اثر بخشی داروی جدید تفسیر کرد. برای مرور کاملی از مطالعات انجام شده در این زمینه وقتی X و Y عضو یک خانواده از توزیع‌ها هستند، به کوتز و همکاران [3] مراجعه کنید.

در بخش ۲، دو برآوردگر ناپارامتری برای θ معرفی می‌شود. واریانس و توزیع مجانبی این برآوردگرها در بخش ۳ مطالعه می‌شود. در بخش ۴، رفتار این دو برآوردگر را وقتی متغیرها توزیع نرمال یا نمایی دارند، با استفاده از شبیه‌سازی مقایسه می‌کنیم.

۲ برآوردگرها

فرض کنید $X_1, \dots, X_{n_1}$ و $Y_1, \dots, Y_{n_2}$ نمونه تصادفی مستقل از دو جامعه به ترتیب با تابع چگالی احتمال پیوسته f و g باشند. توابع توزیع تجمعی متناظر را با F و G نشان می‌دهیم. برآوردگر ناپارامتری θ بر اساس تابع توزیع تجربی عبارتست از

$$\tilde{\theta} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} I(X_i > Y_j). \quad (1)$$

تحت فرض استقلال می‌توان نوشت

$$\theta = P(X > Y) = E(I(X > Y)) = \int \int I(u > v) f(u) g(v) du dv, \quad (2)$$

^۳Reliability

^۴Placebo

که در آن $I(\cdot)$ تابع نشانگر است. یک برآوردگر دیگر برای θ از جایگزینی f و g در رابطه (2) با برآوردگری ناپارامتری به دست می‌آید. به این منظور می‌توان از توابع چگالی هسته‌ای

$$\hat{f}(u) = \frac{1}{n_1 h_1} \sum_{i=1}^m k\left(\frac{u - X_i}{h_1}\right),$$

و

$$\hat{g}(v) = \frac{1}{n_2 h_2} \sum_{j=1}^n k\left(\frac{v - Y_j}{h_2}\right),$$

استفاده کرد که در آن هسته k ، یک تابع چگالی متقارن و پارامترهای هموارسازی h_1 و h_2 عرض باند^۵ نامیده می‌شوند. با به کارگیری $\hat{f}$ و $\hat{g}$ در رابطه (2) داریم

$$\begin{aligned} \hat{\theta} &= \int \int I(u > v) \hat{f}(u) \hat{g}(v) du dv \\ &= \int \int I(u > v) \left[\frac{1}{n_1 h_1} \sum_{i=1}^{n_1} k\left(\frac{u - X_i}{h_1}\right) \right] \left[\frac{1}{n_2 h_2} \sum_{j=1}^{n_2} k\left(\frac{v - Y_j}{h_2}\right) \right] du dv \\ &= \int \int \int \int I(u > v) \left[\frac{1}{h_1} k\left(\frac{u - x}{h_1}\right) \right] \left[\frac{1}{h_2} k\left(\frac{v - y}{h_2}\right) \right] du dv d\hat{F}(x) d\hat{G}(y), \end{aligned}$$

که در آن $\hat{F}$ و $\hat{G}$ توابع توزیع تجربی هستند. با تعویض متغیرهای $r = (u - x)/h_1$ و $s = (v - y)/h_2$ در رابطه اخیر نتیجه می‌شود

$$\begin{aligned} \hat{\theta} &= \int \int \left[\int \int I(h_1 r + x > h_2 s + y) k(r) k(s) dr ds \right] d\hat{F}(x) d\hat{G}(y) \\ &= \int \int H(x - y) d\hat{F}(x) d\hat{G}(y), \end{aligned} \quad (3)$$

که در آن H تابع توزیع $h_2 S - h_1 R$ است و R و S متغیرهای تصادفی مستقل با چگالی مشترک k هستند. اگر k تابع چگالی احتمال متغیر نرمال استاندارد باشد، آنگاه H تابع توزیع یک متغیر تصادفی با میانگین صفر و انحراف استاندارد $t = \sqrt{h_1^2 + h_2^2}$ می‌باشد. در این حالت، از رابطه (3) داریم

$$\hat{\theta} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \Phi\left(\frac{X_i - Y_j}{t}\right), \quad (4)$$

که در آن $\Phi(\cdot)$ تابع توزیع تجمعی متغیر نرمال استاندارد می‌باشد. باکلیزی و ایدوس [1] برآوردگر فوق را در ساختن فاصله اطمینان برای θ به کار بردند اما واریانس و توزیع مجانبی این برآوردگر را مطالعه نکردند.

^۵Bandwidth

۳ ویژگی‌های نظری

در این بخش، ابتدا برخی مفاهیم اساسی در مورد U -آماره‌های ϕ -دونمونه‌ای بیان می‌شود. سپس، دو قضیه در مورد واریانس و توزیع مجانبی این گونه آماره‌ها مطرح کرده و آنها در مورد دو برآوردگر $\hat{\theta}$ و $\tilde{\theta}$ به کار می‌بریم. فرض کنید $\mathcal{P}$ کلاس تمام توزیع‌های ممکن برای X و Y باشد. پارامتر $\vartheta(F, G)$ ، برآوردپذیر از درجه (p, q) است اگر p و q کوچکترین حجم نمونه‌هایی باشند که به ازای آنها، یک آماره $T(X_1, \dots, X_p; Y_1, \dots, Y_q)$ وجود دارد چنانکه

$$E_{F,G}(T(X_1, \dots, X_p; Y_1, \dots, Y_q)) = \vartheta(F, G), \quad \forall F, G \in \mathcal{P}.$$

آماره T در تعریف فوق، هسته برای ϑ نامیده می‌شود و نسخه متقارن شده آن، T_s ، هسته متقارن برای ϑ می‌باشد. بدون از دست رفتن کلیت مسئله، در تعریف بعدی فرض می‌کنیم که T یک هسته متقارن است. فرض کنید $F, G \in \mathcal{P}$ و $\vartheta(F, G)$ یک پارامتر برآوردپذیر از درجه (p, q) باشد. آنگاه U -آماره دونمونه‌ای برای برآورد $\vartheta(F, G)$ به صورت

$$U(\mathbf{X}, \mathbf{Y}) = \frac{1}{\binom{n_1}{p} \binom{n_2}{q}} \sum_{i \in A} \sum_{j \in B} T(X_{i_1}, \dots, X_{i_p}; Y_{j_1}, \dots, Y_{j_q}) \quad (5)$$

تعریف می‌شود که در آن A و B به ترتیب مجموعه تمام زیرمجموعه‌های p و q عضوی از $\{1, \dots, n_1\}$ و $\{1, \dots, n_2\}$ هستند.

در بخش ۲، $X_1, \dots, X_{n_1}$ و $Y_1, \dots, Y_{n_2}$ دو نمونه تصادفی مستقل به ترتیب از F و G هستند. با فرض $\vartheta(F, G) = P(X > Y)$ ، واضح است که $T(X_i; Y_j) = I(X_i > Y_j)$ یک برآوردگر نااریب برای ϑ است. درجه ϑ برابر (۱ و ۱) بوده و U -آماره دونمونه‌ای برای برآورد آن عبارتست از

$$\tilde{\theta} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} T(X_i; Y_j).$$

همچنین، $\hat{\theta}$ یک U -آماره دونمونه‌ای به صورت

$$\hat{\theta} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} T(X_i; Y_j)$$

می‌باشد که در آن $T(X_i; Y_j) = \Phi\left(\frac{X_i - Y_j}{t}\right)$

^۶ U-statistics

واریانس U - آماره دونمونه‌ای تعریف شده در (5) برابر است با

$$Var(U(\mathbf{X}, \mathbf{Y})) = \frac{1}{\binom{n_1}{p}\binom{n_2}{q}} \sum_{c=0}^p \sum_{d=0}^q \binom{p}{c} \binom{n_1-p}{p-c} \binom{q}{d} \binom{n_2-q}{q-d} \zeta_{c,d}$$

که در آن

$$\zeta_{c,d} = Cov \left(T(X_{i_1}, \dots, X_{i_p}; Y_{j_1}, \dots, Y_{j_q}), (T(X_{k_1}, \dots, X_{k_p}; Y_{\ell_1}, \dots, Y_{\ell_q})) \right),$$

و c و d به ترتیب تعداد X ها و Y های مشترک در عبارت فوق است.

با استفاده از قضیه فوق می‌توان واریانس $\tilde{\theta}$ را محاسبه کرد. در این حالت، $p = q = 1$ و داریم

$$Var(\tilde{\theta}) = \frac{1}{n_1 n_2} [(n_1 - 1)\zeta_{s,1} + (n_2 - 1)\zeta_{1,s} + \zeta_{1,1}]$$

که در آن

$$\begin{aligned} \zeta_{s,1} &= Cov(I(X_i > Y_j), I(X_{i'} > Y_j)) \\ &= EE\{I(X_i > Y_j), I(X_{i'} > Y_j) | Y_j\} - \theta^2 \\ &= \int_{-\infty}^{\infty} [1 - F(y)]^2 g(y) dy - \theta^2, \end{aligned} \quad (6)$$

$$\begin{aligned} \zeta_{1,s} &= Cov(I(X_i > Y_j), I(X_i > Y_{j'})) \\ &= EE\{I(X_i > Y_j), I(X_i > Y_{j'}) | X_i\} - \theta^2 \\ &= \int_{-\infty}^{\infty} G^2(x) f(x) dx - \theta^2, \end{aligned} \quad (7)$$

و

$$\begin{aligned} \zeta_{1,1} &= Var(I(X_i > Y_j)) \\ &= \theta - \theta^2. \end{aligned}$$

به طور مشابه، واریانس $\hat{\theta}$ را نیز می‌توان به دست آورد. در اینجا با فرض $\eta = E(\Phi(\frac{X-Y}{t}))$ داریم

$$Var(\hat{\theta}) = \frac{1}{n_1 n_2} [(n_1 - 1)\zeta_{s,1} + (n_2 - 1)\zeta_{1,s} + \zeta_{1,1}]$$

که در آن

$$\begin{aligned}
 \zeta_{\circ, \circ} &= Cov \left(\Phi \left(\frac{X_i - Y_j}{t} \right), \Phi \left(\frac{X_{i'} - Y_j}{t} \right) \right) \\
 &= EE \left\{ \Phi \left(\frac{X_i - Y_j}{t} \right) \Phi \left(\frac{X_{i'} - Y_j}{t} \right) \middle| Y_j \right\} - \eta^\circ \\
 &= \int_{-\infty}^{\infty} E^\circ \Phi \left(\frac{x - y}{t} \right) g(y) dy - \eta^\circ \\
 &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \Phi \left(\frac{x - y}{t} \right) f(x) dx \right]^\circ g(y) dy - \eta^\circ, \tag{۸}
 \end{aligned}$$

$$\begin{aligned}
 \zeta_{\circ, \circ} &= Cov \left(\Phi \left(\frac{X_i - Y_j}{t} \right), \Phi \left(\frac{X_i - Y_{j'}}{t} \right) \right) \\
 &= EE \left\{ \Phi \left(\frac{X_i - Y_j}{t} \right) \Phi \left(\frac{X_i - Y_{j'}}{t} \right) \middle| X_i \right\} - \eta^\circ \\
 &= \int_{-\infty}^{\infty} E^\circ \Phi \left(\frac{x - y}{t} \right) f(x) dx - \eta^\circ \\
 &= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \Phi \left(\frac{x - y}{t} \right) g(y) dy \right]^\circ f(x) dx - \eta^\circ, \tag{۹}
 \end{aligned}$$

و

$$\begin{aligned}
 \zeta_{\circ, \circ} &= Var \left(\Phi \left(\frac{X_i - Y_j}{t} \right) \right) \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Phi^2 \left(\frac{x - y}{t} \right) f(x) g(y) dx dy - \eta^\circ.
 \end{aligned}$$

قضیه بعدی، توزیع مجانبی U-آماره دونمونه‌ای تعریف شده در (5) را نشان می‌دهد. اثبات این قضیه در لهما [۴] آمده است. فرض کنید $X_1, \dots, X_{n_1}$ و $Y_1, \dots, Y_{n_2}$ دو نمونه تصادفی مستقل به ترتیب از F و G باشند و $\vartheta(F, G)$ یک پارامتر برآوردپذیر از درجه (p, q) با هسته متقارن $T_s(X_1, \dots, X_{n_1}; Y_1, \dots, Y_{n_2})$ باشد. اگر $E_{F, G}(T^\circ) < \infty$ و U-آماره دونمونه‌ای برای برآورد ϑ به صورت (5) تعریف شود، آنگاه وقتی

$$n_1, n_2 \rightarrow \infty,$$

$$\sqrt{n_1 + n_2} (U(\mathbf{X}; \mathbf{Y}) - \vartheta(F, G)) \xrightarrow{d} N(\circ, \sigma^\circ)$$

به شرطی که $\lambda \in (\circ, 1)$ و $\frac{n_1}{n_1 + n_2} \rightarrow \lambda$ و $\sigma^\circ = \frac{p^\circ \zeta_{\circ, \circ}}{\lambda} + \frac{q^\circ \zeta_{\circ, 1}}{1 - \lambda} > \circ$. اگر λ همانند قضیه فوق تعریف شده باشد، آنگاه نتیجه می‌شود:

$$\sqrt{n_1 + n_2} (\tilde{\theta} - \theta) \xrightarrow{d} N(\circ, \frac{\zeta_{\circ, \circ}}{\lambda} + \frac{\zeta_{\circ, 1}}{1 - \lambda}),$$

که در آن $\zeta_{1,0}$ و $\zeta_{1,1}$ در (6) و (7) آمده اند. به طور مشابه، می توان دید

$$\sqrt{n_1 + n_2}(\hat{\theta} - \theta) \xrightarrow{d} N(0, \frac{\zeta_{1,0}}{\lambda} + \frac{\zeta_{1,1}}{1-\lambda}),$$

که در آن $\zeta_{1,0}$ و $\zeta_{1,1}$ در (8) و (9) آمده اند.

۴ نتایج عددی

برای مقایسه رفتار $\hat{\theta}$ و $\tilde{\theta}$ ، یک مطالعه شبیه سازی انجام شد که نتایج آن در این بخش آمده است. به این منظور، حالاتی را در نظر گرفتیم که توزیع هر دو جامعه، نرمال یا نمایی باشد. فرض کنید X و $Y - \mu$ متغیرهای تصادفی نرمال استاندارد باشند. آنگاه می توان دید

$$\theta = \Phi\left(\frac{-\mu}{\sqrt{2}}\right).$$

که در آن $\Phi(\cdot)$ تابع توزیع X است. به طور مشابه، برای متغیرهای نمایی استاندارد X و Y/α ($\alpha > 0$) داریم

$$\theta = \frac{1}{1+\alpha}.$$

تحت هر توزیع، ۵ مقدار برای پارامتر متناظر طوری انتخاب شد که به ازای آنها

$$\theta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$$

حاصل شود. مقادیر مناسب پارامتر در جدول 1 آمده است. همچنین، حجم نمونه های انتخاب شده عبارتند از

$$(n_1, n_2) \in \{(3, 3), (3, 7), (5, 5), (10, 10)\}.$$

برای هر ترکیب θ و حجم نمونه ها، ۱۰۰۰۰ جفت نمونه تصادفی از هر توزیع تولید شد. دو برآوردگر از هر جفت نمونه محاسبه شده و بر اساس مقادیر حاصل، میانگین مربع خطای $^{\vee}$ (MSE) برآوردگرها تعیین شد. کارایی نسبی

$^{\wedge}$ (RE) به صورت

$$RE = \frac{\widehat{MSE}(\tilde{\theta})}{\widehat{MSE}(\hat{\theta})}$$

$^{\vee}$ Mean squared error

$^{\wedge}$ Relative efficiency

جدول ۱۹: مقادیر پارامتر متناظر با پارامترهای قابلیت اطمینان متفاوت

θ					
پارامتر	۰,۱	۰,۳	۰,۵	۰,۷	۰,۹
μ	۱,۸۱۲۳۸۸	۰,۷۴۱۶۱۴۳	۰	-۰,۷۴۱۶۱۴۳	-۱,۸۱۲۳۸۸
α	۹	۷/۳	۱	۳/۷	۱/۹

تعریف می‌شود. مقادیر RE بزرگتر از یک نشان می‌دهد که $\hat{\theta}$ از $\tilde{\theta}$ کاراتر است. جدول‌های ۲ و ۳ نتایج را نشان می‌دهند که در آن RE^۱ - RE^۴ به ترتیب بر اساس چهار روش تعیین عرض باند که در ادامه معرفی می‌شوند، محاسبه شده‌اند. روش اول، مینیم کردن میانگین مربع خطای انتگرال‌گیری شده مجانبی^۹ (AMISE) برای برآوردگر هسته‌ای تابع چگالی است. روش دوم، اعتباریابی نااریب^{۱۰} (UCV) است که توسط رادمو [۵] و بومن [۲] پیشنهاد شد. روش سوم، اعتباریابی اریب^{۱۱} (BCV) است که توسط اسکات و جورج [۶] مطالعه شد. روش آخر، روش جایگذاری^{۱۲} (PI) است که توسط شیر و جونز [۸] معرفی شد. جزئیات بیشتر درباره این روش‌ها در شیر [۷] آمده است.

بر اساس جدول‌های ۲ و ۳، نتیجه‌گیری‌های زیر را در مورد مقایسه برآوردگر هسته‌ای قابلیت اعتماد و برآوردگر مشابه بر اساس تابع توزیع تجربی می‌توان انجام داد:

- در همه حالات، دست کم بر اساس یکی از چهار روش تعیین عرض باند، $\hat{\theta}$ از $\tilde{\theta}$ بهتر عمل می‌کند.
- برای هر جفت نمونه مفروض، مقادیر کارایی برای $\theta = ۰.۵$ بزرگتر هستند.
- در برخی موارد، $\hat{\theta}$ از $\tilde{\theta}$ دقت کمتری دارد. برای مثال، نتایج مربوط به $(n_1, n_2) = (۵, ۵), (۱۰, ۱۰)$ و $\theta = ۰.۸, ۰.۹$ را ملاحظه کنید.
- گاهی کارایی حاصل از روش‌های AMISE، UCV و BCV خیلی کمتر از یک است. در این شرایط، روش PI عملکرد بهتری دارد.

^۹Asymptotic mean integrated squared error

^{۱۰}Unbiased cross-validation

^{۱۱}Biased cross-validation

^{۱۲}Plug-in

جدول ۲۰: مقادیر برآورد RE تحت توزیع نرمال

RE۴	RE۳	RE۲	RE۱	θ	(m, n)
۱,۱۳	۰,۵۷	۰,۶۰	۰,۵۷	۰,۱	(۳,۳)
۱,۲۱	۲,۰۷	۲,۰۳	۲,۰۷	۰,۳	
۱,۲۳	۳,۴۲	۳,۱۷	۳,۴۲	۰,۵	
۱,۲۱	۲,۰۷	۲,۰۳	۲,۰۷	۰,۷	
۱,۱۴	۰,۵۸	۰,۶۲	۰,۵۸	۰,۹	
۱,۰۴	۰,۴۳	۰,۵۲	۰,۴۳	۰,۱	(۳,۷)
۱,۲۱	۱,۷۸	۱,۶۶	۱,۷۸	۰,۳	
۱,۲۶	۳,۳۱	۲,۵۲	۳,۳۱	۰,۵	
۱,۲۳	۱,۸۰	۱,۶۹	۱,۸۰	۰,۷	
۱,۰۳	۰,۴۴	۰,۵۳	۰,۴۴	۰,۹	
۱,۰۱	۰,۳۹	۰,۵۴	۰,۳۹	۰,۱	(۵,۵)
۱,۲۳	۱,۶۸	۱,۵۹	۱,۶۸	۰,۳	
۱,۲۸	۳,۴۳	۲,۳۶	۳,۴۳	۰,۵	
۱,۲۴	۱,۷۱	۱,۶۱	۱,۷۱	۰,۷	
۰,۹۹	۰,۳۹	۰,۵۳	۰,۳۹	۰,۹	
۰,۸۵	۰,۲۵	۰,۴۸	۰,۲۵	۰,۱	(۱۰,۱۰)
۱,۲	۱,۲۲	۱,۳۰	۱,۲۲	۰,۳	
۱,۳۰	۳,۰۵	۱,۸۳	۳,۰۵	۰,۵	
۱,۱۹	۱,۱۹	۱,۲۸	۱,۱۹	۰,۷	
۰,۸۴	۰,۲۴	۰,۴۷	۰,۲۴	۰,۹	

جدول ۲۱: مقادیر برآورد RE تحت توزیع نمایی

RE۴	RE۳	RE۲	RE۱	θ	(m, n)
۱,۲۶	۰,۵	۰,۵۳	۰,۵	۰,۱	(۳,۳)
۱,۳۴	۲,۲۶	۲,۲۲	۲,۲۵	۰,۳	
۱,۲۹	۳,۵۲	۳,۲۷	۳,۵۲	۰,۵	
۱,۳۶	۲,۳	۲,۲۷	۲,۳	۰,۷	
۱,۲۷	۰,۴۹	۰,۵۲	۰,۴۹	۰,۹	
۱,۰۲	۰,۲۶	۰,۵۹	۰,۲۶	۰,۱	(۳,۷)
۱,۴۲	۲,۱۴	۱,۹	۲,۱۴	۰,۳	
۱,۳	۳,۵۳	۲,۴	۳,۵۳	۰,۵	
۱,۳۳	۱,۹۱	۱,۸۴	۱,۹۱	۰,۷	
۱,۱۱	۰,۴	۰,۴۴	۰,۴	۰,۹	
۱,۰۲	۰,۲۹	۰,۵۱	۰,۲۶	۰,۱	(۵,۵)
۱,۳۷	۱,۹۵	۱,۸۱	۱,۹۵	۰,۳	
۱,۳۱	۳,۶۲	۲,۲	۳,۶۲	۰,۵	
۱,۳۸	۱,۹۶	۱,۸۳	۱,۹۶	۰,۷	
۱,۰۳	۰,۲۹	۰,۵۱	۰,۲۹	۰,۹	
۰,۷۱	۰,۱۴	۰,۵	۰,۱۴	۰,۱	(۱۰,۱۰)
۱,۳۸	۱,۳۴	۱,۵۲	۱,۳۴	۰,۳	
۱,۳۲	۳,۳۹	۱,۶۱	۳,۳۹	۰,۵	
۱,۳۸	۱,۳۴	۱,۵۲	۱,۳۴	۰,۷	
۰,۷۲	۰,۱۴	۰,۵۱	۰,۱۴	۰,۹	

با توجه به عملکرد بهتر برآوردگرهای هسته‌ای، به عنوان آینده تحقیق قصد داریم از این برآوردگرها در ساختن فاصله اطمینان برای θ استفاده کنیم.

مراجع

- [1] Baklizi, A., and Eidous, O. (2006). *Nonparametric estimation of $P(X < Y)$ using kernel methods*. Metron, LXIV, 47–60.
- [2] Bowman, A.W. (1984). *An alternative method of cross-validation for the smoothing of kernel density estimates*. Biometrika 71, 353–360.
- [3] Kotz, S., Lumelskii, Y., and Pensky, M. (2003). *The stress-strength model and its generalizations. Theory and Applications*. World Scientific, Singapore.
- [4] Lehmann, E.L. (1975). *Nonparametrics: statistical methods based on ranks*. San Francisco: Holden-Day.
- [5] Rudemo, M. (1982). *Empirical choice of histograms and kernel density estimators*. Scandinavian Journal of Statistics 9, 65–78.
- [6] Scott, D.W., and George, R.T. (1987). *Biased and unbiased cross-validation in density estimation*. Journal of the American Statistical Association 82, 1131–1146.
- [7] Sheather, S.J. (2004). *Density estimation*. Statistical Science, 19, 588–597.
- [8] Sheather, S.J., and Jones, M.C. (1991). *A reliable data-based bandwidth selection method for kernel density estimation*. Journal of the Royal Statistical Society, Series B 53, 683–690.

مدل رگرسیون نمایی قطعه ای در تحلیل داده های بقاء بیماران مبتلا به انفارکتوس حاد و مقایسه این مدل با مدل رگرسیون کاکس

هادی پور، م^۱ منوچهری، ط^۲ بانسی، م ر^۳

^۱ مرکز تحقیقات سیاستگذاری سلامت، دانشگاه علوم پزشکی شیراز

^۲ بخش آمار، دانشکده علوم پایه، دانشگاه شیراز

^۳ دانشگاه علوم پزشکی کرمان

چکیده

انفارکتوس حاد میوکارد (Acute Myocardial Infarction) AMI یا سکته قلبی حاد یکی از مهم ترین عامل های مرگ و میر در اغلب کشورهای جهان از جمله ایران می باشد. با توجه به نقش متغیرهای مختلف در روند بیماری، در این تحقیق تلاش شده زمان بقاء بیماران براساس متغیرهای پیشگو مدل بندی شود و همچنین میزان بقاء بیماران مبتلا به سکته قلبی در فواصل مختلف از زمان بقاء بیماران مورد مطالعه قرار گیرد. در این مطالعه همگروهی آینده نگر، ۶۰۷ بیمار با سن بالاتر از ۲۵ سال پذیرش شده در بخش بیمارستان امام رضا (ع) که تشخیص انفارکتوس حاد میوکارد با معیارهای کلینیکی، ECG و آنزیم قلبی در آنها محرز بوده، از سال ۱۳۹۱-۱۳۶۸ مورد پیگیری و بررسی قرار گرفته اند. در این مطالعه ضمن بررسی وضعیت بیماری پس از سکته قلبی، اطلاعاتی در مورد سن، جنس، مصرف سیگار، سابقه دیابت، سابقه پرفشاری خون، سابقه هایپرلیپیدی،

دریافت و عدم دریافت استرپتوکیناز و ... نیز استخراج گردید و اثر آن بر سرانجام بیماری در فاصله های مختلف از زمان بقای بیماران با استفاده از مدل رگرسیون تکه تکه (Piecewise Regression) بررسی و همچنین با رگرسیون کاکس مقایسه شد. داده ها با استفاده از نرم افزار STATA آنالیز گردید.

کلمات کلیدی: سکته قلبی حاد، تحلیل بقا، رگرسیون کاکس، رگرسیون تکه ای، مطالعه کوهورت.

۱ مقدمه

بیماری های قلبی- عروقی در حال حاضر یکی از شایع ترین علت مرگ و میر در اکثر کشورهای جهان است. بیش از ۳۰ درصد مرگ و میر در کشورهای صنعتی بر اثر بیماری های قلبی - عروقی به وقوع می پیوندد. [۱] در حدود ۸۲ درصد از مرگ و میرهای ناشی از بیماری های قلبی مربوط به کشورهای در حال توسعه از جمله ایران، می باشد. در کشورهای در حال توسعه و به خصوص در کشور ما به دلیل پایین بودن فرهنگ عمومی در زمینه آشنایی با عوامل خطرزای بیماری های قلبی - عروقی و عدم رعایت اصول پیش گیری از این بیماری، مرگ و میر آن در حال افزایش است. همچنین با توجه به کاهش سن بروز در افراد و کشندگی بیش تر این بیماری در سنین پایین، هزینه بیش تری بر جامعه تحمیل می گردد.

با توجه به متفاوت بودن امکانات بهداشتی - درمانی و تفاوت های فرهنگی اقتصادی میزان مرگ ناشی از AMI در جوامع مختلف، با هم متفاوت هستند. میزان مرگ و میر انفارکتوس قلبی حدود ۲۵ تا ۳۰ درصد است و بیش از نیمی از این مرگ و میرها مربوط به زمانی هست که هنوز به بیمارستان نرسیده اند. میزان مرگ داخل بیمارستان در اثر AMI، از مقدار ۲,۱۹ درصد تا ۷,۷ درصد متفاوت بوده است. [۲، ۳]

تحلیل بقا در بیماران انفارکتوس قلبی حاد از اهمیت بالایی برخوردار است. مطالعات اندکی وجود دارد که دوره طولانی مدت ۵ ساله را مورد بررسی قرار داده اند. از آنجا که عوامل موثر در بقا بیماران در دوره های بلند مدت متفاوت تر از دوره های کوتاه مدت است، مطالعه دوره طولانی مدت، بسیار اهمیت دارد. در مطالعات انجام شده عواملی چون سن، جنسیت، پرفشاری خون، مصرف سیگار، مصرف استرپتوکیناز، دیابت و هایپرلیپیدمی که به عنوان متغیرهای موثر بر بقاء بیماران سکته قلبی حاد بیان می شود. در این مطالعه علاوه بر متغیرهای ذکر شده، عامل های دیگری مانند روش تشخیص بیماری، سابقه ابتلا در اعضاء درجه اول، سابقه اعتیاد به مواد مخدر، نوع مداخله بالینی، وجود موج Q و سابقه ایسمی به منظور بررسی و پیشگویی وضعیت بقاء پنج ساله بیماران مبتلا به سکته قلبی حاد در بیمارستان امام رضا علیه السلام بین سالهای ۱۳۸۶ تا ۱۳۹۱ پرداخته شد.

۲ پیش زمینه

مطالعه حاضر یک مطالعه از نوع همگروهی (کوهورت) آینده نگر می باشد. تمامی بیماران ۲۵ سال به بالا که بین سالهای ۱۳۸۶ تا ۱۳۹۱، با تشخیص انفارکتوس حاد میوکارد (AMI) در بیمارستان امام رضا (ع) بستری شدند مورد بررسی قرار گرفتند و سپس به مدت پنج سال دنبال شدند. شناسایی سکته قلبی براساس آزمایشات آنزیمی قلب و علائم ECG و در بعضی از موارد براساس تشخیص متخصصان گزارش شد.

برای دستیابی و استفاده از پرونده پزشکی بیماران، از مسئولان مربوطه بیمارستان اجازه نامه کتبی گرفته شد. بمنظور جلوگیری از افزایش داده های گمشده، در پرونده هایی که اطلاعات بیمار ناقص بودند و یا وضعیت بقای وی ثبت نشده بود، با خود فرد بیمار یا خانواده وی تماس گرفته شد و بمنظور عدم ایجاد احساس ناراحتی و آسیب به فرد بیمار یا خانواده وی، نهایت تلاش در برقراری تماس مناسب و اطلاع گرفتن از وضعیت بیمار به صورت غیر مستقیم انجام شد. بیماران بالای ۱۰۵ سال، بیماران بستری شده از سایر شهر و استانهای کشور، بیماران با سابقه قلبی سکته قلبی و بیمارانی که در بخش CCU بستری نشده بودند از مطالعه خارج شدند.

به منظور مدلبندی زمان بقاء پنج ساله بیماران حاضر در مطالعه براساس متغیرهای پیشگو همچون سن، جنسیت، سابقه پرفشاری خون، سابقه دیابت و مصرف سیگار و... از دو مدل رگرسیون کاکس و مدل رگرسیون Piecewise استفاده گردید و سپس این دو مدل از نظر اندازه، جهت و معنی داری ضرایب برآورده شده متغیرهای پیشگو مقایسه شدند. مقدار معنی داری در مطالعه حاضر پنج صدم در نظر گرفته شد. کلیه تجزیه و تحلیل های آماری از توصیف تا تحلیل با استفاده از نرم افزار STATA-SE11 انجام شد.

۳ یافته ها

در مطالعه حاضر تعداد ۶۰۷ بیمار بررسی شدند که از این تعداد ۴۲۳ نفر (۶۹,۷ درصد) مرد بودند. میانگین و انحراف معیار سن بیماران در زمان سکته قلبی، $۱۳,۴۶ \pm ۶۱,۳۴$ سال با کمترین و بیشترین مقدار به ترتیب ۲۶ و ۹۶ سال بود.

میانگین و انحراف معیار مدت زمان بقاء بیماران در مطالعه حاضر $۸۴۰,۸۳ \pm ۱۴۳۶,۸۵$ روز بود. همچنین میانگین و انحراف معیار طول مدت بستری بیماران $۴,۷۳ \pm ۷,۳۹$ با کمترین و بیشترین این مقدار به ترتیب ۱ و ۳۷ روز بود. جدول یک، خصوصیات دموگرافیک بیماران را به تفکیک جنس نشان میدهد.

جهت بررسی عوامل موثر بر رخداد مرگ پس از سکته قلبی حاد از مدل رگرسیون چند متغیره کاکس استفاده شد. ابتدا توسط آزمون کاپلان-مایر متغیرهای موثر بر رخداد مرگ مشخص می شود. در این حالت متغیرهای را که در

جدول ۲۲: خصوصیات دموگرافیک بیماران

جنسیت	شاخص	سن	مدت زمان بستری در بیمارستان	زمان بقا
مرد	تعداد	۴۲۳	۴۲۳	۴۲۳
	میانگین	۵۹,۶۲	۷,۱۳	۱۵۱۸,۸۶
	انحراف استاندارد	۱۳,۴۱۰	۴,۱۲۱	۸۰۸,۱۶۲
	مینیم	۲۶	۱	۰
	ماکزیمم	۸۹	۳۱	۲۲۶۳
زن	تعداد	۱۴۸	۱۴۸	۱۴۸
	میانگین	۶۵,۳۰	۷,۹۷	۱۲۴۸,۳۳
	انحراف استاندارد	۱۲,۷۴۷	۵,۸۸۲	۸۸۵,۳۰۶
	مینیم	۲۶	۱	۰
	ماکزیمم	۹۶	۳۷	۲۲۲۶

آنالیز تک متغیره داراي P-value کمتر از ۰,۲۵ بودند وارد آنالیز چندمتغیره شدند. (جدول شماره دو)

در مرحله بعد، با وارد کردن متغیرهای معنی دار به مدل رگرسیونی piecewise ضرایب رگرسیون و مقدار معنی داری حاصل میشود. تقسیم بندی زمان بقاء به چهار دسته صفر تا ۵۶۸ روز، ۵۶۸ روز تا ۱۱۳۵ روز، ۱۱۳۵ روز تا ۱۶۹۸ روز و ۱۶۹۸ روز تا ۲۲۶۶ روز انجام شد. همانطور که در جدول سه مشاهده میشود برای هریک از بازه های زمانی یک ضریب رگرسیونی برآورد شده است. ریسک فاکتورهایی همچون مداخله انجام شده (بای پس و پیس میکر) و مصرف استرپتوکیناز در مدل رگرسیونی piecewise معنی دار نشد. (جدول شماره سه)

۴ بحث

براساس یافته های تحلیل رگرسیون کاکس بعد از تعدیل اثر سایر متغیرهای مورد بررسی (با حذف اثر متغیرهای بی تاثیر) متغیرهای سن، دیابت، سابقه ایسکمی، نوع مداخله انجام گرفته و روش تشخیصی، مصرف استرپتوکیناز

جدول ۲۳: جدول مدل بندی رگرسیون کاکس

فاکتورها	ضرایب رگرسیون	خطای استاندارد	P-value
سن	۰,۰۴۱	۰,۰۰۶	۰,۰۰۱>
دیابت	۰,۶۶۵	۰,۱۴۷	۰,۰۰۱>
سابقه ایسکمی	۰,۳۵۸	۰,۱۵۲	۰,۰۱۸
انفارکتوس حادترانس مورال با محل غیر مشخص	۱,۲۳۸	۰,۴۳۸	۰,۰۰۵
انفارکتوس حاد غیر مشخص	۰,۴۰۶	۰,۲۱۸	۰,۰۵۰
مصرف استرپتوکیناز	-۰,۳۳۹	۰,۱۱۶	۰,۰۴
مداخله انجام شده- بای پس	-۱,۱۷	۰,۴۵۹	۰,۰۱۷
مداخله انجام شده- پیس میکر	۱,۰۹۱	۰,۲۲۶	۰,۰۰۱

جدول ۲۴: جدول مدل بندی رگرسیون Piecewise

فاکتورها	ضرایب رگرسیون	خطای استاندارد	P-value
سن	۰,۰۶۰	۰,۰۰۶	۰,۰۰۱>
دیابت	۰,۷۲۴	۰,۱۴۵	۰,۰۰۱>
سابقه ایسکمی	۰,۳۶۸	۰,۱۴۶	۰,۰۱۲
انفارکتوس حادترانس مورال با محل غیر مشخص	۱,۰۴۳	۰,۴۳۷	۰,۰۱۷
انفارکتوس حاد غیر مشخص	۰,۵۴۸	۰,۲۱۳	۰,۰۱۰
مصرف استرپتوکیناز	-۰,۶۹۷	۰,۲۰۳	۰,۰۳۰
Time(۰-۵۶۸)	۱,۳۶۰	۰,۲۰۵	۰,۰۰۱>
Time(۵۶۸-۱۱۳۵)	۱,۵۲۷	۰,۲۳۵	۰,۰۰۱>
Time(۱۱۳۵-۱۶۹۸)	۰,۸۱۱	۰,۲۳۱	۰,۰۰۱>

و نوع بیماری در مدل نهایی همچنان معنادار باقی ماندند و در مدل رگرسیونی تکه ای (Piecewise) با در نظر گرفتن چهار بازه مساوی صفر تا ۵۶۸ روز، ۵۶۸ روز تا ۱۱۳۵ روز، ۱۱۳۵ روز تا ۱۶۹۸ روز و ۱۶۹۸ روز تا ۲۲۶۶ روز متغیرهای سن، دیابت، سابقه ایسکمی و نوع بیماری و مصرف استرپتوکیناز در مدل نهایی باقی می ماند. در این مدل رگرسیونی دسته چهارم بعنوان رفرنس کتگوری توسط نرم افزار محسوب شده و در مدل قرار نمی گیرد. با توجه به نتایج بدست آمده از رگرسیون کاکس افزایش سن، داشتن دیابت، سابقه ایسکمی، نوع مداخله انجام شده پیس-میکر، انفارکتوس حاد ترانس مورال با محل غیر مشخص و انفارکتوس حاد غیر مشخص داری ضریب رگرسیونی مثبت معنی دار می باشند به این معنا که داشتن این فاکتورها ریسک خطر بالاتری برای مرگ را نسبت به گروه های دیگر داراست در حالی که متغیرهای مصرف استرپتوکیناز و همچنین مداخله انجام شده بای پس دارای ریسک خطر پایین تر برای مرگ و ضرایب رگرسیونی منفی هستند.

در مدل رگرسیون تکه ای که در سه دسته صفر تا ۵۶۸ روز، ۵۶۸ روز تا ۱۱۳۵ روز، ۱۱۳۵ روز تا ۱۶۹۸ روز معنی دار می باشد، متغیرهای سن، دیابت سابقه ایسکمی، انفارکتوس حاد ترانس مورال با محل غیر مشخص و انفارکتوس حاد غیر مشخص دارای دارای ضریب مثبت معنی دار و ریسک خطر بالا و متغیر مصرف استرپتوکیناز دارای ضریب منفی و ریسک خطر پایین می باشد. جالب توجه است که ریسک فاکتور نوع مداخله اعم از بای پس یا پیس میکر در مدل رگرسیون تکه ای معنی دار نشد. در مقایسه دو مدل رگرسیونی کاکس و تکه ای (Piecewise) درمیابیم که متغیرهای معنی دار در هر دو مدل یکسان هستند بجز ریسک فاکتور نوع مداخله در مدل رگرسیون کاکس معنی دار شده ولی در مدل رگرسیون تکه ای معنی دار نمیباشد. سایر ریسک فاکتورها در هر دو مدل معنی دار و از نظر اندازه و جهت اثرگذاری یکسان هستند.

بعنوان نتیجه گیری از مطالعه حاضر و با توجه به مقایسه دو مدل رگرسیونی کاکس و تکه ای (Piecewise) میتوان گفت در داده های بیمارستانی بخصوص وقتی جمع آوری داده ها از پرونده های پزشکی باشد اگر زمان مرگ بیماران (event in survival data) بطور دقیق ثبت نشده باشد میتوان با در نظر گرفتن زمان فوت بصورت بازه ای مدل رگرسیونی تکه ای (Piecewise) را استفاده کرد.

۵ نتیجه گیری

متغیرهای باقی مانده در مدل نهایی رگرسیون کاکس شامل: سن (با ضریب ۰,۰۴۱، $p\text{-value} < ۰,۰۰۱$)، ابتلا به دیابت (با ضریب ۰,۰۶۵، $p\text{-value} < ۰,۰۰۱$)، داشتن سابقه ایسکمی (با ضریب ۰,۳۵۸، $p\text{-value} = ۰,۰۱۸$)، انفارکتوس حاد غیر مشخص (با ضریب ۱,۲۳۸، $p\text{-value} = ۰,۰۰۵$)، انفارکتوس حاد غیر مشخص (با ضریب ۰,۴۰۶، $p\text{-value} = ۰,۰۵۰$)، مصرف استرپتوکیناز (با ضریب -۰,۳۳۹، $p\text{-value} = ۰,۰۴$)، مداخله انجام شده-بای پس (با ضریب -۱,۱۷، $p\text{-value} = ۰,۰۱۷$) و مداخله انجام شده-پیس میکر (با ضریب ۱,۰۹۱، $p\text{-value} = ۰,۰۰۱$) بوده است.

در حالی که متغیرهای باقی مانده در مدل نهایی رگرسیون تکه ای در چهار دسته معنا دار صفر تا ۵۶۸ روز، ۵۶۸ روز تا ۱۱۳۵ روز، ۱۱۳۵ روز تا ۱۶۹۸ روز و ۱۶۹۸ روز تا ۲۲۶۶ روز شامل: سن (با ضریب ۰,۰۶، $p\text{-value} < ۰,۰۰۱$)، ابتلا به دیابت (با ضریب ۰,۷۲۴، $p\text{-value} < ۰,۰۰۱$)، سابقه ایسکمی (با ضریب ۰,۳۶۸، $p\text{-value} = ۰,۰۱۲$)، انفارکتوس حاد غیر مشخص (با ضریب ۱,۰۴۳، $p\text{-value} = ۰,۰۱۷$)، انفارکتوس حاد غیر مشخص (با ضریب ۰,۵۴۸، $p\text{-value} = ۰,۰۱$) و مصرف استرپتوکیناز (با ضریب -۰,۶۹۷، $p\text{-value} = ۰,۰۳۰$)، تنها متغیر نوع مداخله انجام گرفته که در مدل رگرسیون کاکس معنی دار بود، در مدل رگرسیونی تکه ای حذف شد.

در مطالعه حاضر مدل رگرسیون کاکس و مدل رگرسیون piecewise از نظر شدت و جهت اندازه اثر ریسک فاکتورها، مانند هم عمل کردند. بنظر میرسد جاهایی که زمان بقاء بدرستی و بصورت دقیق ثبت نشده است میتوان از مدل رگرسیون piecewise استفاده کرد.

مراجع

- [1] Allison Paul.(2010) *Survival Analysis using SAS, Practical Guide*, Second Edition.
- [2] Allison Paul.(1995) *How to do Piecewise Exponential Survival Analysis in Stata*, W.M. Mason, Soc 213B, S 02, UCLA 1995.
- [3] Bochud PY, Stauffer JC, Mottet JJ, Regamey C.(1998) *Epidmiology and care of myocardial infarction at the fribourg conton hospital 1995*, Schweiz – Med–Wochonschr. 128(10):363-73.

- [4] Cabeu JP , Gones N , Vaur L , Renault M , Etienne S, Ferrieres J , Danchin N.(1997) *Epidemiology of myocardial infarction in France*, Arch-MolCoeur-Vaiss. 90(11) : 1511-9.
- [5] Ghaffarian Shirazi H, Javan A, Hatamipoor E, Moosavizadeh A, Ghaedi H, Mohammadi Baghmallaie M, et al.(2006) *Survival Rate and its Related Factors in Patients with Acute Myocardial Infarction*. Armaghane-danesh, 11(1):93-104.
- [6] khosravi A, Ebrahimi H. (1387) *Evaluation of 1-Year Survival Rate and Effective Factors in Patients with Acute Myocardial Infarction*. knowlege health J, 3(1):1-7.
- [7] Last J, cummigs S.(1993) *One year survival in acute myocardial infarction*. Lancet, 341:72-75.
- [8] Mosa Farkhani Ehsan, Baneshi Mohammad Reza, Zolala Farzaneh.(2014) *Survival Rate and its Related Factors in Patients with Acute Myocardial Infarction*. Medical Journal of Mashhad University of Medical Sciences.
- [9] Nabaei B, Maghsoodloo M, Dabiran S. (2002) *Evaluation of survival rate and effective factors in acute myocardial infarction patients in Emam Hospital (Year 2000)*. Tehran Univ Med J. 60(4):347-350.