بسم الله الرحمن الرحيم

دانشکده علوم پایه
گروه آمار

پایان‌نامه جهت اخذ درجه کارشناسی ارشد
رشته آمار گرایش ریاضی

عنوان پایان نامه

انواع توزیع های دومتغیره نمایی

استادان راهنما

دکتر داود قزوینی نژاد
دکتر مریم شرفی

نگارش

مریم سیفی

اسفند ماه ۱۳۹۲

دانشکده علوم پایه
گروه آمار

پایان‌نامه جهت اخذ درجه کارشناسی ارشد
رشته آمار گرایش ریاضی

نام دانشجو:
مریم سیفی

تحت عنوان :

انواع توزیع های دومتغیره نمایی و کاربرد آن

در تاریخ ۱۳۹۲/۱۲/۱۲توسط هیأت داوران زیر بررسی و با درجه بسیار خوب به تصویب نهایی رسید.

۱- استاد راهنما:	دکتر داودقزوینی نژاد	بامرتبه ی علمی استادیار	امضاء
۲- استاد راهنما:	دکتر مریم شرفی	بامرتبه ی علمی استادیار	امضاء
۳- استاد داور:	دکتر بهاالدین خالدی	با مرتبه علمی استاد	امضاء
۴- استاد داور:	دکتر عبداله جلیلیان	با مرتبه علمی استادیار	امضاء

چکیده

هرگاه موضوع قابلیت اعتماد مطرح می شود؛ یکی از اولین مسائلی که در ذهن تداعی می گردد شکست یک واحد یا مرگ یک موجود زنده است. از بین توزیع‌های آماری، به دلیل ذات توزیع نمایی، این توزیع می تواند گزینه مناسبی برای مدل بندی شکست‌ها باشد. در حالت دو متغیره ما با وابستگی‌های بین دو مؤلفه رو به رو هستیم. بنابراین یک خانواده وسیعی از توزیع‌های دو متغیره وجود دارد. در این رساله، ما علاقه‌مند به معرفی سه نوع توزیع دو متغیره با نام هایی توزیع نمایی گامبل، توزیع نمایی فروند و توزیع نمایی مارشال-الکین هستیم. همچنین برخی از خواص آماری مدل‌های معرفی شده مانند برآوردگرهای حداکثر درستنمایی پارامترها، توابع مولد گشتاور، ضریب همبستگی و اطلاع فیشر مورد توجه است. در پایان، به عنوان یک مثال کاربردی از توزیع نمایی فروند و مارشال -الکین، به منظور مدل بندی تأخیرهای شبکه استفاده می‌نماییم به طوری که برآوردگرهای ماکسیمم درستنمایی و فواصل اطمینان برای پارامتر جبران تأخیر زمان به دست آورده می شود.

کلید واژه ها:

برآورد ماکسیمم درستنمایی، تأخیر شبکه، تابع بقا، توزیع دو متغیره نمایی، جبران تأخیر زمان، فاصله اطمینان، قابلیت اعتماد، هماهنگ سازی زمان.

پیشگفتار

اواخر دهه ۴۰ و اوایل دهه ۵۰ مساله غیر قابل اطمینان بودن وسائلی که در صنعت هوانوردی بکار برده می شده، مطرح گردید. افزایش دقت و قابلیت اعتماد تجهیزات پیچیده الکترونیکی انگیزه هایی بودند که موجب پیشرفت نظریه قابلیت اعتماد گردید.

نظریه قابلیت اعتماد بخشی ازکنترل کیفیت است که درآن مشخصه مورد مطالعه، طول عمر قطعه است، ولی به دلیل پرهزینه بودن جمع آوری داده های طول عمر، سعی می شود از حداقل داده ها، بیشترین اطلاعات استنتاج گردد. بنابراین در مقابل چارت های کنترلی ساده، تمرکز روی مطالعه انواع توزیعهای موجود در نظریه قابلیت اعتماد وجود دارد. برای تعیین قابلیت اعتماد یک سیستم به شرط داشتن قابلیت اعتماد مؤلفه هایش، معمولا فرض استقلال بین شکست مؤلفه ها را در نظر می گیرند. زیرا داشتن چنین فرضی محاسبه قابلیت اعتماد سیستم را ساده می نماید. اما می دانیم که محاسبه احتمال توأم رخدادها به کمک حاصل ضرب احتمال آن رخدادها بدون در نظرگرفتن استقلال بین آنها معنی دار نخواهد بود و در یک سیستم وابستگی های معینی به دلیل عوامل کارکرد یا همبستگی های درونی بین مؤلفه ها و زیرسیستم های آن وجود دارد. بنابراین در بسیاری از موارد فرض داشتن استقلال بین شکست ها، خطاهای بزرگی را برای پذیرش ایجاد می کند. به منظور گذر از این مشکل بایستی توزیع توأم زمان شکست مؤلفه هائی که ساختار یک سیستم را تشکیل می دهند بدست آورد و آنگاه به کمک آن به محاسبه قابلیت اعتماد آن پرداخت.

دربسیاری از بررسی های آماری که در آن دومتغیرتصادفی وابسته مطالعه می شوند، از توزیع نرمال دومتغیره به عنوان مدل مناسب استفاده می شود اگرچه در این حالت توزیع های حاشیه ای نیز نرمال هستند ولی چنین مدلی در حالت کلی مدل مناسبی برای توصیف زمان های شکست در یک سیستم نمی باشد، بلکه مدل دومتغیره ایی مناسب است که دارای توابع چگالی حاشیه ای به فرم نمایی یا وایبل و یا مشابه با آنها باشد. توزیعهای نمایی نقش اساسی در آزمون طول عمر، قابلیت اعتماد و دیگر زمینه های کاربردی ایفا می کنند. از این توزیع برای مدلبندی عدم اطمینان متغیر تصادفی طول عمر استفاده می کنند. به دلیل چنین کاربرد وسیع از توزیع نمایی یک متغیره، طبیعی است

که از توزیع های نمایی دو (یا چند) متغیره به عنوان مدلهایی برای سیستم هایی با دو(یاچند) مؤلفه استفاده کنیم. اما برخلاف توزیع نرمال، روش خاص و مشخصی برای توسیع توزیع نمایی یک متغیره به دو(یا چند) متغیره در دسترس نیست. بنابراین اغلب برای یافتن توزیع توأم طول عمر مؤلفه های یک سیستم فرض استقلال بین مؤلفه ها را در نظر می گیرند، که دراکثر موارد چنین فرضی سؤال برانگیز بوده یا کاملا غلط است. بنابراین نیاز به توزیعهای نمایی دو (یا چند) متغیره (یعنی توزیع های دومتغیره با چگالی های حاشیه ای نمایی) محسوس است. توزیع دومتغیره نمایی می تواند رفتار شکست یک سیستم دومؤلفه ایی را که طول عمر مؤلفه های آن از هم مستقل نیستند و درعمل به یکدیگر وابسته هستند، شرح دهد. برخی از توزیع های نمایی دومتغیره به کمک روش هایی بدست آمده اند که کاربرد آنها در عمل چندان واضح و روشن نیست وهیچ مدل و یا پایه هایی برای تعیین اینکه چطور این توزیع ها می توانند درعمل ایجاد شوند، در اختیار نیست.

در فصل دوم این رساله سه مدل از توزیع‌های دو متغیره نمایی ارائه شده توسط گامبل معرفی و خصوصیاتی همچون گشتاورهای این توزیع، توابع چگالی حاشیه‌ای و منحنی رگرسیونی این مدل‌ها مورد بحث و بررسی قرار می گیرند. سپس ضریب همبستگی هرکدام ازمدل های متغیرهای تصادفی، برآورد پارامترها از طریق روش برآوردیابی ماکسیمم درستنمایی و نیز اطلاع فیشر بدست آورده خواهد شد. از آنجایی که در سیستم های دومؤلفه ای رخداد ماکسیمم (درسیستم های موازی) یا مینیمم (درسیستم های سری) طول عمر مؤلفه ها معادل با شکست کل سیستم است، مطالعه توزیع آماره های ترتیبی مینیمم و ماکسیمم ضروری به نظر می رسد. و در پایان از طریق شبیه سازی برای مدل نوع اول گامبل به تولید نمونه تصادفی و نیز برآورد پارامتر مورد نظر خواهیم پرداخت.

در فصل سوم این رساله توزیع های دو متغیره نمایی ارائه شده توسط فروند و مارشال-الکین مورد بررسی قرار می گیرند. سپس روش ساخت هر کدام را به طور جداگانه بیان کرده و گشتاورهای متغیرهای تصادفی آنها بدست آورده می شود و همچنین از طریق روش ماکسیمم درستنمایی به برآورد پارامترهای هر کدام از مدل ها پرداخته می شود.

یکی از کاربردهای توزیع دو متغیره نمائی مدل سازی شبکه ها مثل شبکه کامپیوتری است.لذا در فصل چهارم این رساله با استفاده از توزیع دومتغیره نمایی فروند و مارشال-الکین به مدل سازی در

تأخیر شبکه می پردازیم. برآورد پارامتر های آن را بدست آورده و در نهایت فاصله اطمینانی برمبنای این توزیع ها برای جبران تأخیر زمانی محاسبه خواهیم کرد.

فهرست مطالب

فهرست مطالب آ

لیست جداول ت

لیست تصاویر ث

۱ تعاریف و اصطلاحات اولیه ۱

۱.۱ مقدمه ۲

۲.۱ تعاریف ۴

۳.۱ توزیع دومتغیره ۶

۱.۳.۱ توابع گشتاوری ۷

۲.۳.۱ برآورد پارامتر ۸

۲ مدل های گامبل ۱۰

۱.۲ مقدمه ۱۱

۲.۲ مدل گامبل نوع اول ۱۱

۱.۲.۲ ساختار مدل ۱۱

۲.۲.۲ گشتاورها ۱۶

۳.۲.۲ برآورد پارامتر مدل ۲۰

۴.۲.۲ توزیع آماره های ترتیبی مینیمم و ماکسیمم ۲۳

۳.۲ مدل گامبل نوع دوم ۲۸

۱.۳.۲ ساختار مدل ۲۸
۲.۳.۲ گشتاورها ۲۹
۳.۳.۲ برآورد پارامتر مدل ۳۳
۴.۳.۲ توزیع آماره های ترتیبی مینیمم و ماکسیمم ۳۴
۴.۲ مدل گامبل نوع سوم ۳۷
۵.۲ بحث و نتیجه گیری ۳۸
۶.۲ مطالعات شبیه سازی ۳۹
۱.۶.۲ تولید نمونه تصادفی از توزیع دومتغیره نمایی ۳۹
۲.۶.۲ برآورد پارمترهای توزیع دومتغیره نمایی ۴۱

۳ مدل های فروند و مارشال-الکین ۵۰

۱.۳ مقدمه ۵۱
۲.۳ مدل فروند ۵۱
۱.۲.۳ روش ساخت ۵۲
۲.۲.۳ خواص مدل ۵۵
۳.۲.۳ بحث و نتیجه گیری ۶۱
۳.۳ مدل مارشال-الکین BVE ۶۱
۱.۳.۳ روش ساخت ۶۲
۲.۳.۳ خواص مدل ۷۲
۳.۳.۳ بحث و نتیجه گیری ۸۲

۴ کاربرد توزیع های دو متغیره نمایی فروند و مارشال-الکین در تحلیل تأخیر شبکه ها ۸۳

۱.۴ مقدمه ۸۴
۱.۱.۴ تأخیر شبکه ۸۸
۲.۴ توزیع تأخیر شبکه با استفاده از مدل فروند ۹۰

۱.۲.۴ برآورد درستنمائی پارامترهای تأخیرشبکه ۹۲
۳.۴ توزیع تأخیر شبکه با استفاده از مدل مارشال-الکین ۹۳
۱.۳.۴ برآورد درستنمائی پارامترهای تأخیرشبکه ۹۵
۴.۴ فاصله اطمینان برای جبران تأخیر زمانی ۹۶
۱.۴.۴ فاصله اطمینان برای تأخیر شبکه با استفاده از مدل فروند ۹۷
۲.۴.۴ فاصله اطمینان برای تأخیر شبکه با استفاده از مدل مارشال-الکین . . . ۱۰۰
۵.۴ بحث و نتیجه گیری ۱۰۲

نام نامه **۱۰۴**

واژه‌نامه فارسی به انگلیسی **۱۰۷**

مراجع **۱۰۹**

لیست جداول

۱.۴ مقدار ماکسیمم ضریب همبستگی ۹۲

لیست تصاویر

۱.۲ تابع توزیع دومتغیره نمایی مدل نوع اول گامبل ۱۲

۲.۲ تابع چگالی توزیع دو متغیره نمایی نوع اول گامبل ۱۴

۳.۲ منحنی رگرسیون و انحراف استاندارد و ضریب همبستگی برای توزیع دو متغیره نمایی نوع اول گامبل ۱۷

۴.۲ نمودار اطلاع فیشر ۲۲

۵.۲ منحنی رگرسیون برای توزیع دو متغیره نمایی نوع دوم گامبل ۳۰

۱.۴ زمان بندی پیام ها بین دو گره A وB ۸۸

۲.۴ نمودار بلوکی توزیع فروند برای تبادل پیام در یک شبکه دوگره ای ۹۱

۳.۴ نمودار بلوکی قابلیت اعتماد توزیع مارشال-الکین برای یک شبکه دوگره ای . . ۹۴

فصل ۱
تعاریف و اصطلاحات اولیه

۱.۱ مقدمه

در این فصل برای روشن شدن مفهوم توزیع نمایی دومتغیره و نیز بررسی موضوع اصلی در فصل های بعد، از چند تعریف و نماد که تا پایان این رساله مورد استفاده قرار می گیرد بیان می شود.

اواخر دهه ۴۰ و اوایل دهه ۵۰ مساله غیر قابل اعتماد بودن وسائلی که در صنعت هوانوردی بکار برده می شده، مطرح گردید. افزایش دقت و قابلیت اعتماد تجهیزات پیچیده الکترونیکی انگیزه هایی بودند که موجب پیشرفت نظریه قابلیت اعتماد گردید.

نظریه قابلیت اعتماد بخشی ازکنترل کیفیت است که درآن مشخصه مورد مطالعه، طول عمر قطعه است، ولی به دلیل پرهزینه بودن جمع آوری داده های طول عمر، سعی می شود از حداقل داده ها، بیشترین اطلاعات استنتاج گردد. بنابراین در مقابل نمودار های کنترلی ساده، تمرکز روی مطالعه انواع توزیعهای موجود در نظریه قابلیت اعتماد وجود دارد.

واژه قابلیت اعتماد برای درجه بندی از اعتماد به کار میرود که براساس آن، یک سیستم در محیط مشخص و در دوره زمانی معین، کار از پیش تعریف شده ای را انجام می دهد. در مبحث قابلیت اعتماد مشخصه مورد مطالعه زمان شکست یا طول عمر مؤلفه (سیستم یا فرد) است. سیستم یا ماشین نامی بسیار آشناست و از جنبه نحوه کارکرد آن، در زمینه های فنی و مهندسی بسیار پرکاربرد است. موتور یک اتومبیل، یک وسیله الکتریکی مانند انواع لامپها، یک قطعه الکتریکی مانند IC وبسیاری دیگر همه مثال هایی از یک سیستم هستند. طول عمر مؤلفه برابر است با طول فاصله زمانی از آغاز فعالیت تا زمان شکست. با توجه به این که طول عمر همواره در معرض تغییرات است . بنابراین طبیعی است که آن را به عنوان یک متغیرتصادفی غیرمنفی مانند X در نظر بگیریم.

برای تعیین قابلیت اعتماد یک سیستم به شرط داشتن قابلیت اعتماد مؤلفه هایش، معمولا فرض استقلال بین شکست مؤلفه ها را در نظر می گیرند. زیرا داشتن چنین فرضی محاسبه قابلیت اعتماد سیستم را ساده می نماید. اما می دانیم که محاسبه احتمال توأم رخدادها به کمک حاصل ضرب احتمال آن رخدادها بدون در نظرگرفتن استقلال بین آنها معنی دار نخواهد بود و در یک سیستم

وابستگی های معینی به دلیل عوامل کارکرد یا همبستگی های درونی بین مؤلفه ها و زیرسیستم های آن وجود دارد. بنابراین در بسیاری از موارد فرض داشتن استقلال بین شکست ها، خطاهای بزرگی را برای پذیرش ایجاد می کند.

به منظور گذر از این مشکل بایستی توزیع توأم زمان شکست مؤلفه هائی که ساختار یک سیستم را تشکیل می دهند بدست آورد و آنگاه به کمک آن به محاسبه قابلیت اعتماد آن پرداخت.

دربسیاری از بررسی های آماری که در آن دو متغیرتصادفی وابسته مطالعه می شوند، از توزیع نرمال دومتغیره به عنوان مدل مناسب استفاده می شود. اگرچه در این حالت توزیع های حاشیه ای نیز نرمال هستند ولی چنین مدلی در حالت کلی مدل مناسبی برای توصیف زمان های شکست در یک سیستم نمی باشد، بلکه مدل دومتغیره ایی مناسب است که دارای توابع چگالی حاشیه ای به فرم نمایی یا وایبل و یا مشابه با آنها باشد.

توزیع نمائی تجزیه و تحلیل آسانی دارد از اینرو درفیزیک آماری نقش بسزایی ایفا می کند. درسال های اخیر از این توزیع بعنوان اولین روش برای آزمون بقا (قابلیت اعتماد) استفاده کردند.

ازتوزیع دومتغیره نمایی درحالت تعمیم یافته نیز می توان برای آزمون بقای یک سیستم دو مؤلفه ای که یکی از آن مؤلفه ها با شکست مواجهه شده است، استفاده نمود به طور مثال می توان کارایی چشم ،گوش،کلیه یا هرجفت اندام دیگر را اشاره نمود،(فروند،(۱۹۶۱)).

درچهار دهه گذشته توزیع های دومتغیره گوناگونی، مخصوصا تعدادی که برای تحلیل قابلیت اعتماد بکار می روند، ایجاد شده است. مورگنسترن در سال ۱۹۵۶ فرمول کلی برای توزیع های دومتغیره ارائه داد بعد از آن گامبل در خلال سالهای ۱۹۶۷-۱۹۶۰ به مطالعه فرمول کلی برای توزیع های دومتغیره نمائی و لجستیک پرداخت، علاوه بر آن توزیع های دومتغیره نمایی توسط آماردانایی همچون فارلی در سال ۱۹۶۰، فروند(۱۹۶۱)و مارشال -الکین در سال ۱۹۶۷ پیدا شد.

از سال ۱۹۷۰ به بعد مدل بندی توزیع های دومتغیره و چندمتغیره به کمک مدل های شوک در قابلیت اعتماد پایه اصلی تحقیقات آماری چون بمیز، بین و هیگینز، پروشان و سالو، بلاک و باسو، سرکور، مارشال در سال ۱۹۷۵، بلاک ۷۷-۱۹۷۵، فردی و پاتیل(۱۹۷۷)، شیکد(۱۹۷۷)، هالمز (۱۹۸۰) ، سینگوروالا (۱۹۸۶)، باسو(۱۹۸۷)، باسو و کای سان ۹۷-۱۹۹۳، همدانی۱۹۸۰-۱۹۹۵ بوده

است. توزیع های دومتغیره به کمک توزیعهای شرطی در خانواده توزیعهای نمایی نیز توسط افرادی مانند آرنولد و استراس در سال ۱۹۸۸ مورد بررسی قرار گرفت.

یکی از توزیع های آماری بسیارمهمی که در قابلیت اعتماد نقش اساسی ایفا می کند، توزیع نمایی است. چراکه این توزیع برای مدلبندی عدم اعتماد متغیر تصادفی طول عمر بسیار به کار می رود. یکی از کاربردهای توزیع نمایی مطالعه سیستم هایی است که از کارافتادگی و شکست هریک یا تعدادی از مؤلفه های آن باعث از کارافتادگی کل سیستم می شود. همچنین در تجهیزاتی که علل عمده شکست و از کارافتادگی ها بصورت تصادفی است ، بطور مثال می توان به وارد آمدن بار اضافی به سیستم های الکتریکی مانند نوسانات جریان برق اشاره نمود. استفاده از این تابع توزیع می تواند راهگشا باشد. همچنین از توزیع نمایی در تخمین زدن مدت زمان لازم برای رخداد یک پیشامد خاص نیز استفاده می‌شود. برای نمونه، مدت زمان لازم (از هم‌اکنون) تا رخداد یک زمین‌لرزه، آغاز یک جنگ، دریافت یک تماس تلفنی اشتباه و غیره متغیرهای تصادفی با توزیع نمایی می‌باشند. توزیع های نمایی تک متغیره در توصیف طول عمر تک مؤلفه ها اهمیت دارند باسو(۱۹۹۵). درحالی که برای توصیف طول عمر سیستم های دو مؤلفه ای با یکدیگر و ارتباط و همبستگی بین سیستم ها و مؤلفه ها، توزیع دومتغیره با حاشیه نمایی مفید واقع می شود.

۲.۱ تعاریف

در اینجا تعاریف و رابطه های مربوط به کل رساله را بیان می کنیم.

تعریف ۱.۲.۱. سیستم دو مولفه ای: به سیستمی گویند که از دو مؤلفه برای جریان تشکیل شده است. مثل سیستم های موازی و سری.

تعریف ۲.۲.۱. سیستم سری: سیستمی ست که مؤلفه های آن پشت سر هم و در یک خط قرار گرفته باشند و زمانی که یکی از مؤلفه ها متوقف شود کل سیستم متوقف می شود.

تعریف ۳.۲.۱. سیستم موازی : سیستمی ست که مؤلفه های آن بصورت موازی قرار گرفته باشند و زمانی سیستم متوقف می شود که همه مؤلفه های سیستم متوقف شده باشد.

تعریف ۴.۲.۱. فرآیند پواسن: فرآیند پواسن یک فرآیند شمارشی است که تعداد رخداد یک پیشامد خاص را در طول زمان می شمارد. مثالهایی از این نوع فرآیند عبارتند از ورود مشتریان به یک باجه، وقوع زلزله در یک ناحیه معین، قطع برق و غیره. این فرآیند یک ابزار بدیهی برای بسیاری از مسائل کاربردی است. چندین تعریف برای این فرآیند وجود دارد که معادل یکدیگرند.

- تعریف۱: فرآیند شمارشی $\{N(t), t \geq ۰\}$ را یک فرآیند پواسن با نرخ $\lambda > ۰$ گوییم، هرگاه:

۱) $N(t) = ۰$

۲) فرآیند دارای نموهای مستقل باشد.

۳) تعداد پیشامدها در هر فاصله دلخواه به طول t دارای توزیع پواسن با میانگین λtباشد، یعنی $E(N(t)) = \lambda t$. این مطلب به این معنی است که در این فرآیند نموها ایستا است، یعنی به ازای هر $s, t \geq ۰$:

$$P(N(t+s) - N(s) = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

- تعریف دیگر از یک فرآیند پواسن به این صورت است که:

فرآیند شمارشی $\{N(t), t \geq ۰\}$ را یک فرآیند پواسن با نرخ $\lambda > ۰$ گوییم، اگر زمان های بین رخدادهای $X_۱, X_۲, ...$ دارای توزیع نمایی به صورت زیر باشند:

$$P(X_n \leq x) = ۱ - e^{-\lambda x} \qquad , x \geq ۰$$

فرض کنیم $X_۱$ زمان انتظار تا وقوع اولین پیشامد باشد، و به ازای $n \geq ۱$ ، فرض کنیم X_n زمان بین $n-۱$ مین و nمین پیشامد باشد. در اینصورت دنباله $\{X_n, n \geq ۱\}$ را دنباله زمان های بین ورود می نامند. X_nها، $n = ۱, ۲, ...$ متغیرهای تصادفی مستقل و همتوزیع با توزیع نمایی و میانگین $\frac{۱}{\lambda}$ بوده و این فرآیند بی حافظه است، یعنی این فرآیند دارای نموهای مستقل و ایستا است.

تعریف ۵.۲.۱. ضریب همبستگی: شاخصی است آماری که جهت و مقدار رابطه بین دو متغیر را توصیف میکند. ضریب همبستگی درمورد توزیع های دو یا چند متغیره نیز به کار می رود. اگر مقادیر دو متغیر شبیه هم تغییر کند یعنی با کم یا زیاد شدن یکی دیگری نیز کم یا زیاد شود به گونه ای که بتوان رابطه آنها را به صورت یک معادله بیان کرد گوییم بین این دو متغیرهمبستگی وجود دارد.
ضریب همبستگی به عنوان معیاری برای سنجش تغییراتX و Yنسبت به یکدیگر و دارای خواص