بسم الله الرحمن الرحیم

دانشکده علوم ریاضی

گروه آمار

پایان‌نامه

برای دریافت درجه کارشناسی ارشد در رشته

آمار ریاضی

عنوان

ویژگی‌هایی از مدل‌های شکنندگی

استاد راهنما

دکتر محمد امینی

استاد مشاور

دکتر غلامرضا محتشمی برزادران

نگارنده

معصومه بزرگ

زمستان ۱۳۹۱

بسمه تعالی

مشخصات پایان‌نامه تحصیلی دانشجویان

دانشگاه فردوسی مشهد

عنوان: ویژگی‌هایی از مدل‌های شکنندگی

نام نویسنده: معصومه بزرگ

استاد راهنما: دکتر محمد امینی

استاد مشاور: دکتر غلامرضا محتشمی برزادران

دانشکده: علوم ریاضی گروه: آمار	رشته تحصیلی: آمار ریاضی
تاریخ تصویب: ۱۳۹۱/۱۰/۱۹	تاریخ دفاع: ۱۳۹۱/۱۱/۹
مقطع تحصیلی: کارشناسی ارشد	تعداد صفحات: ۱۵۲

چکیده پایان‌نامه : در این پایان‌نامه ابتدا تعاریفی از توابع قابلیت اعتماد، ترتیب‌های تصادفی، ویژگی‌هایی از توزیع طول عمر، مفاهیم وابستگی و تابع مفصل ارائه گردیده است. پس از آن مدل مخاطره متناسب کاکس، مدل زمان شکست شتابیده و مدل شکنندگی معرفی و انواع مدل‌های شکنندگی تبیین گردیده است. سپس مدل‌های آمیخته مخاطره جمعی، وابستگی‌ها و مقایسه‌های تصادفی در این مدل‌ها مورد بحث قرار گرفته است. در ادامه مدل‌های شکنندگی وایبل-گاما، گشتاورهای نامتناهی این مدل‌ها و ارتباط آن با توزیع‌های لگ-لجستیک تعمیم یافته، لجستیک، کوشی و مقدارفرین ارائه گردیده است. علاوه بر این مفصل تولیدشده توسط مدل‌های شکنندگی، شباهت‌ها و تفاوت‌های بین مدل مفصل و مدل‌های شکنندگی مورد توجه قرار گرفته است. در پایان مدل‌های شکنندگی در محیط R تجزیه و تحلیل شده است.

واژگان کلیدی: شکنندگی، مخاطره جمعی، ترتیب تصادفی،مدل آمیخته

امضای استاد راهنما: تاریخ:

اظهارنامه

عنوان پایان‌نامه : ویژگی‌هایی از مدل‌های شکنندگی

اینجانب معصومه بزرگ دانشجوی دوره کارشناسی ارشد دانشکده علوم ریاضی دانشگاه فردوسی مشهد نویسنده پایان‌نامه تحت راهنمایی دکتر محمد امینی متعهد می‌شوم:

آ. تحقیقات در این رساله توسط اینجانب انجام شده و از صحت و اصالت برخوردار است.

ب. در استفاده از نتایج پژوهش‌های محققان دیگر به مرجع مورد استفاده استناد شده است.

ج. مطالب مندرج در این پایان‌نامه تاکنون توسط خود یا فرد دیگری برای دریافت هیچ نوع مدرک یا امتیازی به جایی ارائه نشده است.

د. کلیه حقوق این اثر متعلق به دانشگاه فردوسی مشهد است و مقالات مستخرج با نام ”دانشگاه فردوسی مشهد “ و یا ”Ferdowsi University of Mashhad“ به چاپ خواهد رسید.

ه. حقوق معنوی تمام افرادی که در به دست آمدن نتایج اصلی رساله تاثیرگذار بوده‌اند در مقالات مستخرج از آن رعایت شده است.

و. در کلیه مراحل انجام این رساله، در مواردی که از موجود زنده(یا بافت‌های آن‌ها) استفاده شده، ضوابط و اصول اخلاقی رعایت شده است.

ز. در کلیه مراحل انجام این رساله، در مواردی که به حوزه اطلاعات شخصی افراد دسترسی یافته یا استفاده شده، اصل رازداری، ضوابط و اصول اخلاقی انسانی رعایت شده است.

تاریخ

امضای دانشجو

تقدیم به

ودیعه یزدان، موعود ادیان و منجی انسان

شخصیتی که چون بیاید

بهار به پیشوازش رود

عدل بر رکابش بوسه زند

و جهان یکسره به صلاح و فلاح گراید

و هم اینک نیز که به صلاحدید حق، رخ در نقاب غیبت دارد،

بارش آسمان و رویش زمین، از یمن وجود اوست.

// سپاس گزاری...

سپاس پروردگار بی‌همتا که باران رحمت خود را بر من فرو بارید و نعمت دانش آموزی را نصیبم ساخت.
هم اکنون که به لطف خداوند این دوره تحصیلی را به پایان رسانده‌ام بر خود واجب می دانم از کلیه عزیزانی که در تمامی مراحل این پایان‌نامه مرا راهنمایی و یاری کرده‌اند، سپاس‌گزاری نمایم.

با عرض امتنان از الطاف پر ارج استاد راهنمای ارجمندم، جناب آقای دکتر محمد امینی که همواری این راه، وامدار همیاری و همراهی بزرگوارانه ایشان بود و گام به گام این سفر مرهون هدایت خردمندانه بی‌دریغشان.

با تقدیم سپاس به محضر استاد مشاور فرزانه، جناب آقای دکتر غلامرضا محتشمی که تا همیشه مرهون راهنمایی‌ها و تعالیم شایان امتنانشان خواهم بود.

با تقدیر از عنایات بزرگوارانه اساتید فرهیخته جناب آقای دکتر هادی جباری و جناب آقای دکتر مصطفی رزمخواه که داوری این پایان‌نامه را به عهده گرفتند.

همچنین لازم می‌دانم از خانم دکتر براتپور مدیر گروه آمار و سرکار خانم سلیمانی کارشناس گروه آمار که در طول دوره تحصیل همواره راهنمای اینجانب بوده‌اند، با تمام وجود تشکر و قدردانی نمایم.

در پایان، بوسه می‌زنم بر دستان خداوندگاران مهر و مهربانی، پدر و مادر عزیزم و بعد از خدا، ستایش می‌کنم وجود مقدس‌شان را و تقدیر و تشکر خالصانه‌ی خود را به خواهر و برادران عزیزم و همسر مهربانم تقدیم می‌دارم که اگر همراهی‌ها و همدلی‌های ایشان نبود، ادامه‌ی راه برایم سخت می‌نمود.

معصومه بزرگ
زمستان ۱۳۹۱

فهرست نمادها

احتمال	P
تابع چگالی	$f(.)$
تابع توزیع	$F(.)$
تابع بقا	$\bar{F}(.)$
تابع مخاطره	$\lambda(.)$
تابع مخاطره تجمعی	$\Lambda(.)$
متوسط مخاطره	$H(.)$
تبدیل لاپلاس	$L(.)$
امید ریاضی	$E(.)$
مدل زمان شکست شتابیده	AFT
مدل مخاطره متناسب	PH
نسبت بخت	OR
نسبت درستنمایی نزولی	DLR
نرخ خطر صعودی (نزولی)	$IFR\ (DFR)$
میانگین نرخ خطر صعودی (نزولی)	$IFRA\ (DFRA)$
نو بهتر (بدتر) از کهنه	$NBU\ (NWU)$
امید نو بهتر (بدتر) از کهنه	$NBUE\ (NWUE)$
نو بهتر (بدتر) از کهنه در نفوذ تصادفی دوم	$NBU(2)\ (NWU(2))$
نو بهتر (بدتر) از کهنه در ترتیب تبدیل لاپلاس	$NBU_{Lt}\ (NWU_{Lt})$
میانگین باقیمانده عمر نزولی (صعودی)	$DMRL\ (IMRL)$
نرخ مخاطره معکوس نزولی	$DRHR$
ترتیب تصادفی معمولی	$\leq_{st}$
ترتیب نرخ مخاطره	$\leq_{hr}$
ترتیب نرخ مخاطره معکوس	$\leq_{rh}$
ترتیب مقعر صعودی	$\leq_{icv}$
ترتیب تبدیل لاپلاس	$\leq_{Lt}$
ترتیب نسبت تبدیل لاپلاس	$\leq_{Lt-r}$
به طور کامل مثبت از مرتبه ۲	TP_2
معکوس منظم از مرتبه ۲	RR_2

به طور ربعی وابسته مثبت	PQD
در دم راست صعودی (نزولی)	$RTI\ (RTD)$
به طور منفی وابسته نسبت درستنمایی	$NLRD$
در کنج راست نزولی (صعودی)	$RCSD\ (RCSI)$
ضریب تاو کندال	τ
تابع مفصل دو متغیره	$C(.,.)$

فهرست مطالب

فهرست شکل‌ها

فهرست جدول‌ها

چکیده

در این پایان‌نامه ابتدا تعاریفی از توابع قابلیت اعتماد، ترتیب‌های تصادفی، ویژگی‌هایی از توزیع طول عمر، مفاهیم وابستگی و تابع مفصل ارائه گردیده است. پس از آن مدل مخاطره متناسب کاکس، مدل زمان شکست شتابیده و مدل شکنندگی معرفی و انواع مدل‌های شکنندگی تبیین گردیده است. سپس مدل‌های آمیخته مخاطره جمعی، وابستگی‌ها و مقایسه‌های تصادفی در این مدل‌ها مورد بحث قرار گرفته است. در ادامه مدل‌های شکنندگی وایبل-گاما، گشتاورهای نامتناهی این مدل‌ها و ارتباط آن با توزیع‌های لگ-لجستیک تعمیم یافته، لجستیک، کوشی و مقدارفرین ارائه گردیده است. علاوه بر این مفصل تولیدشده توسط مدل‌های شکنندگی، شباهت‌ها و تفاوت‌های بین مدل مفصل و مدل‌های شکنندگی مورد توجه قرارگرفته است. در پایان مدل‌های شکنندگی در محیط R تجزیه و تحلیل شده است.

پیش‌گفتار

در تحقیقات پزشکی و مطالعات سرطانی، مشاهده‌ی داده‌های زمان شکست خوشه‌ای بسیار رایج است. با چشم پوشی از اثرات خوشه‌ای، ممکن است مدل مخاطره متناسب [۱] (PH) و یا مدل زمان شکست شتابیده [۲](AFT) مستقیما برای تحلیل این مجموعه داده‌ها به کار برده شود، اما این امر بدون در نظر گرفتن همبستگی ممکن در هر خوشه، منجر به امکان وجود اشتباه در نتایج می‌شود. مدل‌های شکنندگی یکی از رایج‌ترین مدل‌های استفاده شده در تحلیل داده‌های زمان شکست خوشه‌ای می‌باشد که در آن همبستگی در هر خوشه توسط یک اثر تصادفی مدل‌بندی می‌شود و بین خوشه‌ها مستقل فرض خواهد شد. در واقع اصطلاح «شکنندگی» در این مدل برای ارزیابی همبستگی در هر خوشه به کار می‌رود.

مدل‌های شکنندگی اولین‌بار در سال ۱۹۷۹ توسط واپل [۳] و همکاران، برای مطالعه برخی ناهمگنی که به وسیله متغیرهای کمکی قابل اندازه‌گیری نمی‌باشند، معرفی شد. کاربرد این مدل‌ها در تحلیل بقا برای ناهمگنی مشاهده‌نشده (پنهان) بین افراد، با فرض اینکه نرخ خطر افراد متأثر از یک کمیت مشخص به نام شکنندگی و یک مخاطره پایه باشد، مورد توجه بسیاری از محققان بوده است.

هوگارد [۴] (۱۹۸۶) مدل‌های شکنندگی را بر پایه یک توزیع پایای مثبت بررسی نمود. کلین [۵] (۱۹۹۲) الگوریتم EM [۶] را برای مدل شکنندگی گاما توسعه بخشید. گوپتا و کرمانی [۷] (۲۰۰۶) مقایسه‌های تصادفی در مدل‌های شکنندگی را از طریق تأثیر مقایسه دو متغیر تصادفی پایه بر روی متغیرهای

[1] Proportional hazard model

[2] Accelerated failure time

[3] Vaupel

[4] Hougaard

[5] Klein

[6] Expectation-Maximization Algoritm

[7] Gupta and Kirmani

شکنندگی متناظر مطالعه کرده و نتایج مفیدی بدست آوردند. ژو و لی [1] (۲۰۰۸) جنبه‌های وابستگی بین متغیر پایه و متغیر شکنندگی در این مدل‌ها را بررسی نمودند. لی و لینگ [2] (۲۰۱۲) مدل‌های شکنندگی جمعی را مطالعه نمودند. موندا [3] و همکاران (۲۰۱۲) به تحلیل مدل‌های شکنندگی پارامتری در محیط R پرداختند.

با توجه به مطالعاتی که تاکنون در زمینه مدل‌های شکنندگی انجام پذیرفته است، در این پایان‌نامه ابتدا به بیان تعاریف و مقدماتی از توابع قابلیت اعتماد، ترتیب‌های تصادفی، برخی ویژگی‌های طول عمر، تابع مفصل و برخی مفاهیم وابستگی می‌پردازیم. پس از آن در فصل دوم انواع مدل‌های آماری در قابلیت اعتماد را ازجمله مدل‌های شکنندگی معرفی می‌کنیم. مدل‌های آمیخته مخاطره‌ی جمعی و وابستگی‌ها و مقایسه‌های تصادفی در این مدل‌ها در فصل سوم مورد بررسی قرار می‌گیرد.

فصل چهارم را به مدل‌های شکنندگی وایبل-گاما، گشتاورهای نامتناهی آن‌ها و ارتباط آن با توزیع‌های لگ-لجستیک تعمیم یافته، لجستیک، کوشی، و مقدار فرین اختصاص می‌دهیم. در ادامه توجه خود را به مفصل تولید شده توسط مدل‌های شکنندگی کلاسیک و خوشه‌ای در فصل پنجم معطوف می‌کنیم و شباهت‌ها و تفاوت‌های میان مدل‌های مفصل و مدل‌های شکنندگی را بیان خواهیم نمود. در پایان از طریق نرم افزار R با چند مثال کاربردی به تحلیل مدل‌های معرفی شده می‌پردازیم.

معصومه بزرگ
بهمن ماه ۱۳۹۱

[1] Xu and Li

[2] Li and Ling

[3] Munda

فصل ۱

مفاهیم و مقدمات

۱.۱ مقدمه

در این فصل برخی مفاهیم اولیه قابلیت اعتماد را یادآوری می‌کنیم. سپس بعضی از مهمترین ترتیب‌های تصادفی مورد نیاز در این پایان‌نامه را معرفی و روابط بین آن‌ها را بیان خواهیم نمود. در ادامه به بیان ویژگی‌هایی از توزیع طول عمر و روابط میان آن‌ها می‌پردازیم. سپس برخی مفاهیم وابستگی را یادآوری و روابط بین آن‌ها را ذکر می‌کنیم. در انتها، به تعریف تابع مفصل می‌پردازیم و توزیع و ضریب کندال مفصل ارشمیدسی را که در فصل ۵ به آن نیازمندیم، محاسبه می‌نماییم.

۲.۱ آشنایی با توابع قابلیت اعتماد

توابع متفاوتی در قابلیت اعتماد مورد استفاده قرار می‌گیرند که در این بخش با برخی از آن ها مانند تابع بقا[۱]، مخاطره[۲]، مخاطره تجمعی[۳] و متوسط مخاطره [۴] آشنا می‌شویم.

پیش از معرفی توابع فوق، ابتدا تعریف تابع مطلقا پیوسته را که در ادامه مورد استفاده قرار می‌گیرد، بیان می‌کنیم.

تعریف ۱.۲.۱. *تابع $f(x)$ در بازه $[a,b]$ که $0 \leq a \leq b \leq +\infty$، مطلقا پیوسته می‌باشد هرگاه برای هر مقدار مثبت ϵ ، $\delta > 0$ ای وجود داشته باشد به طوریکه برای زیردنباله‌های مجزای $[x_k, y_k]$ ، $(k = 1, 2, \cdots, n)$، که در شرط*

$$\sum_{k=1}^{n} |y_k - x_k| < \delta,$$

صدق می‌کنند، داشته باشیم:

$$\sum_{k=1}^{n} |f(y_k) - f(x_k)| < \epsilon.$$

۱.۲.۱ تابع بقا

اولین تعریفی که در قابلیت اعتماد با آن مواجه می‌شویم، تابع قابلیت اعتماد است. متغیر تصادفی نامنفی پیوسته X را که دارای تابع توزیع F و تابع چگالی f باشد در نظر بگیرید. در این صورت می‌توان نوشت،

$$F(x) = P(X \leq x) = \begin{cases} \int_{0}^{x} f(t)dt & \text{اگر} \quad x \geq 0 \\ 0 & \text{اگر} \quad x < 0. \end{cases} \qquad (1.1)$$

[1]Survival function

[2]Hazard rate

[3]Cumulative hazard rate function

[4]Hazard rate average function

تعریف ۲.۲.۱. *برای متغیر تصادفی نامنفی پیوسته X تابع نامنفی $\bar{F}$ را که در آن،*

$$\bar{F}(x) = 1 - F(x) = \int_x^\infty f(t)dt, \qquad (۲.۱)$$

تابع قابلیت اعتماد یا تابع بقا گوییم. اگر تابع $F(.)$ مطلقا پیوسته باشد، آن‌گاه تابع چگالی متغیر تصادفی X وجود دارد و $f(x) = -\bar{F}'(x)$.

۲.۲.۱ تابع مخاطره

تابع مخاطره به دلیل تعبیر احتمالی که در تعریف آن نهفته است و در ادامه آن را بیان می‌کنیم، از اهمیت بالایی در قابلیت اعتماد برخوردار است. به عنوان مثال، استیفنسون[۱] (۱۹۳۰) در علم بیمه از این تابع تحت عنوان شدت میرایی نام برده است.
فاصله زمانی $(x, x+\Delta x]$ را در نظر می‌گیریم. احتمال خرابی قطعه در این فاصله به شرط این‌که در بازه $(0, x]$ خرابی اتفاق نیفتاده باشد، عبارت است از:

$$\begin{aligned} P[x < X < x+\Delta x | X > x] &= \frac{P[x < X < x+\Delta x]}{P[X > x]} \\ &= \frac{F(x+\Delta x) - F(x)}{\bar{F}(x)}. \end{aligned} \qquad (۳.۱)$$

با تقسیم رابطه (۳.۱) بر Δx در فاصله زمانی به طول $\Delta x > 0$ وقتی که $\Delta x \to 0$، تابع مخاطره حاصل می‌شود و آن را با $\lambda(x)$ نمایش می‌دهیم. به عبارت دیگر،

$$\begin{aligned} \lambda(x) &= \lim_{\Delta x \to 0} \frac{1}{\Delta x} \frac{F(x+\Delta x) - F(x)}{\bar{F}(x)} \\ &= \frac{f(x)}{\bar{F}(x)} = -\frac{d}{dx} \ln \bar{F}(x). \end{aligned} \qquad (۴.۱)$$

با انتگرال گیری از رابطه (۴.۱) تابع مخاطره تجمعی به صورت زیر حاصل می‌شود.

$$\Lambda(x) = \int_0^x \lambda(t)dt = -\ln \bar{F}(x). \qquad (۵.۱)$$

[1] Steffensen

متوسط مخاطره نیز به صورت زیر ارائه می‌گردد:

$$H(x) = \frac{1}{x}\int_0^x \lambda(t)dt.$$

همچنین از رابطه (۵.۱) داریم،

$$\bar{F}(x) = \exp\{-\Lambda(x)\}. \tag{۶.۱}$$

۳.۱ ترتیب‌های تصادفی

یکی از راه‌هایی که از گذشته برای مقایسه‌ی دو توزیع یا چگالی بکار می‌رفت، مقایسه‌ی شاخص‌هایی مانند میانگین، پراکندگی، چولگی، کشیدگی و ... بود. اما مان و ویتنی [۱] (۱۹۴۷)، نخستین بار دیدگاه دیگری را مطرح نمودند. آن‌ها برای یك شاخص خاص، ترتیب $X \leq Y$ را معرفی کردند که به صورت دقیق‌تر بیان می‌کند X در آن شاخص خاص کمتر از Y است. امروزه از این ایده با عنوان ترتیب تصادفی یاد می‌کنیم. از جمله معروف‌ترین ترتیب‌های تصادفی، می‌توان ترتیب تصادفی معمولی، ترتیب نرخ مخاطره، ترتیب نرخ مخاطره معکوس و ترتیب نسبت درستنمایی را نام برد، که در ادامه بر اساس شیکد و شانتیکومار [۲] (۲۰۰۷) به تعریف آن‌ها می‌پردازیم.

تعریف ۱.۳.۱. *فرض کنید X و Y دو متغیر تصادفی پیوسته به ترتیب با توابع چگالی f و g، توابع توزیع F و G و توابع بقای $\bar{F}$ و $\bar{G}$ باشند، آن‌گاه*

- *X را در ترتیب نسبت درستنمایی [۳] از Y کوچك‌تر ($X \leq_{lr} Y$) گویند هرگاه $\frac{f(x)}{g(x)}$ نسبت به $x \geq 0$ نزولی باشد.*
- *X را در ترتیب تصادفی معمولی [۴] از Y کوچك‌تر ($X \leq_{st} Y$) گویند، اگر برای هر x، رابطه $\bar{F}(x) \leq \bar{G}(x)$ برقرار باشد.*

---

[1] Mann and Whitney

[2] Shaked and Shanthikumar

[3] Likelihood ratio order

[4] Usual stochastic order