
بسم الله الرحمن الرحیم

پیام دبیر نهمین سمینار ملی آمار و احتمال فازی

سپاس خدای را که هر چه دارم از اوست؛

علم و دانش دو بالی هستند که انسان می‌تواند با آن‌ها تا بیکران پرواز کند. ارزش هر انسان با چگونگی به کارگیری آن‌ها معین می‌شود. علم و دانش، آگاهی و معرفت او را به جایی می‌رسانند که حتی فرشتگان هم بدان دست نمی‌یابند. هر چه دانایی انسان بیشتر شود، نقش رهبری او در زندگی پر رنگ‌تر می شود. در ارزش علم همین بس که خداوند متعال در سوره زمر آیه ۱۱ می فرماید:«قُل هَل یستَوی الذَّینَ یعلَمُونَ و الذین لا یعلَمُونَ ...؛ بگو آیا کسانی که می‌دانند با کسانی که نمی‌دانند برابر هستند؟...» باید گفت؛ پیدایش علم و دانش با خلقت انسان برابری می‌کند و همواره بشر در صدد آن بوده که بفهمد و درک نماید. علم و دانش در زندگی انسان دارای جایگاه ویژه‌ای است. نقش علم در زندگی انسان این است که راه سعادت، تکامل و راه ساختن را به انسان می‌آموزد. علم، انسان را توانا می‌کند که آینده را همان گونه که می‌خواهد بسازد. علم مانند ابزاری در اختیار خواست انسان قرار می‌گیرد و طبیعت را آن چنان که انسان بخواهد و فرمان دهد می‌سازد. با استعانت از خداوند متعال و با کمال مسرت، به اطلاع کلیه دانشگاهیان، پژوهشگران و علاقمندان می‌رساند که دانشکده علوم ریاضی دانشگاه مازندران، در راستای تحقق اهداف کنفرانس‌های تخصصی، به منظور اعتلای سطح دانش آمار و جهت ایجاد محیطی مناسب برای تبادل نظر آماردانان داخل و خارج، نهمین سمینار ملی آمار و احتمال فازی را در روزهای ۱۱ تا ۱۲ اردیبهشت ۹۸ برگزار گردید. بدین‌وسیله از عموم پژوهشگران، اعضای هیأت علمی و دانشجویان گرامی که با حضور فعال و پویای خود، ضمن شرکت در این همایش ملی با ارائه‌ی آخرین دستاوردهای علمی و پژوهشی خود بر غنا و کیفیت علمی این سمینار افزودند، سپاسگزاریم.

با احترام

سید باقر میراشرفی

فهرست سخنرانی‌های عمومی و کارگاه‌های تخصصی

سخنرانی‌های عمومی

کارگاه تخصصی

فهرست مقالات کامل فارسی (به ترتیب حروف الفبا بر اساس نام فامیل نویسنده اول)

فهرست مقالات کامل انگلیسی (به ترتیب حروف الفبا بر اساس نام فایل نویسنده اول)

فهرست چکیده مقالات فارسی (به ترتیب حروف الفبا بر اساس نام فامیل نویسنده اول)

فهرست چکیده مقالات انگلیسی (به ترتیب حروف الفبا بر اساس نام فامیل نویسنده اول)

محورهای سمینار

- احتمال فازی
- رگرسیون فازی
- داده‌کاوی فازی
- کنترل کیفیت فازی
- آمار حیاتی در محیط فازی
- قابلیت اعتماد در محیط فازی
- استنباط آماری در محیط فازی
- وجوه آماری و احتمالی سیستم‌های فازی
- آمار اقتصادی ـ اجتماعی در محیط فازی
- سایر موضوع‌های مرتبط با آمار و احتمال در محیط فازی

اعضای کمیته علمی (به ترتیب حروف الفبا)

- دکتر اکبر اصغرزاده، دانشگاه مازندران
- دکتر احمد پوردرویش، دانشگاه مازندران
- دکتر محمدرضا ربیعی، دانشگاه صنعتی شاهرود (نماینده انجمن آمار ایران)
- دکتر رضا زارعی، دانشگاه گیلان
- دکتر بهرام صادقپور گیلده، دانشگاه فردوسی مشهد
- دکتر سید محمود طاهری، دانشگاه تهران
- دکتر بهروز فتحی واجارگاه، دانشگاه گیلان (نماینده انجمن سیستم‌های فازی)
- دکتر ماشاءالله ماشین‌چی، دانشگاه شهید باهنر کرمان
- دکتر رضا علی محمدپور، دانشگاه علوم پزشکی مازندران
- دکتر سید باقر میراشرفی، دانشگاه مازندران (دبیر کمیته علمی)

اعضای کمیته اجرایی

- دکتر یحیی طالبی رستمی، رئیس دانشگاه مازندران (دبیر سمینار)
- دکتر محمد حسین فاطمی، معاون پژوهشی دانشگاه مازندران
- دکتر الله‌بخش یزدانی، رئیس دانشکده علوم ریاضی دانشگاه مازندران (دبیر کمیته اجرایی)
- دکتر سید باقر میراشرفی، دانشگاه مازندران (دبیر کمیته علمی)
- دکتر سید محمد تقی کامل میرمصطفائی، دانشگاه مازندران
- شهرام پورداد، مدیر امور اداری دانشکده علوم ریاضی دانشگاه مازندران
- سید فاضل باقری، دانشگاه مازندران
- علی یوسفی سوته، دانشگاه مازندران

کادر اجرایی سمینار

- نسیم اندخس
- زهره عظیمی (مسئول دبیرخانه سمینار)
- فائزه کجوری
- هانیه محمدی

حامیان سمینار ملی آمار و احتمال فازی

- دانشگاه مازندران
- انجمن سیستم‌های فازی ایران
- انجمن آمار ایران

9th National Seminar on
Fuzzy Statistics and Probability

سخنرانی عمومی

Monte Carlo computations for fuzzy and crisp systems

Behrouz Fathi-Vajargah[1]

Department of Statistics, Faculty of Mathematical Sciences, University of Guilan, Iran

Abstract: The main subject of this speech is employing the Monte Carlo method to solve the *real and complex fuzzy linear systems* with new techniques. Since, the purpose of this work is obtaining the solution with high accuracy, modifications are exerted on the *Monte Carlo* and *hybrid Monte Carlo* algorithms to find the inverse of matrix under *Strictly Diagonally Dominant* property and in general case so. Specific strategy is implemented on Monte Carlo method for a class of fuzzy linear systems with higher order. We also present a new discussion on the random fuzzy numbers, random number generators and scrambling of them for extra uniformity. We have experimentally indicated that the partitioned rand library function generates the random numbers with more uniformity than library rand function in employed software (such as MATLAB) and programing languages. We also have a survey in Monte Carlo integration problems.

[1]Speaker: fathi@guilan.ac.ir; behrouz.fathi@gmail.com

سخنرانی عمومی

مروری بر نمودارهای کنترل آماری فرایند در محیط فازی

بهرام صادقپور گیلده

گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده: نمودار کنترل آماری فرایند، ابزاری مهم برای شناسایی عوامل تغییر در فرایندها (به ویژه فرایندهای تولید) و انجام اقدامات اصلاحی روی آن‌ها می‌باشد. نمودارهای کنترل اولین بار توسط شوهارت در سال ۱۹۲۴ معرفی شدند، که به دو دسته نمودارهای کنترل برای متغیرهای وصفی و نمودارهای کنترل برای متغیرهای کمی (درحالات تک متغیره و چند متغیره) تقسیم بندی می‌شوند. پس از معرفی نظریه‌ی مجموعه‌ی فازی توسط زاده در سال ۱۹۶۵ و به تبع آن ارائه‌ی مفاهیم جدیدی از آمار در محیط فازی، ساخت نمودارهای کنترلی برای پایش فرایندهایی که دست‌خوش تغییراتی ناشی از دو نوع عدم قطعیت تصادفی و فازی (مانند درجه‌بندی کیفیت کالاهای تولید به صورت عالی، خوب، متوسط، ضعیف و بد) هستند، مورد توجه محققان آماری به ویژه آماردانان صنعتی قرار گرفت. در این سخنرانی با مروری بر اهم کارهای انجام شده در این زمینه، در خصوص ویژگی های مهم نمودارهای کنترل آماری فرآیند فازی صحبت خواهد شد.

واژه‌های کلیدی: نمودار کنترل آماری فرایند، اعداد فازی، غیر فازی سازی، فاصله‌ی بین اعداد فازی.

سخنرانی عمومی

احتمال و امکان: شباهت‌ها و تفاوت‌ها

سید محمود طاهری

دانشکده فنی، دانشگاه تهران

چکیده: نظریه‌ی احتمال و نظریه‌ی امکان، دو نظریه‌ی ریاضی برای بررسی و تحلیل نایقینی (عدم‌اطمینان / عدم‌قطعیت) هستند. هرچند این دو نظریه شباهت‌هایی دارند، ولی تفاوت‌های بنیادی با یکدیگر دارند. در این مقاله، شباهت‌ها و به ویژه تفاوت‌های این دو نظریه بررسی می‌شود. چند مثال مقایسه‌ای ارائه می‌شود تا تفاوت‌های احتمال و امکان به طور ملموس و عینی روشن شود. در پایان، یک پرسش مطرح می‌شود تا خوانندگان با بررسی آن، درباره‌ی تفاوت احتمال و امکان بیشتر تأمل کنند.

واژه‌های کلیدی: اندازه‌ی احتمال، اندازه‌ی امکان، اندازه‌ی لزوم.

کارگاه تخصصی

معرفی بسته fuzzyreg برای تحلیل برخی از روش‌های رگرسیون فازی در R

محمد رضا ربیعی[1]**، میترا خراباتی**

گروه آمار، دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود

چکیده: رگرسیون فازی، زمانی که روابط بین پارامترهای مدل، یا داده‌ها و یا هر دو مورد مبهم (فازی) هستند، جایگزین مناسبی برای رگرسیون آماری است. بنابراین رگرسیون فازی در مواردی که ساختار داده‌ها و یا روابط بین آن‌ها مانع از تجزیه و تحلیل کلاسیک آماری می‌شوند، قابل استفاده است. تا کنون به صورت منسجم تحلیل رگرسیون فازی در نرم افزارهای آماری موجود نیست و کاربران برای استفاده از این روش‌ها باید اقدام به کدنویسی نمایند. در نرم افزار R به علت این‌که افراد می‌توانند دستورات برنامه نویسی خود را به صورت یک بسته ارائه کنند تا دیگران به سهولت از آن استفاده کنند، کاربران زیادی را در سراسر جهان پیداکرده است. تعدادی بسته مرتبط با فازی در این نرم افزار ارائه گردیده است. یکی از این بسته‌ها که به تحلیل برخی از روش‌های رگرسیون فازی اشاره کرده است بسته fuzzyreg است. در این‌جا قصد داریم به معرفی روش‌های رگرسیون خطی فازی مطرح شده در این بسته بپردازیم.

واژه‌های کلیدی: رگرسیون فازی، بسته FuzzyNumber، بسته fuzzyreg.

[1] **نام ارائه دهنده‌ی کارگاه :** rabiei1354@yahoo.com

کشف مشترکین متخلف برق به کمک داده‌کاوی

سید علی خالقی[۱]، **سید باقر میراشرفی**[۲]

[۱] شرکت برق منطقه‌ای مازندران

[۲] گروه آمار، دانشگاه مازندران

چکیده: در سال‌های اخیر انتخاب روش‌های کارآمد برای کشف خسارت و تلفات از موضوعات مهم تحقیقاتی هستند. تلفات غیرفنی یکی از مشکلات اساسی صنعت برق بوده و شناسایی و کاهش آنها از وظایف مهم صنعت برق به شمار می‌آید. هدف اصلی این مقاله پیش‌بینی تلفات غیرفنی با دقت بالا و کشف مشترکین متخلف با استفاده از تکنیک ماشین بردار پشتیبان است. مطالعه‌ی موردی بر داده‌های شرکت توزیع برق مازندران صورت گرفته که از پروفایل بار مشترکین استفاده شده است

واژه‌های کلیدی: تلفات غیرفنی، پروفایل بار، ماشین بردار پشتیبان، پیش‌بینی.

---

۱ مقدمه

با پیشرفت تکنولوژی، امکان ذخیره داده‌ها در حجم بالا فراهم شده است. با افزایش داده‌ها نیاز به ابزارهایی برای تحلیل آنها و کشف دانش از آنها نیز احساس می‌شود. داده‌ها اساس هر سیستم اطلاعاتی هستند و باید مدیریت شوند که طی آن داده‌ها به اطلاعات مفید، دانش و مبنایی برای پشتیبانی تصمیم تبدیل می‌گردند. شرکت برق پایگاه داده عظیمی از اطلاعات مشترکین خود دارد. با بررسی و کاوش پایگاه داده، می‌توان به اطلاعات وسیعی از وضعیت مشترکین دست یافت. از طرفی با توجه به نظر کارشناسان شرکت برق، شناسایی تلفات غیرفنی در صنعت برق ایران بسیار پرهزینه است و نیاز به منابع انسانی و تجهیزات و وقت زیادی دارد. لذا مطالعه در زمینه شناسایی و کشف تلفات غیرفنی برق در ایران از اهمیت بالایی برخوردار است. دستکاری کنتور، نقص کنتور، اتصالات غیرقانونی، بی‌نظمی‌های صورتحساب و صورتحساب‌های پرداخت نشده منشا تلفات غیرفنی هستند. داده‌کاوی ابزاری سریع جهت شناسایی تلفات می‌باشد که باعث صرفه‌جویی در زمان، منابع و هزینه می‌گردد.

۲ ماشین بردار پشتیبان

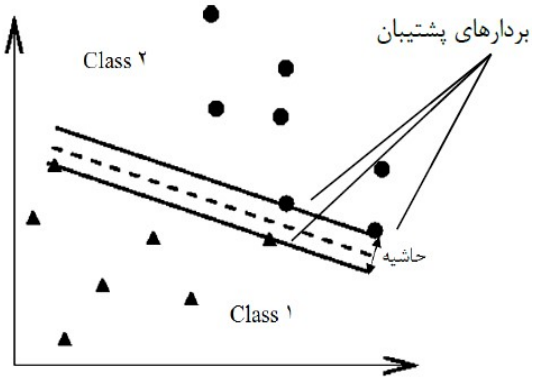
ماشین بردار پشتیبان توسط وپنیک[۱] در سال ۱۹۶۰ بر اساس تئوری یادگیری آماری معرفی شده است ماشین بردار پشتیبان یکی از تکنیک‌های دسته‌بندی است که به دنبال یافتن ابرصفحه‌ای[۲] است. این ابرصفحه به گونه‌ای یافت می‌شود که حاشیه دسته‌بندی را حداکثر نماید که این معادل حداکثر نمودن دقت دسته‌بندی خواهد بود. ابرصفحه‌ها در فضای سه بعدی به‌صورت صفحه و در فضای دو بعدی به

[۱] نام ارائه دهنده‌ی مقاله : Khaleghy.ali@gmail.com

[۱] Vapnik

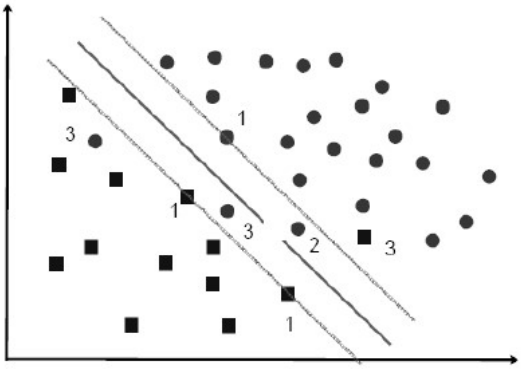
[۲] Hyper-plane

صورت خط می‌باشند. در حالت اصلی، مدل ماشین بردار پشتیبان برای حالت دو دسته‌ای توسعه داده شده است. اگر بتوان داده‌ها را با استفاده از یک ابرصفحه به طور کامل از هم تفکیک نمود در اینصورت به آنها تفکیک‌پذیر خطی[۳] گفته می‌شود (شکل ۱). به نقاطی که دارای کمترین فاصله از ابرصفحه مذکور هستند، بردارهای پشتیبان[۴] گفته می‌شود.



شکل ۱: تفکیک داده‌ها در حالت تفکیک‌پذیر خطی (وپنیک، ۱۹۹۸)

اگر داده‌ها را نتوان با یک ابرصفحه به طور کامل جداسازی نمود به آنها تفکیک‌ناپذیر خطی[۵] گویند (شکل ۲).



شکل ۲: تفکیک داده‌ها در حالت تفکیک‌ناپذیر خطی (وپنیک، ۱۹۹۸)

علیرغم کاربرد ماشین بردار پشتیبان در بسیاری از مسایل دسته‌بندی، دارای نقص‌هایی نیز می‌باشد از جمله مشکل دقت در داده‌های تفکیک ناپذیر خطی. از طرفی این امکان وجود دارد که توزیع برخی داده‌ها به صورت ابرکروی، بیضی و حتی هذلولی شکل باشند که در اینصورت دسته‌بندی آنها بسیار سخت خواهد شد. بنابراین روش جدیدی به نام ماشین بردار پشتیبان قطبی[۶] معرفی خواهد شد که می‌توان به همراه ماشین بردار پشتیبان خطی یا غیرخطی برای حل مسایل مختلف استفاده کرد. ماشین بردار پشتیبان قطبی از دستگاه مختصات قطبی[۷] برای نگاشت داده‌ها استفاده می‌کند. برای نگاشت داده‌ها باید مختصات کلیه داده‌ها با زاویه[۸] و فاصله اقلیدسی[۹] آنها

[۳] Linearly Separable
[۴] Support Vectors
[۵] Non-linearly separable
[۶] Polar support vector machine
[۷] Polar coordinate system
[۸] Angle
[۹] Euclidean distance

جایگزین شود. تابع هدف روش پیشنهادی بصورت زیر بدست می‌آید:

$$min\frac{۱}{۲}\|\omega\|^{۲} + C\sum_{i=۱}^{n}\xi_i$$

$$y_i\left\|\omega^T\theta + b\right\|^{۲} \geq ۱ - \xi_i$$

که $\xi_i \geq ۰$

و θ از رابطه زیر بدست میآید:

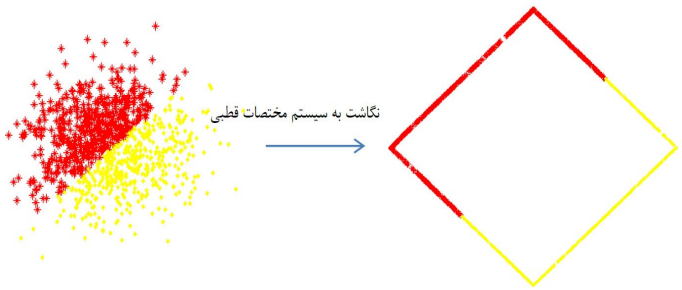
$$x = rcos\theta$$

$$cos\theta = x/r$$

$$\theta = cos^{-۱}(x/r)$$

که اندازه r برابر با فاصله اقلیدسی است.

در شکل زیر میزان خوش‌ساخته شدن داده‌ها بهنگام نگاشت آنها در سیستم مختصات قطبی نشان داده شده است.



شکل ۳: نگاشت از ماشین بردار پشتیبان خطی به سیستم قطبی

۳ پیش پردازش داده ها و آماده سازی آنها

روش ارائه شده در بخش قبل، روی داده‌های واقعی پیاده‌سازی شده و نتایج خروجی بررسی و با هم مقایسه خواهند شد. همانطور که قبلا اشاره شد صورتحساب‌های پرداخت نشده یا بینظمی در صورتحساب منشا تلفات غیرفنی میباشد. اطلاعاتی که از هر مشترک در نظر گرفته شده است مقدار مبلغ کارکرد، تاریخ صدور و وصول قبض، مقدار مبلغ وصولی، میزان مصرف برق در ساعات مختلف از شبانه روز و شماره اشتراک هر یک از مشترکین می‌باشد. برای شناسایی و کشف متخلفین از داده‌های مصرف مشترکین (پروفایل بار) کمک گرفته شده است. از آنجایی که ما برای پیشبینی و آموزش و تست مدل خود نیاز به هر دو گروه از مشترکین، چه سالم و چه متخلف، داریم لذا دو کلاس به نام صفر و یک تعریف کرده‌ایم که مشترکین سالم جز کلاس صفر و مشترکین متخلف جز کلاس یک هستند. در این پژوهش پروفایل بار ۱۰۰۰ مشترک بررسی شده است که ۵۴۵ مورد آن مشترک سالم و ۴۵۵ مورد آن مشترک متخلف می‌باشد. برای پیش‌بینی از اطلاعات سه سال مشترکین شرکت توزیع نیروی برق مازندران (اردیبهشت سال ۹۳ تا اردیبهشت سال ۹۵) استفاده شده است.

۴ پیش بینی با استفاده از ماشین بردار پشتیبان

برای یادگیری ماشین بردار پشتیبان ۷۰ درصد داده‌ها جهت آموزش و ۳۰ درصد داده‌ها جهت تست به ماشین معرفی گردیده است. داده‌ها یکبار با ماشین بردار پشتیبان خطی و سپس با ماشین بردار پشتیبان قطبی اجرا شدند. پس از نتایج به دست آمده از ماشین بردار پشتیبان، دقت پیشبینی هر یک اعتبارسنجی شده است. درصد دقت تست نتایج و زمان آموزش در جدول ۱ ارائه شده است.

جدول ۱: درصد دقت تست نتایج ماشین بردار پشتیبان قطبی

نوع ماشین بردار پشتیبان	دقت تست (درصد)	زمان آموزش (ثانیه)
خطی	۰/۶۵	۰/۷۸۹۱
خطی قطبی	۰/۷۸	۲/۲۸۸۳

نتایج بدست آمده از جدول بالا نشان می‌دهد که ماشین بردار پشتیبان قطبی نسبت به خطی از دقت بالاتری جهت پیشبینی برخوردار است و از طرفی دارای زمان آموزش بالاتری است. این افزایش دقت نشان دهنده این است که روش پیشنهادی توانسته با اجرا بر روی داده‌ها نتیجه خوبی را ارائه دهد. چهار ترکیب ممکن برای تصمیم و خطا وجود دارد: (دو دسته C۱ و C۲ را درنظر بگیرید) (هان و کمبر، ۲۰۰۶)

جدول ۲: class Predicted

	C۱	C۲
C۱	TP	FN
C۲	FP	TN

حساسیت[۱۰] ، شفافیت[۱۱] و دقت[۱۲] بصورت زیر محاسبه می‌گردند:

$$sensitivity = \frac{TP}{TP + FN}$$

$$specificity = \frac{TN}{TN + FP}$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

که FN FP، TN، TP، به ترتیب مثبت درست، منفی درست، مثبت نادرست، منفی نادرست هستند به صورت جدول (۲) تعریف شده‌اند.

منحنی ROC[۱۳] یا مشخصه عملکرد سیستم، ابزاری مفید برای بررسی عملکرد روش‌های بکار برده شده است. منحنی مشخصه عملکرد سیستم یک روش کمی ارزیابی است که در سال ۱۹۵۰ توسعه پیدا کرد. اولین بار برای تشخیص سیگنال رادیویی دارای اغتشاش [۱۴] بکار رفت. از این منحنی در پزشکی، رادیولوژی، بیومتریک و امروزه در پژوهش‌های یادگیری ماشین و داده‌کاوی استفاده می‌شود.

[۱۰] Sensitivity
[۱۱] Specificity
[۱۲] Accuracy
[۱۳] Receiver Operating Characteristic
[۱۴] Noise

جدول ۳: نمادهای استفاده شده و مفاهیم آنها

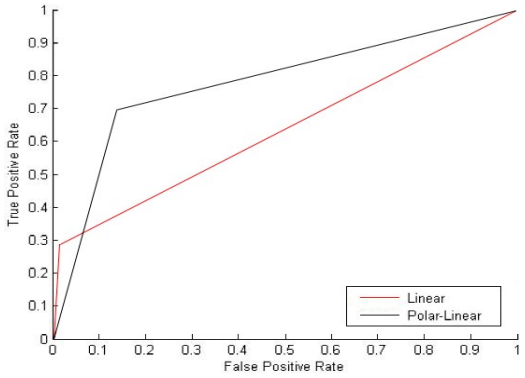
مثبت نادرست	مشاهداتی از دسته C۲ که به نادرستی در دسته C۱ قرار گرفته‌اند.
منفی نادرست	مشاهداتی از دسته C۱ که به نادرستی در دسته C۲ قرار گرفته‌اند.
منفی درست	مشاهداتی از دسته C۲ که بدرستی تشخیص داده شده‌اند.
مثبت درست	مشاهداتی از دسته C۱ که بدرستی تشخیص داده شده‌اند.

در منحنی ROC محور y را[۱۵] TPR یا نرخ مثبت درست و محور x را[۱۶] FPR یا نرخ مثبت نادرست می‌نامند که از روابط زیر بدست می‌آیند:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

در شکل (۴) خروجی حاصل از رسم منحنی نشان داده شده است. هر چه سطح زیر منحنی ROC بیشتر باشد، آن روش از کارایی بهتری برخوردار است. (متز[۱۷] ۱۹۷۸).



شکل ۴: ROC برای مقایسه روش‌های دسته‌بندی

جمع‌بندی

در این مقاله به مطالعه موردی و تشریح کامل فازهای روش پیشنهادی پرداخته شد. از آنجایی که عمل شناسایی و کشف تلفات غیرفنی از حساسیت بالایی برخوردار است پس هر چه دقت پیش‌بینی ماشین بردار پشتیبان بالاتر باشد، نتایج بهتری بدست خواهد آمد. روش پیشنهادی بیان شده بر روی داده‌های واقعی اجرا شد و نتایج حاصل از اجرا بررسی شده و از نظر دقت تست و زمان آموزش مقایسه شدند. بر اساس این تحلیل‌ها، روش پیشنهادی می‌تواند بعنوان یک روش کارآمد مورد استفاده قرار گیرد.

[۱۵] True Positive Rate
[۱۶] False Positive Rate
[۱۷] Metz

مراجع

Cabral J.E., and Gontijo E.M. (2004), Fraud detection in electrical energy consumers using rough sets, *In Proc. of the 2004 IEEE International Conference on Systems, Man and Cybernetics* 4, 3625-3629.

Depuru S.S.S.R., Wang L., and Devabhaktuni V. (2011), Electricity theft: Overview, issues, prevention and a smart meter based approach to control theft, *Energy Policy* 39, 1007-1015.

Depuru S.S.S.R., Wang L., Devabhaktuni V., and Green R.C. (2013), High performance computing for detection of electricity theft, *International Journal of Electrical Power & Energy Systems* 47, 21-30.

Maali Y. and Al-Jumaily A. (2013), Self-advising support vector machine, *Knowledge Based Systems* 52, 214-222.

Yap K.S., Tiong S.K., Nagi J., Koh J. SP. and Nagi F. March. (2012), Comparison of supervised learning techniques for non-technical loss detection in power utility, *International Review on Computers and Software* 7, 126-135.

برآورد پارامترهای رگرسیون لجستیک فازی به روش کمترین مربعات خطا

میترا خراباتی[1]، **محمدرضا ربیعی**

دانشگاه صنعتی شاهرود

چکیده: وقتی متغیر پاسخ گسسته دودویی باشد از رگرسیون لجستیک برای مدل سازی داده‌ها استفاده می‌کنیم. اما در بعضی مواقع ممکن است مشاهدات نادقیق باشند، در این حالت رگرسیون لجستیک دیگر کارآیی ندارد و باید به منطق فازی روی آورد. در این مقاله رگرسیون لجستیک معرفی شده است و روشی برای برآورد پارامترهای مدل ارائه شده است.

واژه‌های کلیدی: متغیرهای دودویی، رگرسیون لجستیک، رگرسیون لجستیک فازی، منطق فازی.

---

۱ مقدمه

در روش مدل‌سازی، زمانی‌که مقادیر متغیر پاسخ، گسسته است دیگر نمی‌توان از روش‌های رگرسیون خطی استفاده کرد. روش رگرسیون لجستیک توانایی مدل‌سازی این گونه مسائل را برای متغیرهای تبیینی پیوسته و گسسته داراست. مدل رگرسیون لجستیکی توسط کاکس (۱۹۷۰) برای توصیف وابستگی یک متغیر دودویی به مجموعه‌ای از متغیرهای پیوسته معرفی شد. در کلاس مدل‌های غیرخطی دسته‌ای از مدل‌ها وجود دارند که ذاتاً خطی هستند؛ یعنی با تبدیلی بر روی متغیرها، رابطه بین آن‌ها خطی می‌شود. مدل رگرسیون لجستیک یکی از این مدل‌هاست که با تبدیل لجیت، خطی می‌شود (اندرسن و مشکانی، ۱۳۸۳).

رگرسیون لجستیک یکی از تکنیک‌های کاربردی برای تحلیل داده‌های رده‌بندی شده است که از آن برای بیان رابطه بین متغیرهای تبیینی با یک متغیر پاسخ از نوع دو سطحی استفاده می‌شود. به‌عنوان نمونه، اگر نتیجه آزمایشی را به‌صورت موفقیت و شکست معرفی کنیم، در این حالت متغیر پاسخ دیگر پیوسته نبوده، و به صورت رده‌بندی شده خواهد بود. در این حالت برای رده‌بندی مشاهدات جدید، می‌توان از مدل رگرسیون لجستیک استفاده کرد که عضوی از رده مدل‌های خطی تعمیم یافته محسوب می‌شوند. در حالت دودویی در متغیرهای وابسته، وضعیت مطلوب را با عدد یک و وضعیت مقابل را با عدد صفر کدگذاری می‌کنیم. به‌طور مثال می‌توان برای مدل‌سازی وضعیت بیماری یک فرد (سالم/بیمار) از مدل لجستیک استفاده کرد (پوراحمد و همکاران، ۱۳۹۰). همان‌طور که گفته شد در رگرسیون لجستیک از تابع لجیت (تبدیل لجیت) که به صورت زیر است:

$$E(Y_i) = P(Y_i = ۱) = \pi_i, \qquad \pi_i(x) = \frac{exp(\beta_۰ + \beta_۱ x_{i۱} + \cdots + \beta_{(p-۱)} x_{i(p-۱)}}{۱ + exp(\beta_۰ + \beta_۱ x + \cdots + \beta_n x)},$$

داریم:

$$Ln(\frac{\pi_i}{۱ - \pi_i}) = \beta_۰ + \beta_۱ x_{i۱} + \cdots + \beta_{(p-۱)} x_{i(p-۱)}, \quad i = ۱, ۲, \cdots, n, \qquad (۱.۱)$$

[1] نام ارائه دهنده‌ی مقاله : kharabatimitra@gmail.com

استفاده می‌کنیم (نتر و همکاران، ۱۹۹۶). هدف به‌دست آوردن $E(Y=۱|X)$ است.

$$E(Y=۱|X=x)=\pi_i(x)=\frac{\beta_۰+\beta_۱x_{i۱}+\cdots+\beta_{(p-۱)}x_{i(p-۱)}}{۱+\beta_۰+\beta_۱x_{i۱}+\cdots+\beta_{(p-۱)}x_{i(p-۱)}.} \quad (۲.۱)$$

برای برآورد پارامترهای رگرسیون لجستیک، باید از روش ماکزیمم درستنمایی استفاده کرد. این مدل‌ها در حجم نمونه کم و روابط بین متغیرهای نادقیق انعطاف پذیری بالا و کارایی بهتری دارند.

۲ رگرسیون لجستیک فازی

رگرسیون فازی همانند رگرسیون کلاسیک، رابطه بین متغیر پاسخ و یک یا چند متغیر تبیینی را مدل‌سازی می‌کند. متغیر مورد مطالعه در رگرسیون کلاسیک، متغیرهای دقیق هستند و مشاهدات مربوط به متغیرها نیز مشاهداتی دقیق می‌باشند. در رگرسیون خطی رابطه خطی بین X_iها (متغیرهای تبیینی) و Y (متغیر پاسخ) وجود دارد، مدل کلی به‌صورت زیر نشان داده می‌شود.

$$Y_i=\beta_۰+\beta_۱x_{i۱}+\cdots+\beta_{(p-۱)}x_{i(p-۱)}+\epsilon_i \qquad i=۱,\cdots,n. \quad (۱.۲)$$

که Y_i متغیر پاسخ برای i امین عنصر و x_{ij} مقدار j امین متغیر تبیینی در i امین عنصر از نمونه است. β_jها ضرایب مدل هستند که بر پایه روش‌های آماری مانند کمترین مربعات خطا برآورد می‌شوند. همچنین ϵ جمله خطا می‌باشد که متغیری تصادفی و غیر قابل مشاهده است. باید توجه داشت که منظور از رابطه خطی، رابطه بین X_iها و Y نیست، بلکه در واقع پارامترهای رگرسیونی به‌طور خطی وارد مدل می‌شوند (طاهری و ماشین‌چی، ۱۳۹۲).

مدل رگرسیون لجستیک، که در این مقاله مورد بحث قرار گرفته است، مدلی است که در آن، مقادیر متغیر پاسخ مبهم بوده و به‌جای اعداد صفر و یک، مجموعه‌های فازی از بازه‌های واحد هستند. در این مدل متغیرهای تبیینی، دقیق فرض می‌شوند. شرایطی را در نظر بگیرید که وضعیت ابتلا به بیماری افراد مبهم است. به‌عبارت دیگر، در تشخیص بیماری یا حالت مورد نظر ابهام وجود دارد. در مدل‌های آماری این نمونه‌های نادقیق از نمونه اصلی حذف می‌شوند. در مواردی که با چنین نمونه‌هایی سروکار داریم برای محاسبه احتمال $P(Y_i=۱)=\pi_i$ ، از مدل لجستیک فازی استفاده می‌کنیم، که در این مدل فازی، امکان به‌جای احتمال مدل‌بندی می‌شود.در حقیقت ابتلا به بیماری را براساس مجموعه‌ای از عوامل مدل سازی می‌شود. این امکان به‌صورت ترم‌های زبانی (مانند خیلی کم، کم، متوسط، زیاد، خیلی زیاد) بیان می‌شوند. هر ترم زبانی یک مجموعه فازی با تابع عضویت بر بازه واحد است که توسط خبره تعریف می‌شود. هر ترم زبانی با $\tilde{\mu_i}$ نشان داده می‌شود، آنگاه براساس اصل گسترش، تبدیل یافته هر ترم زبانی را به‌عنوان متغیر پاسخ در مدل نظر گرفته می‌شود. نسبت $\frac{\tilde{\mu_i}}{۱-\tilde{\mu_i}}$ ، بخت امکانی نامیده می‌شود. لجیت نسبت بخت امکانی یعنی، $\tilde{w_i}=Ln(\frac{\tilde{\mu_i}}{۱-\tilde{\mu_i}})$ به‌عنوان متغیر پاسخ در نظر گرفته می‌شود (سلمانی و همکاران، ۱۳۹۵).

برای مدل بندی یک پاسخ فازی دودویی، بردار مشاهدات به‌صورت $(\tilde{x_{i۰}},\tilde{x_{i۱}},\cdots,\tilde{x_{ip-۱}},\tilde{\mu_i})$ در نظر گرفته می‌شود، که $\tilde{\mu_i}$ امکان وقوع رویداد در فرد iام و $\tilde{x_{ij}}$ مقدار متغیر فازی jام برای فرد iام است. مقدار برآورد شده لجیت به‌صورت زیر است

$$\tilde{\hat{W}}_i=Ln(\frac{\tilde{\mu_i}}{۱-\tilde{\hat{\mu}}_i})=\tilde{b_۰}+\tilde{b_۱}x_{i۱}+\cdots+\tilde{b_{p-۱}}x_{i(p-۱)},i=۱,۲,\cdots,n. \quad (۲.۲)$$

در این مدل $\tilde{b_j}$ پارامترهای برآورد شده مدل با مقدار دقیق است. با استفاده از اصل گسترش می‌توان بخت امکانی وقوع رویداد و همچنین امکان وقوع رویداد را برای هر فرد به‌دست آورد.

۳ برآورد پارامترها

برای برآورد پارامترهای مدل از روش کمترین مربعات استفاده می‌کنیم. در این روش ملاک برآورد پارامترها، مینیمم سازی مجموع توان دوم فاصله بین توابع عضویت‌های ترم‌های تعریف شده مشاهده شده و توابع عضویت‌های ترم‌های تعریف شده پیش بینی شده توسط مدل است، یعنی

$$SSE = \sum_{i=1}^{n} d^2(\tilde{W}_i, \tilde{w}_i).$$

$\tilde{w}_i$ مقدار تبدیل یافته مشاهده ترم‌های زبانی مربوط به هر ویژگی و $\tilde{W}_i$ مقدار برآورد شده متنظر با $\tilde{w}_i$ است که در (۲.۲) تعریف شده است (پوراحمد و همکاران، ۱۳۹۰).

ایده اساسی در اینجا مدل‌سازی مراکز متغیر پاسخ فازی با استفاده از یک مدل رگرسیون خطی است. این ایده را می‌توان تحت شرایط زیر بیان کرد. ابتدا مدل مراکز مشاهده شده و مرزهای پایین و بالای متغیرهای پاسخ را با استفاده از مجموعه‌های مربوط به باقی‌مانده‌ها و مقادیر دیگر تعریف می‌کنیم (کپی و همکاران، ۲۰۰۶).

$$\begin{aligned} m &= \mu + \underline{\epsilon} \\ m - l &= (\mu - \underline{\delta}_L) + \underline{\epsilon}_L \\ m - u &= (\mu - \underline{\delta}_U) + \underline{\epsilon}_U. \end{aligned}$$

در فرمول‌های بالا $\underline{\epsilon}$، $\underline{\epsilon}_L$ و $\underline{\epsilon}_U$ بردار خطاها و μ، $\underline{\delta}_L$ و $\underline{\delta}_U$ بردار مقادیر مراکز و پهناهای متغیر پاسخ است. بنابراین بردارهای معرفی شده براساس مدل رگرسیون به‌صورت زیر بررسی می‌شوند.

$$\begin{aligned} \mu &= X\underline{\beta} \\ \underline{\delta}_L &= \eta_L \mu + \xi_L \mathbf{1} \\ \underline{\delta}_U &= \eta_U \mu + \xi_U \mathbf{1}. \end{aligned} \tag{۱.۳}$$

ماتریس X، ماتریس داده‌ها مشاهده شده x_iها است، پارامترهای $\underline{\beta}$ و $(\eta_L, \eta_U, \xi_L, \xi_U)$ به‌ترتیب ضرایب خطی بین مراکز و توابع مربوط به X_j و رابطه خطی بین پهنای مشاهده شده و مراکز برآورد شده هستند و $\mathbf{1}$ نشان دهنده برداری $(n \times 1)$ با مقدار تکراری یک است.

روشی که برای برآورد پارمترهای فوق استفاده می‌شود مبتنی بر الگوریتم استادارد است که شامل مینمم کردن فاصله بین مقادیر مشاهده شده متغیر

$\tilde{w}_i$ و مقادیر برآورده شده $\tilde{W}_i$ که به‌صورت $\tilde{W}_i = (\mu_i, \delta_{L_i}, \delta_{U_i}), i = 1, \cdots, n$ تعریف می‌شود، مقادیر μ_i، δ_{L_i} و δ_{U_i} در مدل (۱.۳) تعریف شده است.

حال با توجه به معیار کمترین مربعات، پارامترهای مربوط به (۱.۳) باید با به حداقل رساندن فاصله مربعات بین مقادیر مشاهده

شده متغیر پاسخ $\tilde{w}_i$ و مقادیر $\tilde{W}_i$ برآورد شوند. برای این منظور، فاصله اقلیدسی زیر معرفی می‌شود.

$$\begin{aligned} d^{۲}(\tilde{w}_i, \tilde{W}_i) &= d^{۲}\left((m,l,u)_{LR}, (\mu, \underline{\delta}_L, \underline{\delta}_U)_{LR}\right) \\ &= \Delta^{۲}_{LR} = \|m-\mu\|^{۲} + \|(m-\lambda l) - (\mu - \lambda\underline{\delta}_L)\|^{۲} + \|(m+\rho u) - (\mu + \rho\underline{\delta}_U)\|^{۲} \\ &= ۳(m-\mu)'(m-\mu) - ۲\lambda(m-\mu)'(l-\underline{\delta}_L) + \lambda^{۲}(l-\underline{\delta}_L)'(l-\underline{\delta}_L) \\ &\quad + ۲\rho(m-\mu)'(u-\underline{\delta}_U) + \rho^{۲}(u-\underline{\delta}_U)'(u-\underline{\delta}_U) \end{aligned} \tag{۲.۳}$$

در این فاصله توابع وزن λ و ρ با توجه به تابع عضویت LR به‌صورت $\lambda = \int_{۰}^{۱} L^{-۱}(\omega)d\omega$ و $\rho = \int_{۰}^{۱} U^{-۱}(\omega)d\omega$ تعریف می‌شوند، مقادیر مربوط به این توابع وزن در تنظیم مناسب پهنای چپ و راست در هنگام محسابه مرزهای پایین و بالا اعداد فازی $\tilde{w}_i$ و $\tilde{W}_i$ کمک می‌کند (کپی و همکاران، ۲۰۰۶).

در فاصله (۲.۳) می‌توانیم تابع کمترین مربعات را براساس پارامترهای مدل (۱.۳) $\underline{\beta}, \eta_L, \eta_U, \xi_L, \xi_U$ جایگذاری کنیم :

$$\begin{aligned} &min_{\underline{\beta},\eta_L,\eta_U,\xi_L,\xi_U}\, \Delta^{۲}_{LR}(\underline{\beta}, \eta_L, \eta_U, \xi_L, \xi_U) \\ &= ۳(m-X\underline{\beta})'(m-X\underline{\beta}) - ۲\lambda(m-X\underline{\beta})'(l-X\underline{\beta}\eta_L - ۱\xi_L) + \lambda^{۲}(l-X\underline{\beta}\eta_L - ۱\xi_L)'(l-X\underline{\beta}\eta_L - ۱\xi_L) \\ &\quad + ۲\rho(m-X\underline{\beta})'(u-X\underline{\beta}\eta_U - ۱\xi_U) + \rho^{۲}(u-X\underline{\beta}\eta_U - ۱\xi_U)'(u-X\underline{\beta}\eta_U - ۱\xi_U) \\ &= ۳(m'm - ۲m'X\underline{\beta} + \underline{\beta}'X'X\underline{\beta}) - ۲\lambda(m'l - m'X\underline{\beta}\eta_L - m'۱\xi_L + \underline{\beta}'X'l + \underline{\beta}'X'X\underline{\beta}\eta_L + \underline{\beta}'X'۱\xi_L) \\ &\quad + \lambda^{۲}(l'l - ۲l'X\underline{\beta}\eta_L - ۲l'۱\xi_L + \underline{\beta}'X'X\underline{\beta}\eta_L^{۲} + ۲\underline{\beta}'X'۱\eta_L\xi_L + n\xi_L^{۲}) \\ &\quad + ۲\rho(m'u - m'X\underline{\beta}\eta_U - m'۱\xi_U + \underline{\beta}'X'U + \underline{\beta}'X'X\underline{\beta}\eta_U + \underline{\beta}'X'۱\xi_U) \\ &\quad + \rho^{۲}(u'u - ۲u'X\underline{\beta}\eta_U - ۲u'۱\xi_U + \underline{\beta}'X'X\underline{\beta}\eta_U^{۲} + ۲\underline{\beta}'X'۱\eta_U\xi_U + n\xi_U^{۲}). \end{aligned} \tag{۳.۳}$$

از تابع فوق نسبت به پارامترهای مجهول مشتق گرفته و مساوی صفر قرار دادیم و برآوردهای پارامترها به‌صورت زیر حاصل شد.

$$\begin{aligned} \eta_L &= \lambda^{-۱}(\underline{\beta}'X'X\underline{\beta})^{-۱}\left[\lambda(\underline{\beta}'Xl - \underline{\beta}'X۱\xi_L) - (\underline{\beta}'Xm - \underline{\beta}'X'X\underline{\beta})\right], \\ \eta_U &= \rho^{-۱}(\underline{\beta}'X'X\underline{\beta})^{-۱}\left[\rho(\underline{\beta}'Xu - \underline{\beta}'X۱\xi_U) - (\underline{\beta}'Xm - \underline{\beta}'X'X\underline{\beta})\right], \\ \xi_L &= (n\lambda)^{-۱}\left[\lambda۱'(l - X\underline{\beta}\eta_L) - ۱'(m - X\underline{\beta})\right], \\ \xi_U &= (n\rho)^{-۱}\left[\rho۱'(u - X\underline{\beta}\eta_U) - ۱'(m - X\underline{\beta})\right], \\ \underline{\beta} &= [۳ - \lambda\eta_L(۲ - \lambda\eta_L) + \rho\eta_U(۲ + \rho\eta_U)]^{-۱}(X'X)^{-۱}X' \\ &\quad \times \left[۳m - \lambda(m\eta_L + l + ۱\xi_L) + \lambda^{۲}(l\eta_L - ۱\eta_L\xi_L) + \rho(m\eta_U + u + ۱\xi_U) + \rho^{۲}(u\eta_U - ۱\eta_U\xi_U)\right]. \end{aligned}$$

باید به این نکته توجه داشت که برای به‌دست آوردن برآوردهای فوق، پیشنهاد می‌کنیم از یک الگوریتم تکراری با استفاده از چندین نقطه شروع مختلف برای ارزیابی ثبات راه حل، استفاده کنید.

۴ مثال کاربردی

داده‌های عددی این مقاله، از مقاله پوراحمد و همکاران (۱۳۹۰) گرفته شده است، داده‌ها متشکل از ۱۵ نفر زن مشکوک به بیماری لوپوس در بازه سنی ۴۰ − ۱۸ سال از درمانگاه‌های شیراز است. ۵ عامل ابتلا به بیماری لوپوس، از قبیل مواجه شدن با نور خورشید، سابقه خانوادگی، تست‌های آزمایشگاهی ANA، Anti DNA و ESR در اینجا مطرح شدند (پوراحمد و همکاران، ۱۳۹۰).

با توجه به ترم‌های زبانی ((خیلی کم، کم، متوسط، زیاد، خیلی زیاد)) وضعیت عمومی هر یک از ۱۵ فرد توسط پزشک بیان شده است. در همین راستا توابع عضویت هر وضعیت به‌صورت زیر تعریف شدند.

$$VeryLow = \begin{cases} ۱ - \frac{۰/۰۲ - x}{۰/۰۱} & ۰/۰۱ \le x \le ۰/۰۲ \\ ۱ - \frac{x - ۰/۰۲}{۰/۱۸} & ۰/۰۲ \le x \le ۰/۰۱۸ \end{cases} \quad Low = \begin{cases} ۱ - \frac{۰/۲۵ - x}{۰/۱۵} & ۰/۱ \le x \le ۰/۲۵ \\ ۱ - \frac{x - ۰/۰۲}{۰/۱۵} & ۰/۲۵ \le x \le ۰/۴ \end{cases}$$

$$Medium = \begin{cases} ۱ - \frac{۰/۰۵ - x}{۰/۱۵} & ۰/۳۵ \le x \le ۰/۵ \\ ۱ - \frac{x - ۰/۵}{۰/۱۵} & ۰/۵ \le x \le ۰/۶۵ \end{cases} \quad High = \begin{cases} ۱ - \frac{۰/۷۵ - x}{۰/۱۵} & ۰/۶ \le x \le ۰/۷۵ \\ ۱ - \frac{x - ۰/۷۵}{۰/۱۵} & ۰/۷۵ \le x \le ۰/۹ \end{cases}$$

$$VeryHigh = \begin{cases} ۱ - \frac{۰/۹۸ - x}{۰/۱۸} & ۰/۸ \le x \le ۰/۹۸ \\ ۱ - \frac{x - ۰/۹۸}{۰/۰۱} & ۰/۹۸ \le x \le ۰/۹۹ \end{cases}$$

مدل را با استفاده از روش کمترین مربعات که در بخش قبل توضیح دادیم، به داده‌ها برازش می‌دهیم و در نهایت مدل زیر به‌دست می‌آید.

$$\begin{aligned} \hat{\tilde{W}}_i = {} & (-۴/۱۱۱۷۶۴, ۱۲/۲۶۹۸۷۷, ۱۳/۹۶۴۷۷۱)_T + (-۰/۶۵۷۰۷۴, ۷/۵۳۶۵۱۳, ۸/۱۶۷۴۸۹)_T ESRtese \\ & + ۰/۰۰۹۷۳۸ AntiDNAtese \\ & + ۰/۰۰۹۷۳۸ ANAtest + (-۰/۱۴۰۷۲۴, ۷/۷۳۸۳۸۰, ۷/۹۲۳۰۸۷)_T SunExposure \\ & + (۰/۴۱۰۳۴۷, ۳/۳۴۱۹۴۶, ۳/۶۳۹۲۱۵)_T FamilyHistory \end{aligned} \tag{۱.۴}$$

ضرایب مدل اعداد فازی مثلثی نامتقارن هستند. با استفاده از مدل بهینه برازش شده (۱.۴) و با جایگذاری مقادیر مشاهده شده برای هر فرد می‌توان لگاریتم بخت امکانی ابتلا به بیماری در هر فرد را به‌دست آورد. میانگین کمترین مربعات خطای این روش محاسبه شد.

$$MSE = \frac{SSE}{۱۵} = ۰/۱۵۱۹$$

بحث و نتیجه‌گیری

ما در این مقاله، به پیروی از پوراحمد و همکاران (۱۳۹۰)، مدلی مبتنی بر نظریه مجموعه‌های فازی برای مدل‌سازی داده‌های دودویی مبهم ارائه کردیم و پارامترهای مدل لجستیک را به روش کمترین مربعات ارائه شده، برآورد کردیم. مثال کاربردی را به روش ارائه شده مدل‌سازی کردیم. در پایان میانگین را محاسبه کردیم و مقدار نسبتا کم این میانگین نشان داد که روش برآورد پارامترها مناسب است. تمامی عملیات محاسباتی با استفاده از نرم افزار R انجام شده است.

مراجع

اندرسن، ا. و مشکانی، ع. (۱۳۸۳). *تحلیل داده‌های رسته‌ای*، دانشگاه فردوسی مشهد.

پوراحمد، س.، آیت‌اللهی، س.م.ت.، طاهری، س.م. و حبیب آگهی، ز. (۱۳۹۰). مدل‌سازی موقعیت‌های مبهم تشخیص در پزشکی با استفاده از رگرسیون لجستیک به روش کمترین مربعات فازی، *سری سیستم‌های فازی و محاسبات نرم*، **۲**، ۴۳-۱۹.

سلمانی، ف.، طاهری، س.م.، ابدی، ع. و علوی مجد، ح. (۱۳۹۵). انتخاب مدل رگرسیون لجستیک فازی.

طاهری، س.م. و ماشین‌چی، م. (۱۳۹۲). *مقدمه‌ای بر احتمال و آمار فازی*، چاپ دوم، انتشارات دانگاه باهنر کرمان، کرمان.

Coppi R., DUrso P., Giordani P. and Santoro A. (2006), Least squares estimation of a linear regression model with LR fuzzy response, *Computational Statistics & Data Analysis* 51, 267-286.

Diamond P. (1988), Fuzzy least squares, *Information Sciences* 46, 14-157.

Neter J., Kutner M.H., Nachtsheim C.J. and Wasserman W. (1996), *Applied Linear Statistical Models*, Chicago, Irwin.

Tanaka H., Uejima S. and Asai K. (1982), Linear regression analysis with fuzzy model, *IEEE Transactions on Systems, Man, and Cybernetics* 12, 903-907.

رگرسیون کمترین مربعات خطای استوار در محیط فازی

امیر حمزه خمر[1]، محسن عارفی، محمد قاسم اکبری

گروه آمار، دانشگاه بیرجند

چکیده: در این مقاله، با استفاده از تعریف یک فاصله جدید بین اعداد فازی مثلثی، یک رویکرد کمترین مربعات جدید برای مدل رگرسیون فازی پیشنهاد شده است. از مزایای اصلی مدل معرفی شده حساسیت کم آن نسبت به انواع داده های پرت فازی می باشد. با استفاده از یک مثال شبیه سازی، عملکرد مدل پیشنهادی در مواجعه با انواع داده های پرت و عملکرد آن در مقایسه با رگرسیون فازی کمترین مربعات خطا براساس فاصله ژو (۱۹۹۱) مورد بررسی قرار گرفته است. نتایج نشان می دهد که مدل رگرسیون فازی جدید یک مدل استوار نسبت به انواع داده های پرت فازی بوده و از عملکرد بهتری نسبت به رگرسیون فازی کمترین مربعات خطا برخوردار می باشد.

واژه‌های کلیدی: رگرسیون فازی استوار، داده پرت، تابع کرنل، نیکویی برازش.

کد موضوع‌بندی ریاضی (۲۰۱۰): 03E72، 62J86، 62.

---

۱ مقدمه

در دیدگاه کلاسیک، نظریه رگرسیون بر این پایه استوار است که دادههای مشاهده شده کمیتهای عددی دقیقی هستند. اما در بسیاری از موارد دادهها به صورت دقیق مشاهده یا گزارش نمیشوند. در این موارد، با استفاده از نظریه فازی شهودی، میتوانیم برخی مدلهای رگرسیونی را براساس مشاهدات فازی صورتبندی نمائیم. روش های برآوردیابی پارامترهای مدل های رگرسیون فازی کمترین مربعات خطا حساسیت زیادی نسبت به داده های پرت یا دور افتاده (داده های که با روند موجود در بین اکثریت داده ها همخوانی ندارند) دارند. اغلب روش های موجود برآوردیابی پارامترهای این مدل ها با رویکرد کمترین مربعات خطا، تحت تاثیر داده های پرت قرار گرفته و برآوردهایی نامناسب، دور از انتظار و با خطای زیاد ارائه می دهند. رویکرد رگرسیون فازی استوار یک مدل رگرسیون فازی ایجاد می کند که تحت تاثیر داده های پرت کمتر قرار می گیرد. نویسندگان زیادی به مطالعه و تحقیق بر روی مدل های رگرسیونی استوار در محیط فازی پرداخته اند که برخی رویکردهای آنها خالی از کاستی نیست. در برخی رویکردهای ارائه شده ممکن است مدل رگرسیونی فازی نسبت به انواع مختلف داده های پرت استوار نباشد. از این رو، هدف این مقاله ارائه رویکردهای جدید و جامع در استوارسازی مدل های رگرسیون فازی در مواجعه با انواع داده های پرت است. از جمله پژوهش های انجام شده درباره مدل های رگرسیونی استوار در محیط فازی می توان به پیترز (۱۹۹۴) اوساش و اسکوتر (۲۰۰۲)، شون (۲۰۰۵)، چویی و باکلی (۲۰۰۸)، کولا و آپادین (۲۰۰۸)، هاشم پور و همکاران (۲۰۰۹ ،(دورسو و همکاران (۲۰۱۱ ،(طاهری و کلکین نما (۲۰۱۲ ،(چاچی و روزبه (۲۰۱۶)، یابوچی و همکاران (۲۰۱۶) و چاچی (۲۰۱۹) اشاره کرد. در این مقاله می خواهیم با استفاده از یک فاصله جدید (براساس رویکرد تابع کرنل (هسته))

[1] نام ارائه دهنده‌ی مقاله : amir.khammar@birjand.ac.ir

و انتخاب آن به عنوان یک مسئله بهینه سازی (یک تابع هدف) یک روش جدید برای برآورد ضرایب مدل رگرسیون فازی با داده های ورودی معمولی و داده های خروجی فازی ارائه کنیم. با ذکر یک مثال نشان داده می شود که مدل رگرسیون فازی بدست آمده از روش پیشنهادی نسبت به انواع داده های پرت فازی استوار می باشد.

۲ یک فاصله جدید روی مجموعه اعداد مثلثی

فرض کنید $Z = \{z_۱, z_۲, ..., z_n\}$ یک مجموعه ناتهی باشد، بطوریکه $z_i \in \Re^d$. تابع $K : Z \times Z \to \Re$ را یک تابع کرنل (هسته) مرکل می‌نامیم اگر و تنها اگر K تابعی متقارن باشد (یعنی $K(z_i, z_k) = K(z_k, z_i)$) و نامساوی $\sum_{i=۱}^{n} \sum_{k=۱}^{n} c_i c_k K(z_i, z_k) \geq ۰$، $n \geq ۲$، $c_r \in \Re$ برای $r = ۱, ۲, ..., n$ برقرار باشد (مرکل (۱۹۰۹)).

فرض کنید $\phi : Z \to F$ یک نگاشت غیرخطی از فضای ورودی Z به فضای آینده چندبعدی F باشد. با بکارگیری نگاشت ϕ ، ضرب داخلی $z_i^t z_k$ در فضای داخلی به $\phi^t(z_i)\phi(z_k)$ در فضای آینده نگاشته می‌شود. ایده اصلی استفاده از توابع کرنل این است که لازم نیست نگاشت غیرخطی ϕ به صراحت مشخص باشد، زیرا هر تابع کرنل مرکل می‌تواند به فرم $K(z_i, z_k) = \phi^t(z_i)\phi(z_k)$ بیان شود (مولر و همکاران (۲۰۰۱)).

در ادامه فاصله جدید بین دو عدد فازی مثلثی را معرفی می کنیم. با استفاده از ترفند کرنلی فاصله (تعریف شده در فوق)، فاصله جدیدی را برای دو عدد فازی LR به صورت زیر تعریف می‌کنیم:

$$\begin{aligned} D_K(\tilde{A}, \tilde{B}) = \frac{۱}{۳}\{ &\| \phi(m_a) - \phi(m_b) \|^۲ + \| \phi(m_a - \lambda l_a) - \phi(m_b - \lambda l_b) \|^۲ + \| \phi(m_a + \rho r_a) \\ &- \phi(m_b + \rho r_b) \|^۲\} \\ = \frac{۱}{۳}\{ &K(m_a, m_a) - ۲K(m_a, m_b) + K(m_b, m_b) \\ &+ K(m_a - \lambda l_a, m_a - \lambda l_a) - ۲K(m_a - \lambda l_a, m_b - \lambda l_b) + K(m_b - \lambda l_b, m_b - \lambda l_b) \\ &+ K(m_a + \rho r_a, m_a + \rho r_a) - ۲K(m_a + \rho r_a, m_b + \rho r_b) + K(m_b + \rho r_b, m_b + \rho r_b)\}, \end{aligned} \quad (۱.۲)$$

فاصله فوق میانگین مربعات خطاهست که روی فضای اصلی تعریف نشده، بلکه در یک فضای چندبعدی F با بکارگیری نگاشت غیرخطی ϕ روی مراکز و پهناهای چپ و راست دو عدد فازی تعریف شده است. برای بهره‌گیری از فاصله فازی کرنل، ابتدا باید تابع کرنل $K(.)$ و پارامتر هموارساز h تعیین شود. در اینجا، برای تعیین پارامتر هموارساز از یک روش اعتبار سنجی متقابل استفاده شده (توجه ۲.۳) و از تابع کرنل گوسی به فرم زیر بهره گرفته شده است:

$$K(z_i, z_k) = \exp\{-\frac{\| z_i - z_k \|^۲}{۲h^۲}\},$$

که $K(z_i, z_i) = ۱$. با اعمال تابع کرنل گوسی فوق در تعریف فاصله فازی کرنل داریم:

$$D_K(\tilde{A}, \tilde{B}) = ۲ - \frac{۲}{۳}\{K(m_a, m_b) + K(m_a - \lambda l_a, m_b - \lambda l_b) + K(m_a + \rho r_a, m_b + \rho r_b)\}. \quad (۲.۲)$$

برای سه عدد فازی $\tilde{A}$، $\tilde{B}$ و $\tilde{C}$، فاصله D_K در شرایط زیر صدق می‌کند:

۱) $D_K(\tilde{A}, \tilde{B}) = ۰$ اگر و تنها اگر $\tilde{A} = \tilde{B}$،

۲) $D_K(\tilde{A},\tilde{B}) = D_K(\tilde{B},\tilde{A})$،

۳) $D_K(\tilde{A},\tilde{C}) \leq D_K(\tilde{A},\tilde{B}) + D_K(\tilde{B},\tilde{C})$.

۳ رگرسیون فازی براساس رویکرد کمترین مربعات خطا

فرض کنید $X_i = (۱, x_{i۱}, x_{i۲}, ..., x_{ip})$ ، $(۱ \leq i \leq n)$ متغیرهای توضیحی غیرفازی و $\tilde{Y} = (Y, S_l, S_r)_T$ متغیر پاسخ فازی T با مرکز Y و ابهامات چپ و راست S_l و S_r باشند. هدف برازش یک مدل رگرسیونی فازی استوار با ضرایب فازی بر روی مجموعه داده فوق الذکر است، بطوریکه

$$\hat{\tilde{Y}} = \tilde{\beta}_۰ \oplus (\tilde{\beta}_۱ \otimes X_۱), \qquad (۱.۳)$$

که در آن $\tilde{\beta}_j = (\beta_j, \gamma_j, \theta_j)_T$، $j = ۰, ۱, ..., p$، یک عدد فازی مثلثی می‌باشد. برای بدست آوردن بهترین مدل فازی از ۱.۳، با استفاده از فاصله فازی کرنل (۲.۲)، ضرایب رگرسیونی $\tilde{\beta}_j$ را از طریق مینیمم کردن مجموع فواصل کرنل بین $\tilde{Y}_i = (y_i, s_{l_i}, s_{r_i})_T$ و $\hat{\tilde{Y}}_i = (\sum_{j=۰}^{p} \beta_j x_{ij}, \sum_{j=۰}^{p} \gamma_j x_{ij}, \sum_{j=۰}^{p} \theta_j x_{ij})_T$ یعنی $\sum_{i=۱}^{n} D_K(\tilde{Y}_i, \hat{\tilde{Y}}_i)$ (به عنوان تابع هدف) برآورد می‌کنیم، که معادل است با:

$$\begin{aligned} S &= ۲n - \frac{۲}{۳}\sum_{i=۱}^{n}\Big(K(y_i, \sum_{j=۰}^{p}\beta_j x_{ij}) + K(y_i - \lambda s_{l_i}, \sum_{j=۰}^{p}\beta_j x_{ij} - \lambda\sum_{j=۰}^{p}\gamma_j x_{ij}) \\ &\quad + K(y_i + \rho s_{r_i}, \sum_{j=۰}^{p}\beta_j x_{ij} + \rho\sum_{j=۰}^{p}\theta_j x_{ij})\Big), \end{aligned} \qquad (۲.۳)$$

که در آن $x_{i۰} = ۰$ و تابع کرنل $K(.,.)$، تابع کرنل گوسی می‌باشد.

توجه ۱.۳. *برای حل مسئله بهینه‌سازی فوق جهت برآورد ضرایب در مثال‌های عددی از دستور $NMinimize[]$ در نرم‌افزار "$Mathematica$" استفاده شده است. نرم‌افزار "$Mathematica$" قادر است مسئله بهینه‌سازی را به یک مسئله برنامه‌ریزی تبدیل کرده و نتایج مورد نظر را ارائه دهد.*

توجه ۲.۳. *قبل از ورود به بحث بهینه‌سازی تابع هدف معرفی شده در فوق، ابتدا باید پارامتر هموارساز h $(h > ۰)$ را در تابع کرنل برآوزد کنیم. انتخاب یک مقدار مناسب و بهینه برای پارامتر هموارساز یک موضوع مهم در مباحث ناپارامتری و برآوردیابی بر مبنای تابع کرنل است. در این مقاله، برای انتخاب یک مقدار بهینه h یک روش اعتبارسنجی متقابل ارائه می‌کنیم. برای این منظور، با رسم نمودار تابع هدف (۲.۳) به ازای مقادیر مختلف h، می‌توان از نقطه $h = m$ به بعد یک یکنوایی تقریبی را در نمودار مشاهده کرد، بطوریکه تفاوت چندانی در مقادیر تابع هدف با دور شدن از نقطه $h = m$ ملاحظه نمی‌شود. بنابراین می‌توان m را به عنوان یک مقدار مناسب و بهینه برای h در نظر گرفت $(h_{opt} = m)$.*

در ادامه یک مثال شبیه‌سازی جهت بررسی عملکرد مدل پیشنهادی در مواجعه با حالت‌های مختلفی از داده‌های پرت ارائه شده و عملکرد آن با مدل کمترین مربعات بر مبنای فاصله ژو (۱۹۹۱) مقایسه می‌شود.

مثال ۳.۳. *ابتدا یک نمونه ۱۰۰تایی از مجموعه داده $(\tilde{Y}, X_۱)$ طبق مدل رگرسیون خطی فازی زیر شبیه‌سازی کرده و سپس یک نمونه ۲۰تایی از داده‌های پرت را به مجموعه داده شبیه‌سازی شده اضافه می‌کنیم.*

$$\hat{\tilde{Y}} = (۱/۵, ۲/۵)_T \oplus (۳, ۰)_T x_{i۱} + \epsilon_i, \quad i = ۱, ۲, ..., ۱۰۰.$$

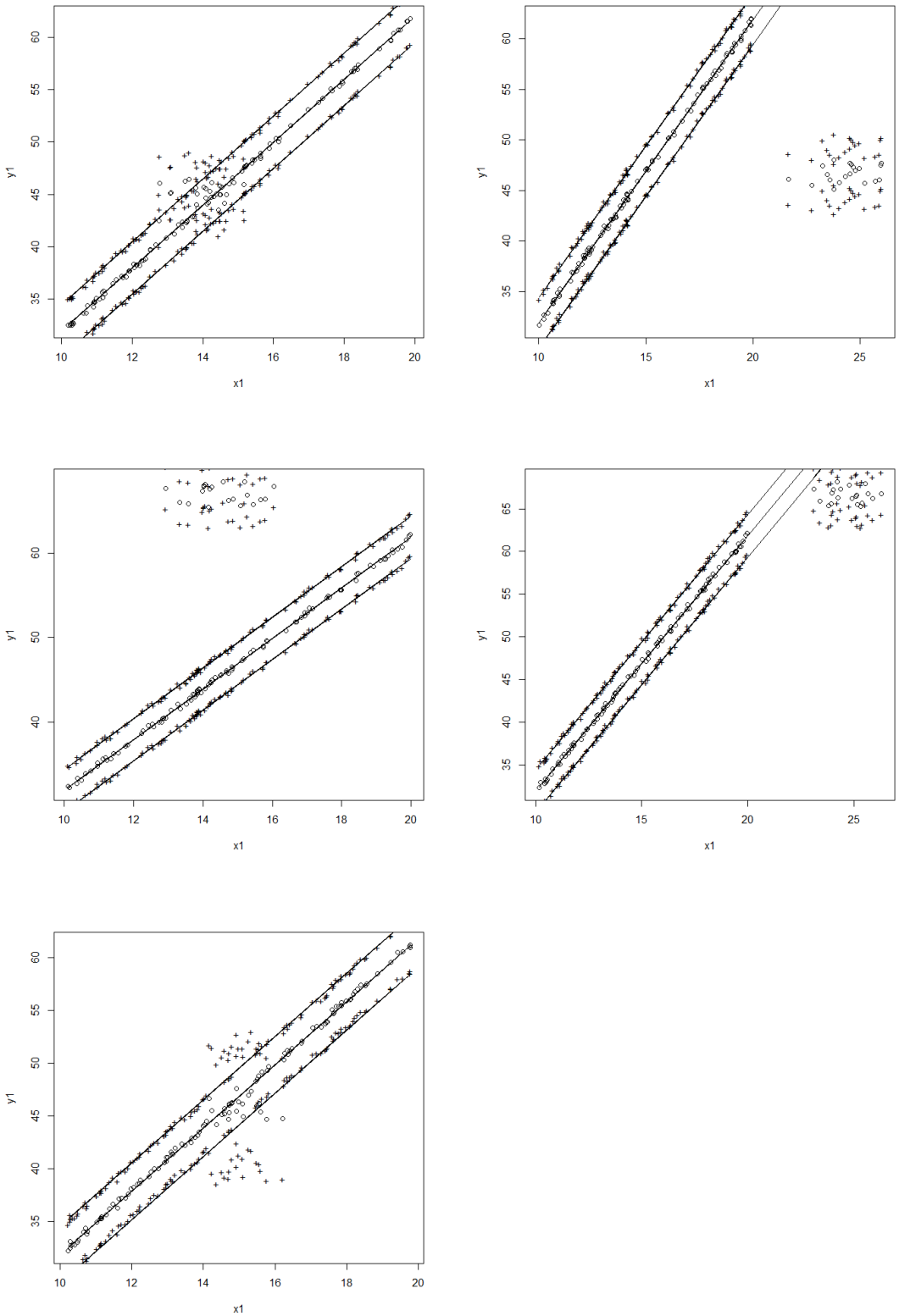
که در آن اعداد فازی از نوع مثلثی متقارن فرض شده و $x_{i1} \sim U(10, 20)$ و $\epsilon_i \sim N(0, 0/4)$، $i = 1, 2, ..., 100$. داده‌های پرت نیز توسط توزیع نرمال دومتغیره $N_2(\mu = (\mu_x, \mu_y), \sum = diag(0/8))$ در چهار نوع مختلف طبق جدول ۱ تولید می‌شوند. بنابراین ما ۵ نوع مختلف از مدل رگرسیون خطی فوق تولید کرده‌ایم که در شکل ۱ نشان داده شده‌اند. پس از انجام مراحل فوق الذکر برآورد ضرایب مدل رگرسیون فازی کمترین مربعات خطا و مدل رگرسیون فازی کمترین مربعات خطای پیشنهادی در مواجعه با هر یک از انواع داده‌های پرت به ترتیب مطابق جداول ۲ و ۳ نشان داده شده است. همچنین مقادیر شاخص‌های نیکویی برازش خطای نسبی ME و اندازه مشابهت MSM (طاهری و کلکین‌نما ۲۰۱۲) برای هر مورد در جداول ۲ و ۳ قابل مشاهده می‌باشد. مدل‌های برازش داده شده بر روی مجموعه داده‌های مورد بررسی در شکل ۱ نمایش داده شده است. با توجه به معیارهای نیکویی برازش و شکل ۱ می‌توان نتیجه گرفت مدل رگرسیون فازی پیشنهاد شده نسبت به انواع داده‌های پرت شبیه‌سازی شده استوار بوده و از عملکرد بهتری نسبت به مدل کمترین مربعات برخوردار می‌باشد.

جدول ۱: پارامترها داده‌های پرت

انواع داده پرت	پارامترها
بدون داده پرت	$N_2(\mu = (\bar{x}, \bar{y}), \sum = \text{diag}(0/8))$
x−جهتی	$N_2(\mu = (\max(x) + 4/5, \bar{y}), \sum = \text{diag}(0/8))$
y−جهتی	$N_2(\mu = (\bar{x}, \max(y) + 4/5), \sum = \text{diag}(0/8))$
نقاط اهرمی	$N_2(\mu = (\max(x) + 4/5, \max(y) + 4/5), \sum = \text{diag}(0/8))$
پهناهای پرت	$U(5, 6)$

جدول ۲: شاخص‌های نیکویی برازش و برآورد پارامترها در مدل پیشنهادی

انواع داده پرت	h_{obt}	$\tilde{\beta}_0 = (\beta_0, \gamma_0)_T$	$\tilde{\beta}_1 = (\beta_1, \gamma_1)_T$	ME	MSM
بدون داده پرت	۲/۲	$(2/023, 2/5)_T$	$(2/997, 0)_T$	۰/۵۹۳	۰/۷۷۴
x−جهتی	۴/۵	$(1/867, 2/5)_T$	$(3/002, 0)_T$	۰/۸۷۲	۰/۷۱۷
y−جهتی	۵/۲	$(1/951, 2/5)_T$	$(2/999, 0)_T$	۰/۸۸۳	۰/۷۱۷
نقاط اهرمی	۲/۲	$(1/989, 2/5)_T$	$(2/993, 0)_T$	۰/۸۸۷	۰/۷۱۷
پهناهای پرت	۰/۹	$(2/01, 2/52)_T$	$(2/99, 0)_T$	۰/۳۷۰	۰/۸۳۱



شکل ۱: انواع داده‌های پرت فازی مطابق مثال ۳.۳

جدول ۳: شاخص‌های نیکویی برازش و برآورد پارامترها در مدل کمترین مربعات

انواع داده پرت	$\tilde{\beta}_0 = (\beta_0, \gamma_0)_T$	$\tilde{\beta}_1 = (\beta_1, \gamma_1)_T$	ME	MSM
بدون داده پرت	$(۲/۵۶۹, ۲/۵)_T$	$(۲/۹۶۹, ۰)_T$	۰/۶۶۵	۰/۷۴۳
x−جهتی	$(۲۷/۹۷۲, ۲/۵)_T$	$(۱/۱۲۵, ۰)_T$	۳/۰۹۱	۰/۱۳۹
y−جهتی	$(۸/۱۲۷, ۲/۵)_T$	$(۲/۸۱۹, ۰)_T$	۳/۴۸۹	۰/۰۴۲
نقاط اهرمی	$(۱۰/۹۳۳, ۲/۵)_T$	$(۲/۳۶۰, ۰)_T$	۲/۰۵۸	۰/۳۴۷
پهناهای پرت	$(۲/۰۱, ۲/۵۲)_T$	$(۲/۹۹, ۰)_T$	۰/۴۵۳	۰/۷۹۵

مراجع

Chachi J. (2019), A weighted least-squares fuzzy regression for crisp input-fuzzy output data, *IEEE Transactions on Fuzzy Systems* 27, 739-748.

Chachi J. and Roozbeh M. (2017), A fuzzy robust regression approach applied to bedload transport data, *Communications in Statistics-Simulation and Computation* 47, 1703-1714.

Choi S.H. and Buckley J.J. (2008), Fuzzy regression using least absolute deviation estimators, *Soft Computing* 12, 257-263.

D'Urso P., Massari R. and Santoro A. (2011), Robust fuzzy regression analysis, *Information Sciences* 181, 4154- 4174.

Hassanpour H., Maleki H.R. and Yaghoobi M.A. (2009), A goal programming approach to fuzzy linear regresion with non-fuzzy input and fuzzy output data, *Asia-Pacific Journal of Operational Research* 26, 587-604.

Kula K.S. and Apaydin A. (2008), Fuzzy robust regression analysis based on the ranking of fuzzy sets, *International Journal Uncertainty, Fuzziness and Knowledge-Based Systems* 16, 663-681.

Mercer J. (1909), Functions of positive and negative type and their connection with the theory of integral equations, *Philosophical Transactions of the Royal Society A* 209, 441-458.

Mueller K.R., Mika S., Raetsch G.R., Tsuda K. and Schloelkopf F. (2001), An introduction to kernel-based learning algorithms, *IEEE Transactions on Neural Networks* 12, 181-202.

Oussalah M. and De Schutter J. (2002), Robust fuzzy linear regression and application for contact identifi-

cation, *Intelligent Automation and Soft Computing* 8, 31-39.

Peters G. (1994), Fuzzy linear regression with fuzzy intervals, *Fuzzy Sets and Systems* 63, 45-55.

Shon B.Y. (2005), Robust fuzzy linear regression based on M-estimators, *Applied Mathematics and Computation* 18, 591-601.

Taheri S.M. and Kelkinnama M. (2012), Fuzzy linear regression based on least absolute deviations, *Irannian Journal of Fuzzy Systems* 9, 121-140.

Tanaka H., Uegima S. and Asai K. (1982), Linear regression analysis with fuzzy model, *IEEE Transactions on Systems* 12, 903-907.

Xu R. and Li C. (2001), Multidimensional least squares fitting with a fuzzy model, *Fuzzy Sets and Systems* 119, 215-223.

نمودار کنترل آماری کیفیت فازی برای داده‌های طول عمر

بهرام صادقپور گیلده، سیده فاطمه اصغری بایگی[1]

گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

چکیده: در دنیای امروزی داده‌های طول عمر از اهمیت زیادی برخوردارند. به همین دلیل مورد توجه آماردانان، به ویژه آماردانان صنعتی می‌باشند. در این مقاله نحوه‌ی رسم نمودارهای کنترل آماری کیفیت برای داده‌های طول عمر، وقتی‌که پارامتر توزیع طول عمر (نمایی) فازی است را بررسی می‌کنیم. برای این منظور ابتدا از یک تابع درجه عضویت برای تبیین تحت‌کنترل یا خارج از کنترل بودن مشاهدات استفاده می‌کنیم، سپس با استفاده از فاصله‌ی $D_{p,q}$ بین اعداد فازی، نمودار کنترلی مشابه نمودار کنترل سنتی رسم می‌کنیم.

واژه‌های کلیدی: توزیع طول عمر، عدد فازی، فاصله بین اعداد فازی، نمودار کنترل آماری کیفیت.

---

۱ مقدمه

نمودارهای کنترل آماری کیفیت و برنامه‌های نمونه‌گیری برای پذیرش ابزارهایی مهم برای اطمینان از تولید محصولات با کیفیت بالا هستند. این نمودار ها که اولین بار توسط شوهارت در سال (۱۹۳۱) معرفی شد، امروزه به یکی از مهم‌ترین ابزارهای بهبود کیفیت تبدیل شده است.

نمودارهای کنترل برای نظارت بر یک فرایند در هنگام تولید اعمال می‌شود تا اقدامات به موقع در مورد فرایند تولید براساس نمودار کنترل آماری کیفیت انجام شود. زمانی‌که تحلیل یک نمودار کنترل نشان دهد که فرایند تولید تحت‌کنترل است هیچ اقدامی انجام نمی‌شود. با این حال اگر نمودار کنترل هشداری مبنی بر وجود تغییر در فرایند بدهد، باید اقدامات مناسب برای تحت‌کنترل درآوردن فرایند انجام شود. بنابراین یک نمودار کنترل آماری کیفیت، زمان مناسب انجام اقدامات اصلاحی را نشان می‌دهد. دو حد کنترل بالایی (UCL)[1] و پایینی (LCL) [2] برای نظارت بر میانگین یا واریانس فرایند تعریف شده است.

نمودارهای کنترل با استفاده از این محدودیت ها برای کاهش محصولات معیوب و در نتیجه افزایش سود شرکت‌های تولیدی بسیار مفید هستند. حفظ کیفیت با استفاده از نمودارهای کنترل شهرت خوبی برای شرکت‌های تولیدی در بازار به ارمغان می‌آورد.

یکی از انواع داده‌ها که به دلیل اهمیت زیاد مورد علاقه‌ی محققین و سرمایه‌گذاران است، داده‌های طول عمر می‌باشد که در واقع نشان دهنده‌ی زمان خرابی یک سیستم است. اما، موقعیت‌های مختلفی وجود دارد که پارامترهای توزیع دقیقا مشخص نیستند که در نتیجه حدود کنترل به طور دقیق در دسترس نیست. برای غلبه بر این مشکل ابتدا یک تابع درجه عضویت برای حدود کنترل در حالتی که پارامترهای نامعلوم به صورت اعداد فازی هستند توسعه می‌دهیم، سپس با استفاده از معیار $D_{p,q}$ یک نمودار کنترل آماری کیفیت مشابه نمودارهای کنترل کیفیت سنتی ارائه می‌دهیم. همچنین با توجه به این‌که در داده‌های طول عمر اغلب زمان شکست اهمیت دارد و داده‌ها

[1] نام ارائه دهنده‌ی مقاله : fatemeasghari0@gmail.com

[1] Upper Control Limit
[2] Lower Control Limit

با طول عمر بالا قابلیت اطمینان خوبی دارند با در نظر گرفتن تنها حد پایینی فرایند نمودارهای کنترلی به صورت فازی و هم‌چنین غیر فازی به کمک معیار $D_{p,q}$ طراحی می‌کنیم. هم‌چنین در پایان معیاری برای بررسی درجه تحت کنترل بودن یک داده‌ی طول عمر معرفی و محاسبه می‌شود.

۲ طراحی نمودار کنترل پیشنهادی

در آزمون فرضیه های فازی،فرضیه ها به صورت زیر اند:

$$\begin{cases} H۰ : \theta \; is \; H_۰(\theta), \\ H۱ : \theta \; is \; H_۱(\theta), \end{cases} \tag{۱.۲}$$

یا

$$\begin{cases} H۰ : \tilde{\theta} = \tilde{\theta_۰} \\ H۱ : \tilde{\theta} = \tilde{\theta_۱}, \end{cases} \tag{۲.۲}$$

که در آن$H_j(\theta), j = ۰, ۱$اشاره دارد که θ یک زیر مجموعه فازی از Θ (فضای پارامتر) با تابع عضویت $H_j(\theta)$ است و می‌نویسیم $H_j : \Theta \longrightarrow [۰, ۱], j = ۰, ۱$.

فرض کنید متغیر تصادفی X دارای تابع چگالی احتمال فازی $(FPDF)\tilde{f}(x, \tilde{\theta})$ که $\tilde{\theta}$ یک پارامتر فازی است، باشد. تحت فرضیه $H_j(\theta), \; j = ۰, ۱$ ، λ برش برای چگالی احتمال فازی برابر است با:

$$\tilde{f}_j(x, \theta_j)\,[\lambda] = \{f_j(x, \theta_j) | \theta_j \epsilon \tilde{\theta}_j\,[\lambda]\} = \{f_j^L\,[\lambda]\,,\; f_j^U\,[\lambda]\} \quad ; \quad j = ۰, ۱, \tag{۳.۲}$$

که

$$\begin{aligned} f_j^L(x)\,[\lambda] &= min\{f_j(x, \theta_j) | \theta_j \epsilon \tilde{\theta}_j\,[\lambda]\}, \\ f_j^U(x)\,[\lambda] &= max\{f_j(x, \theta_j) | \theta_j \epsilon \tilde{\theta}_j\,[\lambda]\}. \end{aligned} \tag{۴.۲}$$

متغیر تصادفی T (که مشخصه‌ی طول عمر می‌باشد) دارای توزیع نمایی با پارامتر فازی و تابع چگالی احتمال زیر است:

$$\tilde{f}_j(x, \theta_j)\,[\lambda] = \{\frac{۱}{\theta_j} e^{-\frac{t}{\theta}} | \theta_j \epsilon \tilde{\theta}_j\,[\lambda]\} \quad ; \quad j = ۰, ۱. \tag{۵.۲}$$

که پارامتر$\tilde{\theta}$ میانگین عمر نامعلوم و فازی این فرایند است. هدف از طراحی نمودار کنترل ابتدا ارائه‌ی طیف وسیعی از تغییراتی که به طور شانسی (تصادفی) برای فرایند طول عمر ایجاد می‌شود و دوم با ادامه‌ی مشاهدات شناسایی علل خاصی که موجب تغییر می‌گردد و درنهایت اقدام جهت اصلاح فرایند می‌باشد. با توجه به اینکه به‌دست آوردن مشاهدات بیشتر در داده‌های طول عمر برای آزمایش محصولات تولیدی نیازمند هزینه و اتلاف وقت بسیار است، مطلوب به نظر می‌رسد که با استفاده از اولین زمان شکست در نمونه‌های داده‌شده نمودار کنترلی براساس زمان دلخواه طراحی نماییم. بنابراین اندازه گیری حداقل طول عمر تاثیر قابل توجهی در کاهش مدت آزمایش و تعداد نمونه‌هایی که از بین می‌رود خواهد داشت. فرض کنید $T_۱, T_۲, ..., T_N$ یک نمونه‌ی N تایی از زمان‌های شکست باشد، در نتیجه زمان مشاهده‌ی اولین شکست به صورت

$$Z = \min(T_۱, T_۲, ..., T_N), \tag{۶.۲}$$

است. متغیر تصادفی Z که در (۶.۲) تعریف شد دارای تابع توزیع احتمال فازی با λ برش زیر است:

$$\tilde{F}_z(z;\theta_j;N)\,[\lambda] = \{۱ - ^{-zN/\theta_j}\,|\theta_j \epsilon \tilde{\theta}_j\,[\lambda]\} \quad ; \quad j = ۰, ۱, \tag{۷.۲}$$

زمان مورد انتظار برای اولین شکست $\tilde{E}(Z)$ برابر است با:

$$\tilde{E}(Z) = \tilde{\theta}/N, \tag{۸.۲}$$

با توجه به مزیت استفاده از اولین زمان‌های شکست، آزمایش را برای M نمونه با اندازه‌ی N انجام می‌دهیم، منتظر می‌مانیم تا اولین شکست رخ‌دهد (به جای منتظر ماندن برای آخرین شکست)، در نتیجه یک نمونه‌ی M تایی از اولین زمان شکست با تابع توزیع (۷.۲) خواهیم داشت.

۳ فاصله‌ی $D_{۲,۱/۲}$ بین اعداد فازی

فرض کنید $\tilde{a}$ و $\tilde{b}$ دو عدد مثلثی متقارن به صورت زیر باشند:

$$\tilde{a} = (a_۱, a_۲, a_۳) \qquad \tilde{b} = (b_۱, b_۲, b_۳), \tag{۱.۳}$$

در اینصورت فاصله‌ی $D_{۲,۱/۲}$ ، بین دو عدد فازی نامنفی $\tilde{a}$ و $\tilde{b}$ در فضای اقلیدسی با استفاده از یگولالمپ (۲۰۰۵) توسط رابطه‌ی زیر به‌دست می آید :

$$D_{۲,۱/۲}(\tilde{a},\tilde{b}) = \sqrt{\frac{۱}{۶}\left[\sum_{i=۱}^{۳}(b_i - a_i)^۲ + (b_۲ - a_۲)^۲ + \sum_{i\epsilon ۱,۲}(b_i - a_i)(b_{i+۱} - a_{i+۱})\right]}, \tag{۲.۳}$$

۴ نمودارهای کنترل برای کوچکترین داده‌ی طول عمر

یک روند نمونه‌گیری مستمر که تعداد نمونه‌های مورد آزمایش آن N باشد را در نظر بگیرید. آزمایش به‌محض مشاهده‌ی اولین شکست خاتمه می‌یابد. $\tilde{\theta}$ میانگین فازی این روند می‌باشد که تنها به N وابسته است. بعد از اینکه M نمونه مورد آزمایش قرار گرفت یک نمونه با اندازه‌ی M از اولین زمان‌های شکست برای تفسیر به دست می‌آید. می‌توان چهار آماره برای طراحی نمودارهای کنترل در نظر گرفت (یگولالمپ، ۲۰۰۵):

۱. اولین زمان شکست در نمونه‌ی N تایی

۲. کوچکترین زمان شکست از میان M اولین زمان‌های شکست نمونه با اندازه‌ی N

۳. بزرگترین زمان شکست از میان M اولین زمان‌های شکست نمونه با اندازه‌ی N

۴. میانگین زمان شکست از میان M اولین زمان‌های شکست نمونه با اندازه‌ی N

در این مقاله حالت اول را بررسی می‌نماییم و سایر حالت‌ها در آینده مورد بررسی قرار خواهد گرفت.

۱.۴ نمودار کنترل برای اولین زمان شکست در نمونه‌با اندازه‌ی N

متغیر تصادفی T، طول عمر محصولات تولیدی با تابع چگالی (۵.۲) است. $T_1, T_2, ..., T_N$ نمونه‌ی N تایی طول عمر می‌باشد، هدف محاسبه‌ی حدود کنترل متغیر تصادفی Z که در (۶.۲) تعریف شد، می‌باشد.

تعریف ۱.۴. *λ برش حدکنترل بالایی و پایینی فازی در سطح اطمینان $(1-\alpha)$ برای متغیر تصادفی Z، با تابع توزیع (۷.۲) که دارای پارامتر فازی است، را به ترتیب با $U\tilde{C}L_Z$ و $L\tilde{C}L_Z$ نشان می‌دهیم و به صورت زیر تعریف می‌کنیم:*

$$L\tilde{C}L_z[\lambda] = \{LCL_z | \int_0^{LCL_z} f(t)dt = \alpha/2 \quad ; \quad \theta \epsilon \tilde{\theta}[\lambda]\}, \tag{۱.۴}$$

و

$$U\tilde{C}L_z[\lambda] = \{UCL_z | \int_{UCL_z}^{\infty} f(t)dt = \alpha/2 \quad ; \quad \theta \epsilon \tilde{\theta}[\lambda]\}, \tag{۲.۴}$$

بدین ترتیب داریم:

$$L\tilde{C}L_z[\lambda] = \left[LCL_z^L, L\tilde{C}L_z^U\right] \quad ; \quad U\tilde{C}L_z[\lambda] = \left[UCL_z^L, U\tilde{C}L_z^U\right], \tag{۳.۴}$$

با توجه به (۵.۲) داریم:

$$\begin{aligned} L\tilde{C}L_z[\lambda] &= \{LCL_z | \int_0^{LCL_z} \frac{N}{\theta}^{-tN/\theta_0} dt = \alpha/2 \quad ; \quad \theta \epsilon \tilde{\theta}[\lambda]\}, \\ &= \{LCL_z | LCL_z = -\frac{\tilde{\theta}}{N}\ln(1-\alpha/2) \quad ; \quad \theta \epsilon \tilde{\theta}[\lambda]\}, \end{aligned} \tag{۴.۴}$$

و

$$\begin{aligned} U\tilde{C}L_z[\lambda] &= \{UCL_z | \int_{UCL_z}^{\inf} \frac{N}{\theta}^{-tN/\theta_0} dt = \alpha/2 \quad ; \quad \theta \epsilon \tilde{\theta}[\lambda]\} \\ &= \{UCL_z | UCL_z = -\frac{\tilde{\theta}}{N}\ln(\alpha/2) \quad ; \quad \theta \epsilon \tilde{\theta}[\lambda]\}, \end{aligned} \tag{۵.۴}$$

۵ نمودار کنترل آماری کیفیت فازی

فرضیه‌های فازی زیر را با استفاده از داده‌های طول عمر به روش شبیه‌سازی آزمون می‌کنیم:

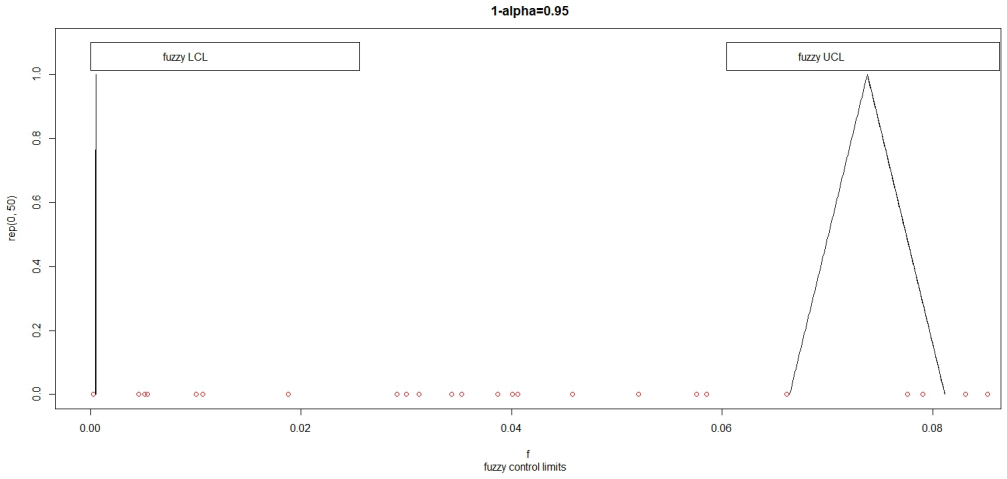
$$\begin{cases} H0 : \tilde{\theta}_0 = 0/4 \\ H1 : \tilde{\theta}_1 = 0/9, \end{cases} \tag{۱.۵}$$

توابع عضویت این فرضیه‌ها را به صورت زیر در نظر می‌گیریم:

$$H_0(\theta) = \begin{cases} \dfrac{\theta - 0/34}{0/4 - 0/34} & , \quad 0/34 \leqslant \theta \leqslant 0/4 \\ \dfrac{0/46 - \theta}{0/46 - 0/4} & , \quad 0/4 \leqslant \theta \leqslant 0/46, \end{cases} \tag{۲.۵}$$

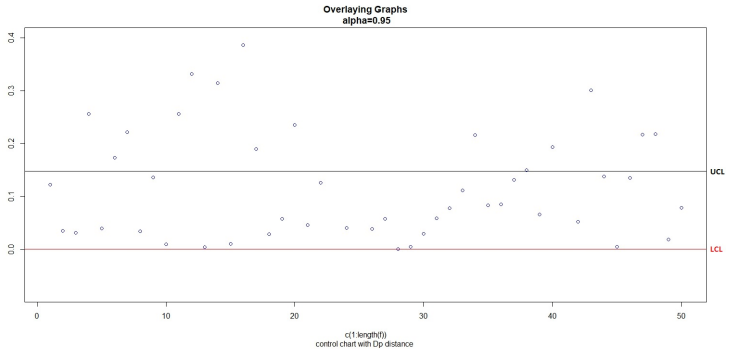
$$H_{۱}(\theta)=\begin{cases}\dfrac{\theta-۰/۶۴}{۰/۹-۰/۸۴} & , \quad ۰/۸۴\leqslant\theta\leqslant۰/۹\\ \dfrac{۰/۹۶-\theta}{۰/۹۶-۰/۹} & , \quad ۰/۹\leqslant\theta\leqslant۰/۹۶,\end{cases} \tag{۳.۵}$$

به ازای $\alpha=۰/۰۵$ نمودار کنترل آماری کیفیت فازی با استفاده از رابطه (۴.۴) و (۵.۴) در شکل ۱ آمده است.



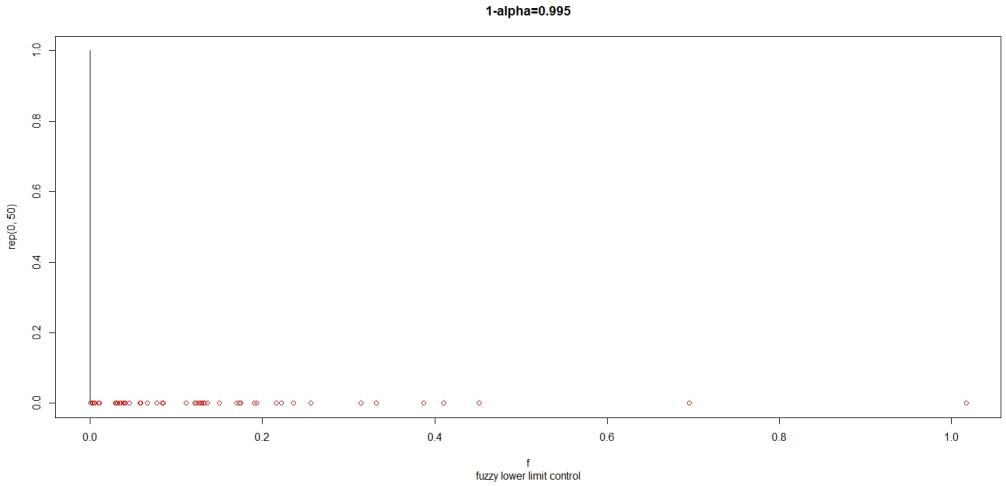
شکل ۱: حدود کنترل فازی برای داده‌های طول عمر

با توجه به شکل ۱ می‌توان در مورد تحت‌کنترل بودن یا نبودن داده‌های طول عمر به کمک حدود کنترل فازی اظهار نظر کرد. با استفاده از فاصله‌ی $D_{۲,۱/۲}$ نمودار کنترل آماری کیفیت داده‌های شکل ۱ را به دست می‌آوریم. ابتدا یک مقدار واقعی صفر را به صورت عدد فازی مثلثاتی $(۰,۰,۰)$ در نظر می‌گیریم، فاصله‌ی حد کنترل بالایی و پایینی از آن را به کمک رابطه‌ی (۷.۲) محاسبه نموده و مقادیر واقعی برای LCL_z و UCL_z به‌دست می‌آوریم و نمودار شکل ۲ حاصل می‌شود.



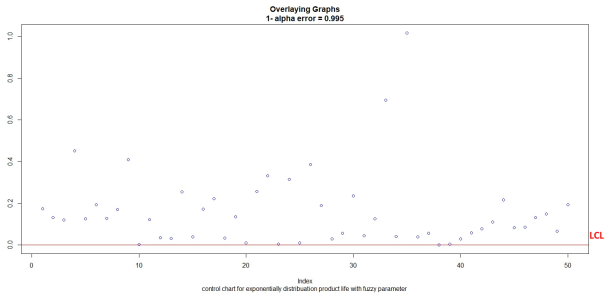
شکل ۲: حدود کنترل واقعی با استفاده از فاصله‌ی $D_{۲,۱/۲}$ برای داده‌های طول عمر

در داده‌های طول عمر اغلب، داده‌هایی که از حد بالا بزرگتر هستند خارج از کنترل محسوب نمی‌شوند بلکه برای فرایند مفید نیز هستند. اما اگر داده از حد پایینی کمتر باشند خارج از کنترل محسوب می‌شود و درصد این داده‌ها برای انجام اقدامات اصلاحی اهمیت دارد. به‌همین دلیل در اینجا نمودار کنترلی را با در نظر گرفتن تنها حد کنترل پایینی و سطح اطمینان $۱-\alpha=۰/۹۹۵$ طراحی می‌نماییم. نمودار ۳ برای این منظور طراحی شده است.



شکل ۳: حد کنترل پایینی فازی برای داده‌های طول عمر

نمودار با استفاده از فاصله‌ی $D_{۲,۱/۲}$ برای $۱ - \alpha = ۰/۹۹۵$ در شکل ۴ نمایش داده شده است.



شکل ۴: حدود کنترل واقعی با استفاده از فاصله‌ی $D_{۲,۱/۲}$ برای داده‌های طول عمر

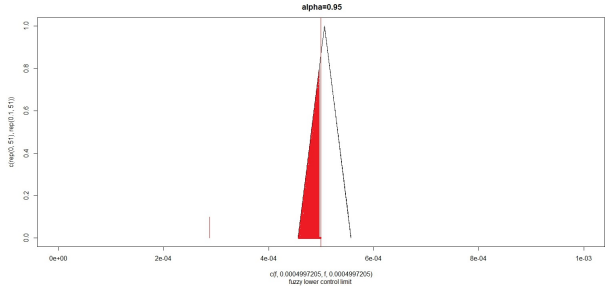
۶ بررسی یک معیار برای تحت کنترل بودن داده‌ها در نمودار کنترل آماری کیفیت فازی

با استفاده از مساحت زیر منحنی نمودار کنترل آماری کیفیت فازی می‌توان معیاری تحت عنوان درجه تحت‌کنترل بودن که آن را با $d_{in-control}$ نشان می‌دهیم، ارائه کرد. در این حالت دیگر نیازی به محاسبه‌ی معیارهایی مثل فاصله‌ی $D_{۲,۱/۲}$ نیست و می‌توان درجه تحت‌کنترل بودن را محاسبه نمود. فرض کنید مساحت زیر نمودار LCL فازی را با $A_{L\tilde{C}L}$ ، مساحت زیر نمودار UCL فازی را با $A_{U\tilde{C}L}$ نشان دهیم، آنگاه با توجه به شکل ۵ درجه تحت کنترل بودن برای مشاهده‌ای که روی تکیه‌گاه LCL قرار گیرد، از روی حد پایینی به‌صورت زیر محاسبه می‌شود:

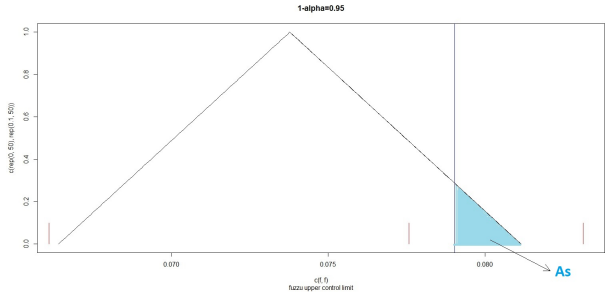
$$d_{in-control} = \frac{A_s}{A_{L\tilde{C}L}}, \qquad (۱.۶)$$

که در آن منظور از A_s مساحت قسمت قرمز رنگ در نمودار ۵ است، در این حالت مساحت سمت چپ نمودار را در نظر می‌گیریم.

برای نقطه‌ی G در شکل ۵، $d_{in-control} = ۰/۳۷۳۶۴۹۷$ محاسبه شده است.



شکل ۵: نمودار کنترل برای محاسبه‌ی $d_{in-control}$ برای نقطه‌ی G مفروض



شکل ۶: نمودار کنترل برای محاسبه‌ی $d_{in-control}$ برای نقطه‌ی G مفروض

به علاوه به کمک حد کنترل بالایی فازی این معیار به صورت زیر است:

$$d_{in-control} = \frac{A_s}{A_{U\tilde{C}L}}, \tag{۲.۶}$$

توجه داشته باشید که در محاسبه درجه‌ی تحت کنترل بودن به کمک حد بالایی، مساحت سمت راست مثلث را در نظر می‌گیریم. برای نقطه‌ی G در نمودار ۶، $d_{in-control} = ۰/۱۰۲۳۳۰۶$ محاسبه شده است. که در آن منظور از A_s مساحت قسمت آبی رنگ در نمودار ۶ می‌باشد. سه حالت خواهیم داشت:

-اگر $d_{in-control} = ۰$ آنگاه داده‌ی مورد نظر خارج از کنترل است،

- اگر $d_{in-control} = ۱$ آنگاه داده‌ی طول عمر تحت کنترل می‌باشد،

- زمانی‌که $۰ < d_{in-control} < ۱$ داده‌ی طول عمر مورد نظربا درجه‌ی خاصی تحت کنترل است.

جدول ۱ درجه تحت کنترل بودن را برای ۵۰ داده‌ی طول عمربررسی کرده است، به کمک این جدول می‌توان در مورد تحت‌کنترل بودن یا نبودن هر داده اظهار نظر کرد.

جدول ۱: درجه‌ی تحت کنترل بودن برای داده‌های طول عمر

شماره داده	درجه‌ی تحت کنترل بودن	شماره‌ی داده	درجه ی تحت کنترل بودن ۸
۱	۰	۲۶	۱
۲	۰/۳۷۳۶۴۹۷	۲۷	۱
۳	۱	۲۸	۰
۴	۰	۲۹	۱
۵	۱	۳۰	۱
۶	۰	۳۱	۱
۷	۰	۳۲	۰/۱۲۳۸۰۱
۸		۳۳	۰
۹	۰	۳۴	۰
۱۰	۱	۳۵	۰
۱۱	۰	۳۶	۰
۱۲	۰	۳۷	۰
۱۳	۱	۳۸	۰
۱۴	۰	۳۹	۱
۱۵	۱	۴۰	۰
۱۶	۰	۴۱	۰
۱۷	۰	۴۲	۱
۱۸	۱	۴۳	۰
۱۹	۱	۴۴	۰
۲۰	۰	۴۵	۱
۲۱	۱	۴۶	۰
۲۲	۰	۴۷	۰
۲۳	۰	۴۸	۰
۲۴	۱	۴۹	۱
۲۵	۰	۵۰	۰/۱۰۲۳۳۰۶

مراجع

Sadeghpour-Gildeh B. and Gien D. (2001), Measurement error effect on the performance of fuzzy capability index, *Actes du Quatrieme Congres Qualite et Surete de Fonctionnement*, Annecy, France, 222-230.

Yegulalp T.M. (2005), Control Charts for exponetially distribuation product life, *Naval Research Logistics Quarterly* 22, 696-712.

پیش‌بینی رتبه‌ی سهمیه‌ی داوطلبان آزمون ورودی دانشگاه‌ها، با استفاده از روش‌های داده‌کاوی

سید مهدی علوی[۱]، سید باقر میراشرفی[۲]، بهروز مینایی بیدگلی[۳]، آزاده بدریان[۴]

[۱] سازمان سنجش آموزش کشور

[۲] دانشگاه مازندران

[۳] دانشگاه علم و صنعت ایران

[۴] سازمان سنجش آموزش کشور

چکیده: امروزه، دانشگاه‌ها و مراکز آموزش عالی اصلی‌ترین منبع تأمین نیروی متخصص در تمام حوزه‌های کشور هستند از این رو نظام آموزش عالی کشور با افزایش روزافزون تعداد متقاضیان در آزمون‌های ورودی دوره‌های مختلف از جمله کنکور سراسری روبروست. رشد فراوان این متقاضیان، اهمیت گزینش علمی دقیق و عادلانه آنان را بیش‌از پیش نمایان نموده است. این آزمون ها طی سالهای اخیر دستخوش تغییرات بسیار شده، از جمله این تغییرات در کنکور سراسری مصوبه سال ۱۳۸۶ مجلس شورای اسلامی در خصوص حذف آزمون ورودی دانشگاه‌ها بود که در این راستا نمرات سوابق تحصیلی داوطلبان در زمینه سنجش و تفکیک آنان از اهمیت ویژه‌ای پیدا کرد. این مقاله قصد دارد با اتکا به نحوه اجرای این مصوبه و بکارگیری روشهای داده‌کاوی، با بررسی داده‌ها و نمرات سوابق تحصیلی دوره دبیرستان و پیش دانشگاهی، جنسیت، محل اخذ مدرک دیپلم و پیش‌دانشگاهی، سهمیه‌نهایی و ... پاسخ این سوال را بیابد که "آیا می‌توان با کمک روش‌های داده کاوی، مدلی برای پیش‌بینی رتبه سهمیه‌نهایی داوطلبان براساس داده ها و نمرات سوابق تحصیلی دیپلم و پیش دانشگاهی آنان، ارائه نمود؟"

واژه‌های کلیدی: داده‌کاوی، کنکور، سوابق تحصیلی، درخت تصمیم، پیش‌بینی رتبه.

کد موضوع‌بندی ریاضی (۲۰۱۰): 60A86، 62C86.

---

۱ مقدمه

در سالهای اخیر اهمیت داده ها بر کسی پوشیده نیست. این داده ها و اطلاعات بعنوان منبعی حیاتی در سازمان‌ها محسوب می‌شوند. داده‌کاوی سامانه‌اي هوشمند زیرمجموعه هوش مصنوعی است که در آن داده‌های موجود در پایگاه داده تحلیل و بررسی می نماید که منجر به کشف دانش‌ها، روندهای احتمالی، ارتباط‌هاي غیر محسوس و الگوهای مخفی در میان انبوه داده‌ها می‌شود (بیگاس ، ۱۹۹۶). وقتی در پایگاه داده یک سازمان، اطلاعات متنوعی وجود داشته باشد، با استفاده از داده‌کاوی می‌توان الگوهایی را بر روی آن کشف نمود که با روابط درون پایگاه داده نمی‌توان به آن پی برد (اُسمانبگوویچ و سولجیچ ، ۲۰۱۲).

این مقاله قصد دارد از منظری دیگر (متفاوت از میزان موفقیت داوطلبان)، با کمک روشهای داده‌کاوی بررسی نماید که آیا با استفاده از

[۱] نام ارائه دهنده‌ی مقاله : Alavimahdi61@yahoo.com

داده‌های مربوط به سوابق تحصیلی داوطلبان در سال‌های گذشته، می‌توان یک مدل صحیح برای پیش‌بینی میزان تأثیر حاصل از اعمال سوابق تحصیلی در رتبه سهمیه‌نهایی این داوطلبان در سال‌های آتی، ارائه نمود؟ در راستای دسترسی به اهداف داده‌کاوی در پژوهش فعلی از الگوریتم‌های C&R ، QUEST و C۵,۰ در درخت‌تصمیم براي کسب نتایج بهتر استفاده شده است.

۲ روش‌شناسی طرح

از جمله روش‌های داده‌کاوی مورد استفاده در این طرح درخت تصمیم است که مجموعه‌ای از نودهای تصمیم‌گیری[۱] بوده که با شاخه‌ها[۲] به هم متصل شده‌اند و از نود ریشه[۳] آغاز و به نودهای برگ[۴] ختم می‌گردند. ایجاد درخت تصمیم شامل دو مرحله؛ ایجاد و رشد درخت و هرس آن با هدف کاهش خطای پیش‌بینی است.

تمام الگوریتم‌های درخت، با نگرش بالا به پایین اجرا می‌شود. یکی از روش‌های معمول ایجاد درخت، انتخاب معیاری برای انشعاب گره های بالایی به تعدادی زیر گره می‌باشد.

فرآیند ساختن درخت تصمیم با در نظر گرفتن یک گره ریشه در بالای آن شروع شده و در ادامه یک رکورد جدید در گره ریشه وارد شده و با انجام آزمونی بر روی آن مشخص می‌شود که این رکورد به کدام گره فرزند متعلق است. روش‌های مختلفی برای انتخاب این آزمون وجود دارد که تمام آنها هدف یکسانی دارند و آن، تعیین روشی برای بهترین جدا سازی میان داده‌ها است. این عمل آنقدر ادامه می‌یابد تا تمام داده‌ها در گره‌ها قرار داده شوند. یعنی به نود برگ در درخت برسیم. در این صورت تمام رکوردهایی که در مسیر رسیدن به یک برگ قرار دارند، در یک دسته قرار می‌گیرند (حج فروش ، ۱۳۸۱).

معیاری که برای اثربخشی یک درخت تصمیم سنجیده می‌شود، درصد داده‌هایی است که درست دسته‌بندی می‌شوند و دسته پیش بینی شده آنها با دسته واقعی آنها یکسان است. هر راه ایجاد شده از ریشه به برگ معادل یک قاعده است.گاهی اوقات بریدن برخی از شاخه‌های ضعیف، بهبود پیش‌بینی در شاخه‌های دیگر را در پي دارد. مهمترین هدف از دسته‌بندی، ایجاد مدلی برای پیش بینی است.

۳ مدلسازی

داده‌های مورداستفاده در مقاله، پایگاه داده ثبت‌نام آزمون سراسری سال‌های ۹۵ ، ۹۶ و سوابق تحصیلی ارائه‌شده از سوی وزارت آموزش پرورش بوده است. جداول داده‌ای مربوط به ثبت‌نام و سوابق تحصیلی داوطلبان در آزمون سراسری سال‌های ۹۵ ، ۹۶ به ترتیب دارای ۱۱۵۱۲۹۲،۱۰۷۵۶۸۲ رکورد و بیش از ۵۰۰۰ فیلد مختلف، در قالب فایل Spss، به‌عنوان داده‌های اولیه تحقیق مورداستفاده قرار می‌گیرد. هر رکورد از این پایگاه داده حاوی اطلاعات یک داوطلب است. با توجه به شباهت کلی داده‌ها، در این تحقیق اطلاعات داوطلبان شرکت‌کننده در گروه آزمایشی علوم تجربی به‌صورت خاص استفاده خواهد شد.

با توجه به استفاده از الگوریتم‌های مختلف درخت تصمیم و ضرورت تولید متغیرهای کلاسه‌بندی اعم از نمره‌کل نهایی، رتبه‌کل در سهمیه، در داده‌های سال ۱۳۹۵ براساس دهک های تعیین شده متغیرهای NKOL۱Class و RTBKOL۱Class با کرانه‌های نشان داده شده در ذیل، ساخت شده است.

---

[۱] Decision Nodes
[۲] Branches
[۳] Root Node
[۴] Leaf Nodes

جدول ۱: دهک‌بندی نمره کل نهایی و رتبه در سهمیه‌نهایی منطقه یک در سال ۱۳۹۵

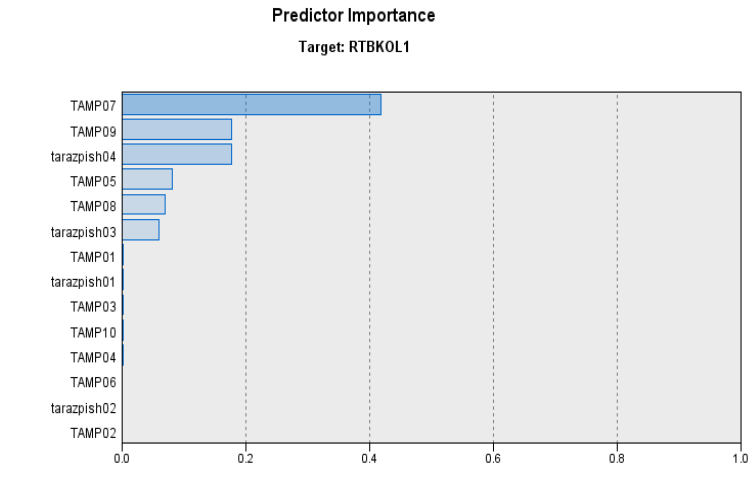
دهک	نهایی کل نمره	سهمیه در رتبه
۰	۴۲۲۶	۵۸۷۹
۱	۴۶۸۱	۱۱۹۴۹
۲	۵۰۸۷	۱۷۹۷۵
۳	۵۴۹۰	۲۴۲۶۸
۴	۵۹۵۲	۳۰۷۵۱
۵	۶۴۴۷	۳۷۶۹۳
۶	۷۰۳۰	۴۵۰۸۷
۷	۷۶۹۲	۵۳۵۳۳
۸	۸۵۹۰	۶۳۵۷۸
۹	بزرگتر از ۸۵۹۱	بزرگتر از ۶۳۵۷۹

پس از کلاسه‌بندی و بررسی بدست آمده از مدلهای مختلف به دلیل محدوده بزرگ دهک‌های مشخص شده، داده‌ها براساس سهمیه‌نهایی تفکیک و مجددا کلاسه‌بندی شدند که نتایج آن در ادامه آورده می‌شود:

در این پژوهش براي يافتن ارتباط ميان رتبه در سهميه‌نهايی داوطلب با نمره‌هاي سوابق ديپلم و پيش دانشگاهی آنان از الگوريتم C۵,۰ با تعداد ۲۱ متغير به عنوان ورودي و يك متغير رتبه در سهميه‌نهايی به عنوان خروجي استفاده گرديد.

نتايج حاصل از اين الگوريتم در ادامه ارائه شده است.

شكل ۱ نشان مي‌دهد كه ۷ متغيرهای تراز نمرات شيمی ۳ و ازمايشگاه، زيست شناسی ۲ و آزمايشگاه، فيزيک پيش دانشگاهی،



شكل ۱: نمودار تاثير متغيرها در وضعيت رتبه در سهميه‌نهايی داوطلبان با درخت تصميم

زبان خارجی ۳، رياضی ۳، زيست شناسی پيش دانشگاهی و تعليمات ديني و قرآن ۳ بيشترين تاثير را در متغير هدف ما يعني رتبه در سهميه‌نهايی داوطلب، دارد. با بررسی ميزان تاثير هريک از متغيرهای گفته در بالا، در بهترين حالت بررسی شده به عنوان مدل، نمرات

جدول ۲: متغیرهاي ورودي سوابق تحصيلي در مدل سازي

ردیف	نام فیلد ورودی	شرح فیلد
۱	Tamp۰۱	تراز درس ۰۱ دیپلم (تعلیمات دینی و قرآن ۳)
۲	Tamp۰۲	تراز درس ۰۲ دیپلم (زبان فارسی ۳)
۳	Tamp۰۳	تراز درس ۰۳ دیپلم (ا دبیات فارسی ۳)
۴	Tamp۰۴	تراز درس ۰۴ دیپلم(عربی ۳)
۵	Tamp۰۵	تراز درس ۰۵ دیپلم(زبان خارجی ۳)
۶	Tamp۰۶	تراز درس ۰۶ دیپلم(فیزیک ۳ و آزمایشگاه)
۷	Tamp۰۷	تراز درس ۰۷ دیپلم (شیمی ۳ و آزمایشگاه)
۸	Tamp۰۸	تراز درس ۰۸ دیپلم(ریاضی ۳)
۹	Tamp۰۹	تراز درس ۰۹ دیپلم (زیست شناسی ۲ و آزمایشگاه)
۱۰	Tamp۱۰	تراز درس ۱۰ دیپلم (زمین شناسی)
۱۱	TarazPish۰۱	تراز درس ۰۱ پیش دانشگاهی (زبان فارسی)
۱۲	TarazPish۰۲	تراز درس ۰۲ پیش دانشگاهی(معارف اسلامی)
۱۳	TarazPish۰۳	تراز درس ۰۳ پیش دانشگاهی(زیست شناسی)
۱۴	TarazPish۰۴	تراز درس ۰۴ پیش دانشگاهی(فیزیک)
۱۵	JNC	جنسیت
۱۶	BDIP	کد بخش پیش‌دانشگاهی
۱۷	BDIP۱	کد بخش دیپلم
۱۸	BDIP۲	کد بخش ماقبل از دیپلم
۱۹	BTVL	کد بخش تولد
۲۰	CHAP	چپ دست
۲۱	SAHMN	سهمیه‌نهایی
۲۲	Age	سن

تراز سوابق دیپلم و پیش دانشگاهی بیشترین تاثیر را در پیش بینی رتبه در سهمیه داوطلبان داشته است.
به طور مثال يكي از قوانين حاصل از درخت تصميم نشان مي‌دهد داوطلباني كه نمرات تراز درس شيمي ۳ و آزمايشگاه آنان كوچكتر يا مساوی ۶۱۹۲ و نمره تراز درس زیست شناسی ۲ و آزمایشگاه آنان کوچکتر یا مساوی ۵۱۲۹ و نمره تراز فيزيك ۳ و آزمايشگاه آنان كوچكتر يا مساوی از ۴۲۳۷ و نمره تراز درس زبان فارسی ۳ آنان کوچکتر یا مساوی ۴۴۸۳ و نمره تراز درس معارف اسلامی پیش دانشگاهی آنان کوچکتر یا مساوی ۲۱۹۲ باشد به میزان ۷۵/۳۱ درصد رتبه در سهمیه آنان کمتر از ۶۲۵۷۴ خواهد بود.

۴ نتیجه گیری

با توجه به نتایج بدست آمده با تفکیک داده‌های مناطق ۱، ۲ و ۳ در سالهای ۹۵ و ۹۶ ، ابتدا دادهها براساس فیلد سهمیه‌نهایی کلاس‌بندی‌شده و سپس رتبه در سهمیه‌نهایی داوطلبان پیش بینی گردید. با تعیین میزان اختلاف رتبه پیش بینی شده با رتبه‌واقعی کسب شده در سهمیه‌نهایی توسط این داوطلبان، به نتایج زیر دست یافتیم. نتایج بدست آمده از الگوریتم C&R نشان داد که اختلاف یاد شده شامل بازه نا متقارن صفر تا صد هزار بوده و در الگوریتم QUEST نیز حدود ۲۵ درصد از رتبه های پیش بینی شده با رتبه های واقعی کسب شده مطابقت داشت که نشان دهنده مناسب نبودن مدل و وجود درصد خطای بالا است. اما در الگوریتم C۵ حدود ۸۶/۴ درصد پیش بینی‌های انجام شده با واقعیت موجود اختلاف کلاس کمتر از ۱ رتبه داشت.

مراجع

حج فروش، ا. (۱۳۸۱)، آسیب‌های کنکور بر هدف‌های دورهٔ متوسطه و ارائه پیشنهاد براي رفع آسیب‌ها، *مجموعه مقالات سمینار بررسی روش‌ها و مسائل آزمون‌های ورودی دانشگاه‌ها*، اصفهان.

Bigus J.P. (1996), *Data Mining with Neural Networks: Solving Business Problems from Application Development to Decision Support*, McGraw-Hill, New York.

Osmanbegović E. and Suljić M. (2012), Data mining approach for predicting student performance, *Economic Review-Journal of Economics and Business* 10, 3-12.

رگرسیون استوار فازی با استفاده از جریمه‌ی لاسو در حضور متغیرهای پاسخ، توضیحی و ضرایب فازی

وحید گودرزی[1]، محمدرضا ربیعی، محمد آرشی

شاهرود-دانشگاه صنعتی شاهرود-دانشکده علوم ریاضی-گروه آمار-کدپستی ۳۶۱۹۹۹۵۱۸۱-۳۱۶

چکیده: در مسائل کاربردی که مولفه‌های نادقیق اجتناب ناپذیر هستند، برای مدل‌های رگرسیونی فازی با ورودی، خروجی و ضرایب فازی که ابزارهای کلاسیک نمی‌توانند معیار مناسبی باشند، برای برطرف سازی برخی از محدودیت‌های فعلی از مجموعه‌های فازی استفاده می‌کنیم. از طرفی اگر بین متغیرهای توضیحی فازی همخطی وجود داشته باشد، روش رگرسیون فازی معمولی برای برآورد ضرایب کارا نخواهد بود. در این مقاله، با ارائه‌ی یک روش پیشنهادی فازی با استفاده از جریمه‌ی لاسو و متر قدرمطلقی فازی که در مقابل داده‌های‌پرت نیز استوار است، تقریبی برای برآورد ضرایب رگرسیونی به دست آورده و پارامترهای مدل رگرسیونی را هنگامی که متغیرهای ورودی، خروجی و ضرایب فازی مقدار هستند، برآورد می‌کنیم. کاربرد این روش را در داده‌های سیمان پورتلند مورد ارزیابی قرار می‌دهیم. اگرچه در رگرسیون کلاسیک جریمه‌ی لاسو در شرایط داده‌های همخط ناکاراست، اما روش پیشنهادی در داده‌های مورد استفاده که دارای همخطی بالایی می‌باشند، کارایی قابل توجهی نسبت به روش‌های کمترین توان‌های دوم و استوار فازی دارد.

واژه‌های کلیدی: رگرسیون استوار فازی، رگرسیون لاسو فازی، متر قدر مطلقی، همخطی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 62J07، 62J86.

---

۱ مقدمه

در رگرسیون خطی کلاسیک در صورت عدم وجود همخطی، برآوردگر رگرسیونی معمولی کمترین توان‌های دوم، بهترین برآوردگر خطی نااریب (BLUE[1]) است. در صورت وجود همخطی بین متغیرهای پیشگو یکی از موثرترین روش‌های مقابله با آن رگرسیون ریج معرفی شده توسط هورل و کنارد (۱۹۷۰) است. همچنین برای کاهش بعد متغیرهای پیشگو (انتخاب متغیر) تیبشیرانی (۱۹۹۶) روش جریمه‌ی لاسو را پیشنهاد کرد.

تحلیل رگرسیونی فازی یک روش قدرتمند برای بررسی ارتباط بین پدیده‌هایی است که در آن‌ها متغیر خروجی (پاسخ) به متغیر یا متغیرهای ورودی (توضیحی)، در محیط فازی، وابسته است. مدل رگرسیون خطی فازی اولین بار توسط تاناکا و همکاران (۱۹۸۲) معرفی شد. در زمینه رگرسیون فازی مطالعات زیادی صورت گرفته است که از آن جمله می‌توان به ایشیبوشی و تاناکا (۱۹۹۲) اشاره کرد که مدل‌های رگرسیون فازی غیرخطی را با توجه به کاربرد آن در شبکه‌های عصبی پیشنهاد دادند، دیاموند (۱۹۸۸) نیز روش دیگری را بر پایه‌ی کمترین توان‌های دوم ارائه کرد. برای اولین بار هانگ و همکاران (۲۰۰۴) الگوریتم یادگیری رگرسیون ریج را برای مدل‌های رگرسیون فازی با مقادیر ورودی دقیق و مقادیر خروجی اعداد فازی نرمال تعمیم دادند، همچنین دونسو وهمکاران (۲۰۰۷) رگرسیون

[1] نام ارائه دهنده‌ی مقاله : v1992gi@gmail.com

[1] Best Linear Unbiased Estimator

ریج فازی را برای مدل‌های درجه دوم پیشنهاد دادند، اخیراً نیز ربیعی و همکاران (۱۳۹۶) مدلی را براساس رگرسیون ریج فازی برای تنها ضرایب و خروجی فازی و در ادامه ربیعی و همکاران (۲۰۱۸) رگرسیون ریج فازی را برای متغیرهای ورودی و خروجی و ضرایب فازی بررسی نمودند، گودرزی و همکاران (۱۳۹۷) نیز نسخه استوار از این برآوردگرها ارائه کردند.

در این‌جا مدل‌بندی رگرسیون لاسو در محیطی فازی مورد مطالعه قرار خواهد گرفت و در آن از متر حسن‌پور و همکاران (۲۰۱۰) استفاده خواهد شد که نسبت به داده‌های دورافتاده نیز استوار است. در بخش ۲ به تعاریف و مفاهیم اولیه پرداخته و در ادامه مدل رگرسیون لاسو فازی پیشنهادی را معرفی کرده و شاخص‌های ارزیابی مدل را تعریف خواهیم کرد. سپس با استفاده از داده‌های واقعی سیمان پورتلند، مدل پیشنهادی را بررسی خواهیم کرد. بخش آخر بحث و نتیجه گیری خواهد بود.

۲ تعاریف و مفاهیم فازی

ابتدا به تعریف مفاهیم فازی می‌پردازیم و سپس با استفاده از خواص برآوردگر لاسو، برآوردگر پیشنهادی را ارائه می‌دهیم.

مجموعه فازی $\tilde{A}$ از مجموعه مرجع $\mathcal{X}$ به صورت $\mathcal{F} = \{(x, A(x))|x \in \mathcal{X}\}$ که در آن $A(x) : \mathcal{X} \to [۰, ۱]$ تابع عضویت x در $\tilde{A}$ است. گوییم مجموعه فازی $\tilde{A}$ زیر مجموعه‌ی فازی $\tilde{B}$ است، اگر برای هر $x \in \mathcal{X}$، $A(x) \leq B(x)$ یعنی

$$\tilde{A} \subseteq \tilde{B} \Leftrightarrow A(x) \leq B(x), \forall x \in \mathcal{X}.$$

تعریف ۱.۲. *اگر عدد فازی $\tilde{A}$ دارای تابع عضویت زیر باشد*

$$A(x) = \begin{cases} L\left(\frac{m-x}{l}\right) & x \leq m \\ R\left(\frac{x-m}{r}\right) & x > m \end{cases} \tag{۱.۲}$$

که در آن L و R توابعی غیرصعودی از $\mathbb{R}^+$ به $[۰, ۱]$ هستند و $L(۰) = R(۰) = ۱$، آنگاه $\tilde{A}$ را یک عدد فازی LR نامیده و با نماد $\tilde{A} = (m, l, r)_{LR}$ نشان می‌دهیم. عدد m را مقدار نما(میانه) و اعداد مثبت l و r را به ترتیب پهنای چپ و پهنای راست $\tilde{A}$ می‌نامیم. L و R نیز توابع مرجع(یا توابع شکل) نامیده می‌شوند. در حالت خاص اگر $L(x) = R(x) = ۱ - |x|$ باشد، عدد فازی به‌دست آمده را مثلثی نامند که با $A(x) = (m, l, r)_T$ نشان می‌دهند. $T(\mathbb{R})$ را مجموعه اعداد مثلثی فازی گویند.

اعمال حسابی در بین اعداد فازی بر اساس اصل گسترش قابل تعریف است، برای به‌دست آوردن اطلاعات بیشتر در این زمینه به دووا و پراد (۱۹۸۰) و طاهری و ماشین‌چی (۱۳۹۲) مراجعه نمایید. در زیر به جمع دو عدد فازی و همچنین به ضرب تقریبی آن‌ها می‌پردازیم.

قضیه ۲.۲. *(دووا و پراد، ۱۹۸۰) اگر $\tilde{A} = (m_A, l_A, r_A)_{LR}$ و $\tilde{B} = (m_B, l_B, r_B)_{LR}$ آنگاه بر اساس اصل گسترش داریم*

$$\tilde{A} \oplus \tilde{B} = (m_A + m_B, l_A + l_B, r_A + r_B)_{LR} \tag{۲.۲}$$

قضیه ۳.۲. *(دووا و پراد، ۱۹۸۰) اگر $\tilde{A} = (m_A, l_A, r_A)_{LR}$ و $\tilde{B} = (m_B, l_B, r_B)_{LR}$ آنگاه*

$$\tilde{A} \otimes \tilde{B} \simeq \begin{cases} (m_A m_B, m_A l_B + m_B l_A, m_A r_B + m_B r_A)_{LR} & \tilde{A} \succ ۰, \tilde{B} \succ ۰ \\ (m_A m_B, m_A l_B - m_B r_A, m_A r_B - m_B l_A)_{RL} & \tilde{A} \succ ۰, \tilde{B} \prec ۰ \\ (m_A m_B, -m_B r_A - m_A r_B, -m_B l_A - m_A l_B)_{RL} & \tilde{A} \prec ۰, \tilde{B} \prec ۰ \end{cases} \tag{۳.۲}$$

که در آن $\tilde{A} \succ 0$ *و* $\tilde{B} \prec 0$ *به‌ترتیب نشان‌دهنده اعداد فازی مثبت و منفی هستند.*

قضیه ۴.۲. *(دووا و پراد، ۱۹۸۰) فرض کنید* $\tilde{A}$، $\tilde{B}$ *و* $\tilde{C}$ *اعداد فازی باشند، در این صورت*

$$(\tilde{A} \oplus \tilde{B}) \otimes \tilde{C} \subseteq (\tilde{A} \otimes \tilde{C}) \oplus (\tilde{B} \otimes \tilde{C}) \tag{۴.۲}$$

و تساوی زمانی برقرار است اگر $\tilde{A}, \tilde{B}, \tilde{C} \succ 0$.

تعریف ۵.۲. *(حسن‌پور و همکاران، ۲۰۱۰) فرض کنید* $\tilde{A} = (m_A, l_A, r_A)_{LR}$ *و* $\tilde{B} = (m_B, l_B, r_B)_{LR}$ *دو عدد فازی در* $T(\mathbb{R})$ *باشند،* $d : T(\mathbb{R}) \times T(\mathbb{R}) \to \mathbb{R}^{\geq 0}$ *را به صورت زیر تعریف می‌کنیم*

$$d(\tilde{A}, \tilde{B}) = |m_A - m_B| + |l_A - l_B| + |r_A - r_B| \tag{۵.۲}$$

۳ برآوردگر لاسو

از دیدگاه آماری، لاسو مخفف عملگر انتخاب‌کننده، منقبض‌کننده و مینیمم‌کننده از جنس قدرمطلق است. برآوردگر لاسو نوع جریمه‌شده برآوردگر OLS است و از حل مسئله بهینه‌سازی زیر به‌دست می‌آید

$$\hat{\boldsymbol{\beta}}_{\text{LASSO}} = \min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right) \quad \text{s.t.} \quad \sum_{j=1}^{p} |\beta_j| \leq t \tag{۱.۳}$$

که در آن $t > 0$ یک مقدار ثابت است. اگر $t = 0$، آنگاه متغیری در مدل وجود نخواهد داشت، درحالی‌که اگر $t = \infty$، مدل کامل خواهد بود. فرض کنید $\hat{\boldsymbol{\beta}}^* = (\hat{\beta}_1^*, \ldots, \hat{\beta}_p^*)^\top$ یک برآوردگر اولیه برای $\boldsymbol{\beta}$ است که معمولا این برآوردگر را $\hat{\boldsymbol{\beta}}_{\text{OLS}}$ (برآوردگر کمترین توان‌های دوم) درنظر می‌گیرند. اگر $t > \sum_{j=1}^{p} |\hat{\beta}_j^*|$، آنگاه روش لاسو برآوردگر OLS را نتیجه می‌دهد. با انتخاب $0 < t < \sum_{j=1}^{p} |\hat{\beta}_j^{\text{OLS}}|$، مسأله‌ی بهینه‌سازی (۱.۳) معادل است با

$$\hat{\boldsymbol{\beta}}_{\text{LASSO}} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^p} \left((\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + k \sum_{j=1}^{p} |\beta_j| \right)$$

که در آن k پارامتر تنظیم‌کننده است و سطح تنکی (تعداد پارامترهای صفر) را در برآوردگر لاسو کنترل می‌کند. وقتی $k \to 0$ برآوردگر لاسو به برآوردگر OLS میل می‌کند و وقتی $k \to \infty$، برآوردگر لاسو به سمت صفر میل می‌کند. به‌عبارت دیگر، هرچه جریمه بزرگتری به‌کار برده شود، تعداد بیشتری از ضرایب به‌سمت صفر منقبض می‌شوند. در حقیقت k و t رفتار یا رابطه‌ای معکوس با یکدیگر دارند. از این‌رو برآوردگر لاسو می‌تواند متغیرهای اضافی را از مدل حذف نماید. در واقع برآوردگر لاسو، هم‌زمان هم انتخاب متغیر انجام می‌دهد و هم ضرایب را منقبض می‌کند. همچنین به‌دلیل تابع جریمه‌ی قدرمطلق، جواب صریحی برای برآوردگر لاسو وجود ندارد. تیبشیرانی (۱۹۹۶) این برآوردگر را از طریق برنامه‌ریزی درجه دوم به‌دست آورده بود.

۴ شاخص‌های ارزیابی مدل

برای ارزیابی عملکرد مدل رگرسیونی فازی، کیم و بیشوا (۱۹۹۸) معیار زیر را به عنوان اختلاف بین دو تابع عضویت $Y_i(x)$ و $y_i(x)$(iامین برآورد $Y_i(x)$) را معرفی کردند که در آن $S_{\tilde{Y}_i}$ و $S_{\tilde{y}_i}$ به ترتیب تکیه‌گاه‌های اعداد فازی مثلثی $\tilde{Y}_i$ و $\tilde{y}_i$ هستند،

$$E_i = \int_{S_{\tilde{Y}_i} \cup S_{\tilde{y}_i}} \left| Y_i(x) - y_i(x) \right| \mathrm{d}x, \quad i = ۱, ۲, \ldots, n \tag{۱.۴}$$

و $E = \sum_{i=۱}^{n} E_i$، مقدار کل خطای برآورد است.

همچنین برای ارزیابی عملکرد مدل رگرسیونی فازی، حسن‌پور و همکاران (۲۰۰۹، ۲۰۱۰) معیار دیگری بنام λ را معرفی کردند که اختلاف بین مراکز مشاهدات و پاسخ‌های برآورد شده را به صورت $\lambda = \sum_{i=۱}^{n} \lambda_i$ محاسبه می‌کنند که در آن $\lambda_i = \left| m_{y_i} - \sum_{j=۰}^{p} m_{A_j} m_{X_{ij}} \right|$ است.

۵ برآوردگر استوار فازی بر پایه فاصله قدرمطلقی

اکنون حالتی را درنظر بگیرید که مجموعه داده‌های فازی به‌صورت زیر باشند

$$\left\{ \left(\tilde{x}_{i۱}, \tilde{x}_{i۲}, \ldots, \tilde{x}_{ip}, \tilde{y}_i \right) \middle| i = ۱, \ldots, n \right\}$$

که در آن $\tilde{x}_{ij}, \tilde{y}_i \in T(\mathbb{R})$ و $\tilde{x}_{ij} \succ ۰$ ، $i = ۱, \ldots, n$ و $j = ۱, \ldots, p$. در این صورت مدل رگرسیونی خطی فازی به صورت زیر تبدیل می‌شود

$$\tilde{Y}_i = \tilde{A}_۰ \oplus (\tilde{A}_۱ \otimes \tilde{x}_{۱i}) \oplus \cdots \oplus (\tilde{A}_p \otimes \tilde{x}_{pi}), \quad i = ۱, \ldots, n \tag{۱.۵}$$

از آنجا که برای برآورد ضرایب رگرسیونی، محاسبه ضرب دو عدد فازی مثلثی $\tilde{A}_j$ و $\tilde{x}_j$ اجتناب‌ناپذیر است، بنابراین طبق قضیه‌ی ۳.۲ این ضرب جواب دقیق نداشته و تقریب آن نیز به علامت اعداد فازی مثلثی وابسته است. در ریاضیات کلاسیک می‌دانیم که هر عدد حقیقی مانند a را می‌توان به‌صورت $a = a' - a''$ نوشت که در آن a' و a'' دو عدد حقیقی مثبت‌اند. با استفاده از این ترفند، می‌توان ضرب‌های معادله‌ی (۱.۵) را با استفاده از قضیه‌های ۳.۲ و ۴.۲ به‌دست آورد.

لم ۱.۵. *در مدل (۱.۵) فرض کنید* $\tilde{X}_{ij} = (m_{X_{ij}}, l_{X_{ij}}, r_{X_{ij}})_T$ *و* $\tilde{Y}_i = (m_{Y_i}, l_{Y_i}, r_{Y_i})_T$ *به ازای* $i = ۱, \ldots, n$ *و* $j = ۱, \ldots, p$ *و* $\tilde{A}_۰ = (m_{A_۰}, l_{A_۰}, r_{A_۰})_T$ *باشند. اگر* $\tilde{A}_j = \tilde{A}'_j \oplus \tilde{A}''_j$ *که در آن* $\tilde{A}'_j, \tilde{A}''_j \in T(\mathbb{R})$ *و* $\tilde{A}'_j = (m_{A'_j}, l_{A'_j}, r_{A'_j})_T, \tilde{A}''_j = (-m_{A''_j}, l_{A''_j}, r_{A''_j})_T$ *، در این صورت داریم* $(j = ۱, \ldots, p)$،

$$\begin{aligned} \tilde{Y}_i \subseteq \Bigg(& m_{A_۰} + \sum_{j=۱}^{p} (m'_{A_j} - m''_{A_j}) m_{X_{ij}}, l_{A_۰} + \sum_{j=۱}^{p} (m'_{A_j} l_{X_{ij}} + l'_{A_j} m_{X_{ij}} + m''_{A_j} r_{X_{ij}} + l''_{A_j} m_{X_{ij}}) \\ & , r_{A_۰} + \sum_{j=۱}^{p} (m'_{A_j} r_{X_{ij}} + r'_{A_j} m_{X_{ij}} + m''_{A_j} l_{X_{ij}} + r''_{A_j} m_{X_{ij}}) \Bigg) \end{aligned} \tag{۲.۵}$$

طرف راست معادله‌ی (۲.۵) را با $\tilde{Y}_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}'')$ نشان می‌دهیم، که در آن $\tilde{\mathbf{A}}'$ و $\tilde{\mathbf{A}}''$ بردارهای $p-$بعدی هستند و به ترتیب $\tilde{A}'_j$ و $\tilde{A}''_j$، j امین عضو می‌باشد. همانطور که می‌دانیم $\tilde{Y}_i$ لزوما یک عدد فازی مثلثی نیست، اگرچه برای هر انتخاب $\tilde{\mathbf{A}}'$ و $\tilde{\mathbf{A}}''$،

$\tilde{Y}_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}'')$ یک برآورد مثلثی از $\tilde{Y}_i$ است که مستقل از علامت ضرایب رگرسیونی می‌باشد.

برای دستیابی به بهترین انتخاب، تابع زیر را با روش‌های عددی به کمک نرم‌افزار Mathematica 11 کمینه می‌کنیم

$$\sum_{i=1}^{n} d\big(\tilde{Y}_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}''), \tilde{Y}_i\big). \tag{۳.۵}$$

به شرط آن که $m'_A \geq l'_A$ و $m''_A \geq r''_A$ و همچنین $m'_A, m''_A, l'_A, l''_A, r'_A, r''_A \geq 0$ ، با استفاده از متر حسن‌پور و همکاران (۲۰۱۰) داریم

$$\sum_{i=1}^{n} d\big(\tilde{Y}_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}''), \tilde{Y}_i\big) = \sum_{i=1}^{n} \Big(\big|m_{Y_i} - m_{Y_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}'')}\big| + \big|l_{Y_i} - l_{Y_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}'')}\big| + \big|r_{Y_i} - r_{Y_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}'')}\big|\Big). \tag{۴.۵}$$

حسن‌پور و همکاران (۲۰۱۰) نشان دادند که این برآوردگر نسبت به حضور داده دورافتاده در مجموعه داده‌ها استوار است.

زمانیکه همخطی بین متغیرهای توضیحی وجود دارد، معیاری برای اندازه‌گیری این همخطی، شاخص عامل تورم واریانس تعمیم‌یافته است که به‌صورت زیر تعریف می‌شود.

تعریف ۲.۵. *(فرخی، ۱۳۹۵) برای شناسایی همخطی بین متغیرهای ورودی فازی از شاخص VIF تعمیم‌یافته که به‌صورت زیر تعریف می‌شود، استفاده می‌کنیم*

$$\mathrm{VIF}_{\mathrm{G_j}} = \frac{1}{1 - R^2_{G_j}}, \quad j = 1, 2, \ldots, p-1 \tag{۵.۵}$$

که در آن $R^2_{G_j} = \frac{\sum_{i=1}^{n} d^2(\hat{\tilde{X}}_{ij}, \bar{\tilde{X}}_j)}{\sum_{i=1}^{n} d^2(\tilde{X}_{ij}, \bar{\tilde{X}}_j)}$ *ضریب تعیین تعمیم‌یافته است و* $\hat{\tilde{X}}_{ij}$ *مقادیر رگرسیون‌شده متغیر* $\tilde{X}_j$ *روی دیگر متغیرهای توضیحی است. اگر* $\tilde{X}_{ij} = (m_{X_{ij}}, l_{X_{ij}}, r_{X_{ij}})_T$ *به ازای* $i = 1, \ldots, n$ *و* $j = 1, \ldots, p$*، آن‌گاه* $\bar{\tilde{X}}_j$ *از رابطه‌ی زیر به‌دست می‌آید*

$$\bar{\tilde{X}}_j = \left(\frac{\sum_{i=1}^{n} m_{X_{ij}}}{n}, \frac{\sum_{i=1}^{n} l_{X_{ij}}}{n}, \frac{\sum_{i=1}^{n} r_{X_{ij}}}{n}\right)$$

۶ مدل رگرسیون فازی پیشنهادی از دیدگاه رگرسیون لاسو کلاسیک

در این بخش سعی در ارائه برآوردگری داریم که بتواند در حضور داده‌های پرت و وجود همخطی میان متغیرهای توضیحی، زمانی که متغیر پاسخ، توضیحی و ضرایب فازی هستند، کارایی مطلوب را داشته باشد. در این برآوردگر استوار از جریمه لاسو برای مقابله با مشکل همخطی استفاده می‌نماییم.

اکنون مدل رگرسیون فازی ۱.۵ را درنظر بگیرید، برای برآورد ضرایب مدل زمانی که همخطی بین متغیرهای توضیحی وجود دارد و در حضور داده‌های پرت، تابع هدف زیر را با استفاده از جریمه لاسو معرفی می‌کنیم

$$\sum_{i=1}^{n} \mathrm{d}\big(\tilde{Y}_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}''), \tilde{Y}_i\big) + k||\tilde{\mathbf{A}}|| \tag{۱.۶}$$

که در آن k پارامتر تنظیم‌کننده لاسو، $\tilde{Y}_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}'')$ طرف راست معادله‌ی (۲.۵) خواهد بود و اگر $\tilde{0} = (0, 0, 0)_T$ باشد، $||\tilde{A}||$ را به صورت زیر تعریف می‌کنیم

$$||\tilde{A}|| = \mathrm{d}(\tilde{A}, \tilde{0}) = |m_A - 0| + |l_A - 0| + |r_A - 0| \tag{۲.۶}$$

با استفاده از معادله‌ی (۲.۶) و توضیحات بخش ۵ تعریف می‌کنیم

$$\begin{aligned}||\tilde{\mathbf{A}}|| &= \sum_{j=1}^{p}||\tilde{A}_j|| = \sum_{j=1}^{p}||(m'_{A_j} - m''_{A_j}, l'_{A_j} + l''_{A_j}, r'_{A_j} + r''_{A_j})|| \\ &= \sum_{j=1}^{p}(|m'_{A_j} - m''_{A_j}| + |l'_{A_j} + l''_{A_j}| + |r'_{A_j} + r''_{A_j}|)\end{aligned} \tag{۳.۶}$$

اگر داشته باشیم $\tilde{A}_0 = (m_{A_0}, l_{A_0}, r_{A_0})_T$ با استفاده از تعاریف بالا می‌توان (۱.۶) را به صورت زیر بازنویسی کرد

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{d}\big(\tilde{Y}_i(\tilde{\mathbf{A}}', \tilde{\mathbf{A}}''), \tilde{Y}_i\big) + k||\tilde{\mathbf{A}}|| = \sum_{i=1}^{n} \Bigg[&\Big| m_{Y_i} - \Big(m_{A_0} + \sum_{j=1}^{p}(m'_{A_j} - m''_{A_j})m_{X_{ij}}\Big)\Big| \\ &+ \Big| l_{Y_i} - \Big(l_{A_0} + \sum_{j=1}^{p}(m'_{A_j} l_{X_{ij}} + l'_{A_j} m_{X_{ij}} + m''_{A_j} r_{X_{ij}} + l''_{A_j} m_{X_{ij}})\Big)\Big| \\ &+ \Big| r_{Y_i} - \Big(r_{A_0} + \sum_{j=1}^{p}(m'_{A_j} r_{X_{ij}} + r'_{A_j} m_{X_{ij}} + m''_{A_j} l_{X_{ij}} + r''_{A_j} m_{X_{ij}})\Big)\Big| \Bigg] \\ &+ k\Bigg[|m_{A_0}| + |l_{A_0}| + |r_{A_0}| + \sum_{j=1}^{p}\Big(|m'_{A_j} - m''_{A_j}| + |l'_{A_j} + l''_{A_j}| + |r'_{A_j} + r''_{A_j}|\Big)\Bigg]\end{aligned} \tag{۴.۶}$$

با کمینه کردن رابطه‌ی (۴.۶) تحت قیود زیر

$$m'_{A_j} \geq l'_{A_j}, m''_{A_j} \geq r''_{A_j}, \quad l_{A_0}, r_{A_0}, m'_{A_j}, m''_{A_j}, l'_{A_j}, l''_{A_j}, r'_{A_j}, r''_{A_j} \geq 0, \quad j = 1, 2, \ldots, p \tag{۵.۶}$$

برآورد ضرایب مدل رگرسیونی محاسبه می‌گردد. برای کمینه کردن این معادله از نرم افزار Mathematica 11 استفاده می‌کنیم.

۱.۶ مثال کاربردی

داده‌های جدول ۱ مربوط به سیمان پُرتلند هستند که در وودز و همکاران (۱۹۳۲) آورده شده است. این داده‌ها از مطالعه تاثیر ترکیبات مختلف روی حرارت بیرون آمده در طول سخت شدن سیمان پورتلند به‌دست می‌آید. در اینجا داده‌هایی که از چهار ترکیب و میزان گرمای بیرون آمده بعد از ۸۰ روز به‌دست آمده‌اند را بررسی می‌کنیم. این چهار ترکیب، تری‌سدیم آلومینات 3CaO.Al2O3، تری‌کلسیم سیلیکات 3CaO.SiO2، تتراکلسیوم آلمینوفریت 4CaO.Al2O3.Fe2O3 و دی‌کلسیوم سیلیکات 2CaO.SiO2 هستند که به ترتیب X_1، X_2، X_3 و X_4 و گرمای بیرون آمده Y، تعریف می‌شوند.

جدول ۱: داده‌های فازی شده سیمان پُرتلند

i	$\tilde{X}_1$	$\tilde{X}_2$	$\tilde{X}_3$	$\tilde{X}_4$	$\tilde{Y}$
۱	(۷, ۰/۷۷, ۰/۹۱)	(۲۶, ۲/۸۶, ۳/۳۸)	(۶, ۰/۶۶, ۰/۷۸)	(۶۰, ۶/۶, ۷/۸)	(۷۸/۵, ۸/۶۳۵, ۱۰/۲۰۵)
⋮	⋮	⋮	⋮	⋮	⋮
۱۲	(۱۱, ۱/۵۴, ۱/۳۲)	(۶۶, ۹/۲۴, ۷/۹۲)	(۹, ۱/۲۶, ۱/۰۸)	(۱۲, ۱/۶۸, ۱/۴۴)	(۱۱۳/۳, ۱۵/۸۶۲, ۱۳/۵۹۶)
۱۳	(۱۰, ۱/۷, ۱/۱)	(۶۸, ۱۱/۵۶, ۷/۴۸)	(۸, ۱/۳۶, ۰/۸۸)	(۱۲, ۲/۰۴, ۱/۳۲)	(۱۰۹/۴, ۱۸/۵۹, ۱۲/۰۳)

با توجه به برخی ابهامات در اندازه‌گیری‌های مربوط به متغیرهای ورودی و مشاهدات متغیر پاسخ، با مشورت با یک فرد متخصص با استفاده از یک عدد فازی مثلثی با پهناهای راست و چپ متناسب برای مراکز $\tilde{X}_1$، $\tilde{X}_2$، $\tilde{X}_3$ و $\tilde{X}_4$ و همچنین $\tilde{Y}$، فازی سازی شده‌اند

(فرخی، ۱۳۹۵). از آن‌جا که متغیرهای ورودی فازی هستند، از معیار $\mathrm{VIF_G}$ برای تشخیص همخطی بین متغیرها استفاده می‌کنیم. در این صورت

$$VIF_{G_۱} = ۳/۷۰۵, \quad VIF_{G_۲} = ۴۱/۴۲۶۳, \quad VIF_{G_۳} = ۴/۲۲۲۵ \quad VIF_{G_۴} = ۱۵/۸۴۷$$

که همخطی بین متغیرهارا نشان می‌دهد $(\exists i = ۱, ۲, ۳, ۴, \ VIF_{G_i} \geq ۵)$. لذا باید از روش رگرسیون استوار لاسو فازی پیشنهادی استفاده کرد. با استفاده از برآوردگر لاسو استوار فازی و انتخاب پارامتر $k = ۲/۵$، با مینیمم کردن رابطه‌ی (۴.۶) تحت قیود (۵.۶) با استفاده از نرم‌افزار Mathematica 11 مدل رگرسیون لاسو فازی به‌صورت زیر بدست می‌آید

$$\tilde{\mathbf{Y}} = (۰/۰۰۰۰۴, ۰, ۰) + (۲/۲۰, ۰, ۰)\tilde{\mathbf{X}}_۱ + (۱/۱۵, ۰, ۰)\tilde{\mathbf{X}}_۲ + (۰/۸۳, ۰, ۰)\tilde{\mathbf{X}}_۳ + (۰/۴۶۹, ۰, ۰)\tilde{\mathbf{X}}_۴ \qquad (۶.۶)$$

همچنین شاخص‌های ارزیابی مدل را برای این مسئله به‌دست می‌آوریم و نتایج به‌دست آمده برای مدل رگرسیون کمترین توان‌های دوم معمولی فازی، رگرسیون کمترین توان‌های دوم استوار فازی و رگرسیون لاسو فازی استوار را در جدول ۲ خلاصه می‌کنیم.

جدول ۲: شاخص‌های ارزیابی مدل برای رگرسیون لاسو فازی برای مجموعه داده سیمان پُرتلند

مدل رگرسیونی فازی	k	شاخص E	شاخص λ
OLS	۰	۳۷/۷۳	۱۹/۸۲
کمترین توان‌های دوم استوار	۰	۳۶/۳۷	۱۹/۱۴
روش پیشنهادی	۲/۵	۳۵/۱۲	۱۸/۶۲

بحث و نتیجه‌گیری

همان‌طور که در جدول ۲ نشان داده شد، اگر مدل رگرسیون کمترین توان‌های دوم استوار فازی نسبت به برآوردگر کمترین توان‌های دوم معمولی فازی عملکرد بهتری را نشان می‌دهند، به این دلیل است که در داده‌های سیمان پورتلند بزرگی بعضی مقادیر ممکن است در برآورد ضرایب تاثیر گذار بوده و بعنوان داده‌پرت تلقی شوند. اما برآوردگر لاسو استوار فازی، با بهبود دادن برآوردگر استوار فازی، در حضور همخطی بین متغیرهای فازی، عملکرد بهتری را نشان می‌دهد.

مراجع

ربیعی، م.ر.، فرخی، م. و آرشی م. (۱۳۹۶)، رگرسیون ریج فازی در حضور داده‌های پرت برای ضرایب و خروجی فازی، *هفتمین سمینار آمار و احتمال فازی دانشگاه بیرجند*، ۷۲-۶۴.

طاهری، م. و ماشین‌چی، م. (۱۳۹۲)، *مقدمه‌ای بر احتمال و آمار فازی*، چاپ دوم، انتشارات دانشگاه شهید باهنر کرمان، کرمان.

فرخی، م. (۱۳۹۵)، *مطالعه چندین روش رگرسیون ریج در محیط فازی*، پایان نامه کارشناسی ارشد، دانشگاه صنعتی شاهرود.

گودرزی، و.، ربیعی، م.ر. و آرشی، م. (۱۳۹۷)، رگرسیون ریج استوار با متغیرهای پاسخ، توضیحی و ضرایب فازی، *چهاردهمین کنفرانس آمار ایران*، شاهرود، ایران.

Diamond P. (1988), Fuzzy least squares, *Information Sciences* 46, 141-157.

Donso S., Marin N. and Vila M. (2007), Fuzzy ridge regression with non symmetric membership functions and quadratic models, *International Conference on Intelligent Data Engineering and Automated Learning*, Springer, 135-144.

Dubois D.J. and Prade, H. (1980), *Fuzzy Sets and Systems: Theory and Applications*, Academic Press, New York.

Hassanpour H., Maleki H.R. and Yaghoobi M.A. (2009), A note on evaluation of fuzzy linear regression models by comparing membership functions, *Iranian Journal of Fuzzy Systems* 6, 1-6.

Hassanpour H., Maleki H.R. and Yaghoobi M.A. (2010), Fuzzy linear regression model with crisp coefficients: a goal programming approach, *Iranian Journal of Fuzzy Systems* 7, 19-39.

Hoerl A.E. and Kennard R.W. (1970), Ridge regression biased estimation for nonorthogonal problems, *Technometrics* 12, 55-67.

Hong D.H., Hwang C. and Ahn C. (2004), Ridge estimation for regression models with crisp inputs and gaussian fuzzy output, *Fuzzy Sets and Systems* 142, 307-319.

Ishibuchi H. and Tanaka H. (1992), Fuzzy regression analysis using neural networks, *Fuzzy Sets and Systems* 50, 257-265.

Kim B. and Bishu R.R. (1998), Evaluation of fuzzy linear regression models by comparing membership functions, *Fuzzy Sets and Systems* 100, 343-352.

Rabiei M. R., Arashi A., Goodarzi V. and Kouhsar F.P. (2018), Fuzzy ridge linear regression analysis for fuzzy input output data, *Proceeding of Seminar on Fuzzy Statistics and Probability* 8, 156-164.

Tanaka H., Uejima S. and Asai K. (1982), Linear regression analysis with fuzzy model, *IEEE Transaction System Man and Cybermatics* 12, 903-907.

Tibshirani R. (1996) Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B* 58, 267-288.

Woods H., Steinour H.H. and Starke H.R. (1932), Effect of composition of Portland cement on heat evolved during hardening, *Industrial and Engineering Chemistry* 24, 1207-1214.

9th National Seminar on
Fuzzy Statistics and Probability

Some Fixed point theorems for contractions in non-triangular non-archimedean extended fuzzy b-metric spaces

Z. Bagheri*

Department of Mathematics, Azadshahr Branch, Islamic Azad University, Azadshahr, Iran

Abstract

In this paper, we investigate the existence of fixed points under various contractive conditions in the setup of extended fuzzy b-metric spaces. Moreover, some examples are provided here to illustrate the usability of the obtained results.

Keywords: Fixed Point, Non-triangular non-archimedean spaces, Fuzzy b-metric spaces.

Mathematics Subject Classification (2010): 47H10, 54H25.

1 Introduction and preliminaries

Grabiec (1988) defined contractive mappings on a fuzzy metric space and extended fixed point theorems of Banach and Edelstein in such spaces. Successively, George and Veeramani (1994) slightly modified the notion of a fuzzy metric space introduced by Kramosil and Michálek and then defined a Hausdorff and first countable topology on it.

Definition 1.1. *A binary operation $\star : [0,1] \times [0,1] \to [0,1]$ is called a continuous t-norm if it satisfies the following assertions:*

(T1) $\star$ is commutative and associative;

(T2) $\star$ is continuous;

(T3) $a \star 1 = a$ for all $a \in [0,1]$;

(T4) $a \star b \leq c \star d$ when $a \leq c$ and $b \leq d$, with $a, b, c, d \in [0,1]$.

Four basic examples of the continuous t-norm are $a \star_1 b = min\{a, b\}$, $a \star_2 b = \frac{ab}{max\{a,b,\lambda\}}$ for $\lambda \in (0,1)$, $a \star_3 b = ab$ and $a \star_4 b = \max\{a + b - 1, 0\}$ for all $a, b \in [0,1]$.

Definition 1.2. *A 3-tuple $(X, M, *)$ is said to be a fuzzy metric space if X is an arbitrary set, $*$ is a continuous t-norm and M is a fuzzy set on $X^2 \times (0, \infty)$ satisfying the following conditions, for all $x, y, z \in X$ and $t, s > 0$,*

(i) $M(x, y, t) > 0$;

(ii) $M(x, y, t) = 1$ for all $t > 0$ if and only if $x = y$;

(iii) $M(x, y, t) = M(y, x, t)$;

*Speaker: zohrehbagheri@yahoo.com

(iv) $M(x,y,t) * M(y,z,s) \leq M(x,z,t+s)$;

(v) $M(x,y,.) : (0,\infty) \to [0,1]$ *is continuous;*

The function $M(x,y,t)$ *denotes the degree of nearness between* x *and* y *with respect to* t.

If, in the above definition, the triangular inequality (iv) is replaced by the following condition:

$$(NA) \qquad M(x,y,t) * M(y,z,s) \leq M(x,z,\max\{t,s\})$$

for all $x,y,z \in X$ and $t,s > 0$ then the triple $(X,M,*)$ is called a non-Archimedean fuzzy metric space. It is easy to check that (NA) implies (iv), that is, every non-Archimedean fuzzy metric space is itself a fuzzy metric space.

Definition 1.3. *Hussain et al. (2015) A fuzzy b-metric space is an ordered triple* $(X,B,\star)$ *such that* X *is a nonempty set,* $\star$ *is a continuous t-norm and* B *is a fuzzy set on* $X \times X \times (0,+\infty)$ *satisfying the following conditions, for all* $x,y,z \in X$ *and* $t,s > 0$*:*

(F1) $B(x,y,t) > 0$;

(F2) $B(x,y,t) = 1$ *if and only if* $x = y$;

(F3) $B(x,y,t) = B(y,x,t)$;

(F4) $B(x,y,t) \star B(y,z,s) \leq B(x,z,\lambda(t+s))$ *where* $\lambda \geq 1$;

(F5) $B(x,y,\cdot) : (0,+\infty) \to (0,1]$ *is left continuous.*

Definition 1.4. *Let* $(X,B,\star)$ *be a fuzzy b-metric space. Then*

(i) a sequence $\{x_n\}$ *converges to* $x \in X$*, if and only if* $\lim_{n\to+\infty} B(x_n,x,t) = 1$ *for all* $t > 0$;

(ii) a sequence $\{x_n\}$ *in* X *is a Cauchy sequence if and only if for all* $\epsilon \in (0,1)$ *and for all* $t > 0$*, there exists* n_0 *such that* $B(x_n,x_m,t) > 1-\epsilon$ *for all* $m,n \geq n_0$;

(iii) the fuzzy b-metric space $(X,B,\star)$ *is called complete if every Cauchy sequence in it converges to some* $x \in X$.

Definition 1.5. *Let* X *be a (nonempty) set. A function* $\widetilde{d} : X \times X \to R^+$ *is an extended b-metric if there exists a strictly increasing continuous function* $\Omega : [0,\infty) \to [0,\infty)$ *with* $\Omega^{-1}(t) \leq t \leq \Omega(t)$ *such that for all* $x,y,z \in X$*, the following conditions hold:*

(p_1) $\widetilde{d}(x,y) = 0$ *iff* $x = y$,

(p_2) $\widetilde{d}(x,y) = \widetilde{d}(y,x)$,

(p_3) $\widetilde{d}(x,z) \leq \Omega(\widetilde{d}(x,y) + \widetilde{d}(y,z))$.

In this case, the pair $(X,\widetilde{d})$ *is called an extended b-metric space.*

Proposition 1.6. *Let* (X,d) *be a b-metric space (with parameter* $\lambda \geq 1$*) and there exists a strictly increasing continuous function* $\Omega : [0,\infty) \to [0,\infty)$ *with* $\lambda\Omega^{-1}(t) \leq t \leq \Omega(\frac{t}{\lambda})$ *for all* $t > 0$ *and* $\Omega^{-1}(0) = 0 = \Omega(0)$*. Let* $M_\Omega : X \times X \times (0,\infty) \to [0,1]$ *be defined by* $M_\Omega(x,y,t) = \frac{\lambda\Omega^{-1}(t)}{\lambda\Omega^{-1}(t)+d(x,y)}$*, for all* $t > 0$*. Then* $(X,M_\Omega,\wedge,\Omega)$ *is an extended fuzzy b-metric space.* M_Ω *will be called the standard extended fuzzy b-metric.*

Example 1.7. *Let* (X,d) *be a metric space (with parameter* $\lambda \geq 1$*).*

1. *If* $\Omega(t) = e^t - 1$*, then* $\Omega^{-1}(u) = \ln(1+u)$*. So,* $M_\Omega : X \times X \times (0,\infty) \to [0,1]$ *defined by* $M_\Omega(x,y,t) = \frac{\ln(1+t)}{\ln(1+t)+d(x,y)}$ *is an extended fuzzy b-metric.*
2. *If* $\Omega(t) = te^t$*, then* $\Omega^{-1}(u) = W(u)$*, for* $u \geq 0$*, where* W *is the Lambert* W*-function (see, e.g., Dence (2013)). Let* $M_\Omega : X \times X \times (0,\infty) \to [0,1]$ *be defined by* $M_\Omega(x,y,t) = \frac{W(t)}{W(t)+d(x,y)}$*. Then* $(X,M_\Omega,\wedge,\Omega)$ *is an extended fuzzy b-metric space.*

2 Main results

2.1 Results under Geraghty-type conditions

Theorem 2.1. *Let $(X, \preceq)$ be a partially ordered set and $(X, M, \star, \Omega)$ be a complete non-Archimedean extended fuzzy b-metric space. Suppose that $T : X \to X$ be a self-mapping satisfying the following assertions:*

(i) T is an extended fuzzy Geraghty contractive type mapping;

(ii) There exists $x_0 \in X$ such that $x_0 \preceq Tx_0$,

(iii) T is continuous.

Then T has a fixed point, that is, there exists $x_0 \in X$ such that $Tx_0 = x_0$.

2.2 Fixed point results via comparison functions

Lemma 2.2. *If $\phi \in \Phi$, then the following are satisfied.*
(a) $\phi(t) < t$ for all $t > 0$;
(b) $\phi(0) = 0$.

Theorem 2.3. *Let $(X, \preceq)$ be a partially ordered set and $(X, M, \star, \Omega)$ be a complete extended fuzzy b-metric space. Suppose that $T : X \to X$ be a self-mapping satisfying the following assertions:*

(i) T is an extended fuzzy comparison contractive mapping;

(ii) There exists $x_0 \in X$ such that $x_0 \leq Tx_0$,

(iii) T is continuous.

Then T has a fixed point, that is, there exists $x_0 \in X$ such that $Tx_0 = x_0$.

2.3 Fixed point results related to JS-contractions

Theorem 2.4. *Let (X, d) be a complete metric space and let $T : X \to X$ be a given mapping. Suppose that there exist $\theta \in \Theta_0$ and $k \in (0, 1)$ such that*

$$x, y \in X, \quad d(Tx, Ty) \neq 0 \implies \theta\big(d(Tx, Ty)\big) \leq \theta\big(d(x, y)\big)^k. \tag{1}$$

Then T has a unique fixed point.

Theorem 2.5. *Let $(X, \preceq)$ be a partially ordered set and $(X, M, \star, \Omega)$ be a complete non-Archimedean extended fuzzy b-metric space. Suppose that $T : X \to X$ be a self-mapping satisfying the following assertions:*

(i) T is an extended-fuzzy JS-contraction type mapping;

(111) There exists $x_0 \in X$ such that $x_0 \preceq Tx_0$,

(iv) T is continuous.

Then f has a fixed point.

References

George A. and Veeramani P. (1994), On some results in fuzzy metric spaces, *Fuzzy Sets and Systems* 64, 395-399.

Grabiec M. (1988), Fixed points in fuzzy metric spaces, *Fuzzy Sets and Systems* 27, 385-389.

Jleli M. and Samet B. (2014), A new generalization of the Banach contraction principle, *J. Inequal. Appl.* 38, 8 pages.

Hussain N., Salimi P. and Parvaneh V. (2015), Fixed point results for various contractions in parametric and fuzzy b-metric spaces, *J. Nonlinear Sci. Appl.* 8, 719-739.

Dence T.P. (2013), A brief look into Lambert W function, *Applied Math.* 4, 887-892.

9th National Seminar on
Fuzzy Statistics and Probability

Some fixed point theorems for various contractions in parametric and fuzzy b-metric spaces

Z. Bagheri*

Department of Mathematics, Azadshahr Branch, Islamic Azad University, Azadshahr, Iran

Abstract

The notion of parametric metric spaces being a natural generalization of metric spaces was recently introduced and studied by Hussain et al. In this paper we introduce the concept of parametric b-metric space and investigate the existence of fixed points under various contractive conditions in such spaces. As applications, we derive some new fixed point results in triangular partially ordered fuzzy b-metric spaces. Moreover, some examples are provided here to illustrate the usability of the obtained results.

Keywords: Fixed Point, Contractive condition, Fuzzy b-metric spaces.

Mathematics Subject Classification (2010): 47H10, 54H25.

1 Introduction and preliminaries

Grabiec (1988) defined contractive mappings on a fuzzy metric space and extended fixed point theorems of Banach and Edelstein in such spaces. Successively, George and Veeramani (1994) slightly modified the notion of a fuzzy metric space introduced by Kramosil and Michálek and then defined a Hausdorff and first countable topology on it.

Definition 1.1. *A 3-tuple $(X, M, *)$ is said to be a fuzzy metric space if X is an arbitrary set, $*$ is a continuous t-norm and M is a fuzzy set on $X^2 \times (0, \infty)$ satisfying the following conditions, for all $x, y, z \in X$ and $t, s > 0$,*

(i) $M(x, y, t) > 0$;

(ii) $M(x, y, t) = 1$ *for all* $t > 0$ *if and only if* $x = y$;

(iii) $M(x, y, t) = M(y, x, t)$;

(iv) $M(x, y, t) * M(y, z, s) \leq M(x, z, t + s)$;

(v) $M(x, y, .) : (0, \infty) \to [0, 1]$ *is continuous;*

The function $M(x, y, t)$ denotes the degree of nearness between x and y with respect to t.

Example 1.2. *Let (X, d) be a metric space. Then $(X, M, *)$ is a fuzzy metric space, where $a \star b = ab$ for all $a, b \in [0, 1]$ and $M(x, y, t) = \frac{t}{t+d(x,y)}$ for all $x, y \in X$ and for all $t > 0$. We call this M as the standard fuzzy metric induced by the metric d. Even if we define $a \star b = min\{a, b\}$ for all $a, b \in [0, 1]$, then the triple $(X, M, *)$ will be a fuzzy metric space.*

Definition 1.3. *Let $(X, B, \star)$ be a fuzzy b-metric space. Then*

*Speaker: zohrehbagheri@yahoo.com

(i) a sequence $\{x_n\}$ *converges to* $x \in X$*, if and only if* $\lim_{n\to+\infty} B(x_n, x, t) = 1$ *for all* $t > 0$;

(ii) a sequence $\{x_n\}$ *in* X *is a Cauchy sequence if and only if for all* $\epsilon \in (0,1)$ *and for all* $t > 0$*, there exists* n_0 *such that* $B(x_n, x_m, t) > 1 - \epsilon$ *for all* $m, n \geq n_0$*;*

(iii) the fuzzy b-metric space $(X, B, \star)$ *is called complete if every Cauchy sequence in it converges to some* $x \in X$*.*

Proposition 1.4. *Let* (X, d) *be a b-metric space with coefficient* $s \geq 1$ *and let* $\rho(x, y) = \xi(d(x, y))$ *where* $\xi : [0, \infty) \to [0, \infty)$ *is a strictly increasing continuous function with* $t \leq \xi(t)$ *for* $t \geq 0$ *and* $\xi(0) = 0$*. Then* ρ *is a p-metric with* $\Omega(t) = \xi(st)$*.*

Proposition 1.5. *Let (X, d) be a b-metric space (with parameter* $\lambda \geq 1$*) and there exists a strictly increasing continuous function* $\Omega : [0, \infty) \to [0, \infty)$ *with* $\lambda\Omega^{-1}(t) \leq t \leq \Omega(\frac{t}{\lambda})$ *for all* $t > 0$ *and* $\Omega^{-1}(0) = 0 = \Omega(0)$*. Let* $M_\Omega : X \times X \times [0, \infty) \to [0, 1]$ *be defined by* $M_\Omega(x, y, t) = \frac{\lambda\Omega^{-1}(t)}{\lambda\Omega^{-1}(t)+d(x,y)}$*,* $if t > 0$ *and* $M_\Omega(x, y, t) = 0$ *if* $t = 0$*. Then* $(X, M_\Omega, \wedge, \Omega)$ *is an extended fuzzy b-metric space.* M_Ω *will be called standard extended fuzzy b-metric.*

Definition 1.6. *Let* $(X, M, \star)$ *be a fuzzy metric space. The fuzzy metric* M *is said to be triangular whenever,* $(\frac{1}{M(x,y,t)} - 1) \leq (\frac{1}{M(x,z,t)} - 1) + (\frac{1}{M(y,z,t)} - 1)$ *for allx, y, z* $\in X$ *and any* $t > 0$*. Finally, let* X *be a nonempty set. If* $(X, M, \star)$ *is a fuzzy metric space and* $(X, \preceq)$ *is a partially ordered set, then* $(X, M, \star, \preceq)$ *is called an ordered fuzzy metric space. Also,* $x, y \in X$ *are called comparable if* $x \preceq y$ *or* $y \preceq x$ *holds. Let*$(X, \preceq)$ *be a partially ordered set and* $T : X \to X$ *be a mapping.* T *is called a non decreasing mapping if*$Tx \preceq Ty$*, whenever* $x \preceq y$ *for all* $x, y \in X$*.*

2 Main results

2.1 Results under Geraghty-type conditions

Theorem 2.1. *Let* $(X, \preceq)$ *be a partially ordered set and* $(X, M, \star, \Omega)$ *be a complete non-Archimedean extended fuzzy b-metric space. Suppose that* $T : X \to X$ *be a self-mapping satisfying the following assertions:*

(i) T *is an extended fuzzy Geraghty contractive type mapping;*

(ii) There exists $x_0 \in X$ *such that* $x_0 \preceq Tx_0$*,*

(iii) T *is continuous.*

Then T *has a fixed point, that is, there exists* $x_0 \in X$ *such that* $Tx_0 = x_0$*.*

2.2 Fixed point results via comparison functions

Definition 2.2. *Let* $(X, M, \star, \Omega)$ *be a non-Archimedean extended fuzzy b-metric space. We say that* $T : X \to X$ *is an extended-fuzzy comparison contractive mapping if there exist two functions* ϕ *and* Ω *such that:*

$$\frac{1}{M(Tx, Ty, \Omega^{-2}(t))} - 1 \leq \phi(\frac{1}{M(x, y, t)} - 1)$$

for all $t > 0$ *and for all comparable elements* $x, y \in X$*.*

Theorem 2.3. *Let* $(X, \preceq)$ *be a partially ordered set and* $(X, M, \star, \Omega)$ *be a complete extended fuzzy b-metric space. Suppose that* $T : X \to X$ *be a self-mapping satisfying the following assertions:*

(i) T is an extended fuzzy comparison contractive mapping;

(ii) There exists $x_0 \in X$ such that $x_0 \preceq Tx_0$,

(iii) T is continuous.

Then T has a fixed point, that is, there exists $x_0 \in X$ such that $Tx_0 = x_0$.

2.3 Fixed point results related to JS-contractions

Theorem 2.4. *Let (X, d) be a complete metric space and let $T : X \to X$ be a given mapping. Suppose that there exist $\theta \in \Theta_0$ and $k \in (0, 1)$ such that*

$$x, y \in X, \quad d(Tx, Ty) \neq 0 \implies \theta\big(d(Tx, Ty)\big) \leq \theta\big(d(x, y)\big)^k. \tag{1}$$

Then T has a unique fixed point.

Remark 2.5. *It is clear that $f(t) = e^t$ does not belong to Θ_0, but it belongs to Θ. Other examples are $f(t) = \cosh t$, $f(t) = \frac{2\cosh t}{1+\cosh t}$, $f(t) = 1 + \ln(1 + t)$, $f(t) = \frac{2+2\ln(1+t)}{2+\ln(1+t)}$, $f(t) = e^{te^t}$ and $f(t) = \frac{2e^{te^t}}{1+e^{te^t}}$, for all $t \geq 0$.*

Theorem 2.6. *Let $(X, \preceq)$ be a partially ordered set and $(X, M, \star, \Omega)$ be a complete non-Archimedean extended fuzzy b-metric space. Suppose that $f : X \to X$ be a self-mapping satisfying the following assertions:*

(i) f is an extended-fuzzy JS-contraction type mapping;

(111) There exists $x_0 \in X$ such that $x_0 \preceq fx_0$,

(iv) f is continuous.

Then f has a fixed point.

Example 2.7. *Let $X = [0, 1.2)$ and $(X, \widetilde{d})$ be an extended b-metric space with $d(x, y) = |x - y|^2$ for all $x, y \in X$ and $\Omega(x) = e^{2x} - 1$. Then the triple $(X, M, *, \Omega)$ is an extended fuzzy b-metric space, where $a \star b = ab$ for all $a, b \in [0, 1.2]$ and $M(x, y, t) = \frac{t}{t+[d(x,y)]}$ for all $x, y \in X$ and for all $t > 0$.*

Example 2.8. *Let (X, d) be a metric space. Then the triple (X, M, d) is a extended fuzzy b-metricspace, where $a \star b = ab$ for all $a, b \in [0, 1]$ and $M(x, y, t) = \frac{t}{t+\Omega d(x,y)}$ for all$x, y \in X$ and for all$t > 0$.$(X, M, \star, \Omega)$ is a complete extended fuzzy b-metricspace equipped with the $d(x, y) = e^{|x-y|} - 1$ for all $x, y \in X$, where $\Omega(x) = e^x - 1$.*

References

Geraghty, M. (1973), On contractive mappings, *Proc. Amer. Math. Soc.* 40, 604-608.

Đukić D., Kadelburg Z. and Radenović S. (2011), Fixed points of Geraghty-type mappings in various generalized metric spaces, *Abstr. Appl. Anal.*, Article ID 561245, 13 pages.

George A. and Veeramani P. (1994), On some results in fuzzy metric spaces, *Fuzzy Sets and Systems* 64, 395-399.

Grabiec M. (1988), Fixed points in fuzzy metric spaces, *Fuzzy Sets and Systems* 27, 385-389.

9th National Seminar on
Fuzzy Statistics and Probability

Solving fully fuzzy linear programming problems with unrestricted variables

Elham Hosseinzadeh*[1] , Hassan Hassanpour[2], Javad Tayyebi[3]

[1]Department of Mathematics, Kosar university of Bojnord, Bojnord, Iran

[2]Department of Mathematics, University of Birjand, Birjand, I.R. Iran

[3]Department of Industrial Engineering, Birjand University of Technology, Birjand, Iran

Abstract

In the past few years, a growing interest has been shown in fuzzy linear programming and currently there are several methods for solving FFLP. However, due to the limitation of these methods, they cannot be applied for solving fully fuzzy linear programming (FFLP) with unrestricted fuzzy coefficients and fuzzy variables. In this paper a new efficient method for FFLP has been proposed, in order to obtain the fuzzy optimal solution with unrestricted variables and parameters. To show the efficiency of the proposed method a numerical example have been illustrated.

Keywords: Fully fuzzy linear programming problems, Triangular fuzzy numbers, Ranking function, Fuzzy optimal solution.

Mathematics Subject Classification (2010): 90C70, 90C05.

1 Introduction

Linear Programming (LP) is one of the most important techniques in operational research. Many real-world problems can be transformed into a linear programming model. Hence this model is an indispensable tool for today's applications, such as energy, transportation and manufacturing. The fuzzy linear programming problems in which all the parameters as well as the variables are represented by fuzzy numbers is known as FFLP problems. FFLP problems can be divided into two categories:

(1) FFLP problems with inequality constraints (2) FFLP problems with equality constraints.

Some authors have proposed different methods for solving FFLP problems with inequality constraints. In all these methods firstly the FFLP problem is converted into crisp linear programming problem and then the obtained crisp linear programming problem is solved to find the fuzzy optimal solution of the FFLP problems. Dehghan et al. (2006) proposed a fuzzy linear programming approach for finding the exact solution of fully fuzzy linear system (FFLS) of equations. Lotfi et al. (2009) proposed a method to obtain the approximate solution of FFLP problems. Shamooshaki et al. (2015) using L-R fuzzy numbers and ranking function established a new scheme for FFLP. Ezzati et al. (2015) using a new ordering on triangular fuzzy numbers and converting FFLP to a multiobjective linear programming (MOLP) problem, presented a new method to solve FFLP. To the best of our knowledge, till now there is no method in the literature to obtain the exact solution of FFLP problems with equality constraints.

*Speaker: e.hosseinzadeh@yahoo.com

The paper is organized as follows: A review on notions of fuzzy set theory is presented in Section 2, the proposed approach is introduced in Section 3, a numerical exampleis presented in Section 4, and Section 5 is devoted to conclusion.

2 Preliminaries

In this section, some necessary backgrounds and notions of fuzzy set theory are reviewed.

Definition 2.1. *A triangular fuzzy number (TFN), say $\tilde{A}$, is denoted by $\tilde{A} = (a, \alpha, \beta)$ where a, α, and β are the center, left, and right spread of $\tilde{A}$, respectively. a is a real number and α, $\beta > 0$. The membership function of $\tilde{A}$ is as follows:*

$$\tilde{A}(x) = \begin{cases} \frac{x-(a-\alpha)}{\alpha} & a-\alpha \le x \le a, \\ \frac{a+\beta-x}{\beta} & a \le x \le a+\beta, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The set of all TFNs is denoted by $T(\mathbb{R})$.

The following formulas for addition of two TFNs and multiplication of a TFN by a scaler are drawn from extension principle of Zadeh (1987).

If $\tilde{A}_1 = (a_1, \alpha_1, \beta_1)$ and $\tilde{A}_2 = (a_2, \alpha_2, \beta_2)$ be in $T(\mathbb{R})$ and $r \in \mathbb{R}$, then:

$$\tilde{A}_1 + \tilde{A}_2 = (a_1 + a_2, \alpha_1 + \alpha_2, \beta_1 + \beta_2), \tag{2}$$

$$r\tilde{A}_1 = \begin{cases} (ra_1, r\alpha_1, r\beta_1) & r \ge 0, \\ (ra_1, -r\beta_1, -r\alpha_1) & r < 0. \end{cases} \tag{3}$$

The multiplication of two TFNs based on extension principle is not necessarily a TFN (Dobois and Prade 1980). The following approximation for the product of two TFNs $\tilde{A}_1 = (a_1, \alpha_1, \beta_1)$ and $\tilde{A}_2 = (a_2, \alpha_2, \beta_2)$ is proposed by Dobois and Prade (1980):

$$\tilde{A}_1 \otimes \tilde{A}_2 \simeq \begin{cases} (a_1a_2, a_1\alpha_2 + a_2\alpha_1, a_1\beta_2 + a_2\beta_1) & \tilde{A}_1 > 0, \tilde{A}_2 > 0, \\ (a_1a_2, a_2\alpha_1 - a_1\beta_2, a_2\beta_1 - a_1\alpha_2) & \tilde{A}_1 < 0, \tilde{A}_2 > 0, \\ (a_1a_2, -a_2\beta_1 - a_1\beta_2, -a_2\alpha_1 - a_1\alpha_2) & \tilde{A}_1 < 0, \tilde{A}_2 < 0, \end{cases} \tag{4}$$

where $\tilde{A}_i > 0 (< 0)$ means $a_i - \alpha_i \ge 0 (a_i + \beta_i \le 0)$, $i = 1, 2$. Note that, we let $\tilde{0} = (0, 0, 0)$ as a zero fuzzy number.

Definition 2.2. *Sakawa (1993) Let $\tilde{A}$ and $\tilde{B}$ be two fuzzy sets. $\tilde{A}$ is said to be a subset of $\tilde{B}$, denoted by $\tilde{A} \subseteq \tilde{B}$, if and only if its membership function is less than or equal to that of $\tilde{B}$ everywhere on X:*

$$\tilde{A} \subseteq \tilde{B} \Leftrightarrow \tilde{A}(x) \le \tilde{B}(x) \quad \forall\, x \in X. \tag{5}$$

Proposition 2.3. *Suppose $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are fuzzy numbers. Then*

$$(\tilde{A} + \tilde{B}) \otimes \tilde{C} \subseteq \tilde{A} \otimes \tilde{C} + \tilde{B} \otimes \tilde{C}. \tag{6}$$

The equality holds if $\tilde{A}$, $\tilde{B}, \tilde{C} > 0$.

Definition 2.4. *A ranking function is a function $\mathcal{R} : T(\mathbb{R}) \to \mathbb{R}$ which maps each fuzzy number into the real line, where a natural order exists. Let $\tilde{A} = (a, \alpha, \beta)$ be a triangular fuzzy number, then $\mathcal{R}(\tilde{A}) = a + \frac{1}{4}(\beta - \alpha)$.*

Definition 2.5. *: Let $\tilde{A} = (a, \alpha, \beta)$ and $\tilde{B} = (b, \gamma, \delta)$ be two triangular fuzzy numbers, then*
(i) $\tilde{A} \le \tilde{B}$ iff $\mathcal{R}(\tilde{A}) \le \mathcal{R}(\tilde{B})$,
(ii) $\tilde{A} < \tilde{B}$ iff $\mathcal{R}(\tilde{A}) < \mathcal{R}(\tilde{B})$.

3 Proposed method

Consider the following fully fuzzy linear programming (FFLP) with m constraints and n variables:

$$\begin{aligned} Min \quad & \tilde{z} = \tilde{C} \otimes \tilde{x} \\ s.t. \quad & \tilde{A} \otimes \tilde{x} = \tilde{b}, \quad \tilde{x} \text{ is unrestricted fuzzy number.} \end{aligned} \tag{7}$$

where $\tilde{C}^T = [\tilde{c}_j]_{1\times n}$ and $rank(\tilde{A}, \tilde{b}) = rank(\tilde{A}) = m$. Due to the limitation of the existing methods, these methods cannot be applied for solving above FFLP problem. Here, to overcome the limitations of the existing methods, a new method is proposed.

Remind that a real number a has infinite representations in the form of $a = a' - a''$, where a' and a'' are nonnegative real numbers. Suppose $\tilde{A} = (a, \gamma, \delta)$ and $\tilde{x} = (x, \alpha, \beta,)$ are two TFNs where $\tilde{x}$ is unrestricted. Set $\tilde{x} = \tilde{x}' + \tilde{x}''$ in which $\tilde{x}' = (x', \alpha', \beta')$ and $\tilde{x}'' = (-x'', \alpha'', \beta'')$ where $x = x' - x''$, $\alpha = \alpha' + \alpha''$, $\beta = \beta' + \beta''$, $x' \geq \alpha'$, $x'' \geq \beta''$, and $a', a'', \alpha', \alpha'', \beta', \beta'' \geq 0$. By using Proposition 2.3, approximation (4), and addition (2) we have:

$$\tilde{A} \otimes \tilde{x} = \tilde{A} \otimes (\tilde{x}' + \tilde{x}'') \subseteq \tilde{A} \otimes \tilde{x}' + \tilde{A} \otimes \tilde{x}'' \tag{8}$$

$$\tilde{A} \otimes \tilde{x}' + \tilde{A} \otimes \tilde{x}'' = \begin{cases} (ax' - ax'', x'\gamma + \alpha' a + x''\delta + \alpha'' a, x'\delta + \beta' a + x''\gamma + \beta'' a) & \tilde{A} \geq 0, \\ (ax' - ax'', x'\gamma - \beta' a + x''\delta - \beta'' a, x'\delta - \alpha' a + x''\gamma - \alpha'' a) & \tilde{A} < 0, \end{cases} \tag{9}$$

Using (8) the FFLP problem (10) can be written as follows:

$$Min \quad \tilde{z} = \tilde{C} \otimes \tilde{x}' + \tilde{C} \otimes \tilde{x}'' \tag{10}$$

$$s.t. \quad \tilde{A} \otimes \tilde{x}' + \tilde{A} \otimes \tilde{x}'' = \tilde{b}, \tag{11}$$

$$\tilde{x}' \otimes \tilde{x}'' = \tilde{0}, \tag{12}$$

$$\tilde{x}' \geq \tilde{0}, \ , \tilde{x}'' \leq 0. \tag{13}$$

by considering the constraint (12), the relation of subsetting in proposition (2.2) can be held as equality, since at least one of $\tilde{x}'$ and $\tilde{x}''$ is zero. Also, using Definition (2.4), the fuzzy objective function of the FFLP problem, can be converted into the crisp objective function. To illustrate this procedure, let us consider the following example.

4 Numerical example

Consider the following FFLP problem

$$\begin{aligned} Max \quad & \tilde{z} = (2,3,1) \otimes \tilde{x}_1 + (3,1,1) \otimes \tilde{x}_2 \\ s.t. \quad & (-2,1,2) \otimes \tilde{x}_1 + (7,1,1) \otimes \tilde{x}_2 = (-17,13,14), \\ & (4,2,2) \otimes \tilde{x}_1 + (-1,1,3) \otimes \tilde{x}_2 = (8,12,10), \end{aligned}$$

where $\tilde{x}_1$ and $\tilde{x}_2$ are unrestricted triangular fuzzy numbers.

By using Definition (2.4) the fuzzy objective function of the FFLP problem, can be converted into the crisp objective function. If set $\tilde{x}_1' = (x_1', \alpha_1', \beta_1')$, $\tilde{x}_1'' = (-x_1'', \alpha_1'', \beta_1'')$ where $\tilde{x}_1 = \tilde{x}_1' + \tilde{x}_1''$ and $\tilde{x}_2' = (x_2', \alpha_2', \beta_2')$, $\tilde{x}_2'' = (-x_2'', \alpha_2'', \beta_2'')$ where $\tilde{x}_2 = \tilde{x}_2' + \tilde{x}_2''$, then by using relation (9) the obtained fuzzy

linear programming problem can be written as follows:

$$\begin{aligned}
Max\ \ z = \mathcal{R}[&(2x_1' - 2x_1'', 3x_1' - 2\beta_1' + x_1'' - 2\beta_1'', x_1' - 2\alpha_1' + 3x_1'' - 2\alpha_1'')+\\
&(3x_2' - 3x_2'', x_2' + 3\alpha_2' + x_2'' + 3\alpha_2'', x_2' + 3\beta_2' + x_2'' + 3\beta_2'')]\\
s.t.\quad &(-2x_1' + 2x_1'', x_1' + 2\beta_1' + 2x_1'' + 2\beta_1'', 2x_1' + 2\alpha_1' + x_1'' + 2\alpha_1'')+\\
&(7x_2' - 7x_2'', x_2' + 7\alpha_2' + x_2'' + 7\alpha_2'', x_2' + 7\beta_2' + x_2'' + 7\beta_2'') = (-17, 13, 14),\\
&(4x_1' - 4x_1'', 2x_1' + 4\alpha_1' + 2x_1'' + 4\alpha_1'', 2x_1' + 4\beta_1' + 2x_1'' + 4\beta_1'')+\\
&(-x_2' + x_2'', x_2' + \beta_2' + 3x_2'' + \beta_2'', 3x_2' + \alpha_2' + x_2'' + \alpha_2'') = (8, 12, 10),\\
&\tilde{x}_1' \otimes \tilde{x''}_1 = 0, \qquad \tilde{x}_2' \otimes \tilde{x}_2'' = 0,\\
&\tilde{x}_1', \tilde{x}_1'', \tilde{x}_2', \tilde{x}_2'' \geq 0.
\end{aligned}$$

Using the arithmetic operations, Definition (2.4) and definition of zero fuzzy number, the above problem can be written as follows:

$$\begin{aligned}
Max\ \ z = &\frac{3}{2}x_1' - \frac{3}{2}x_1'' + 3x_2' - 3x_2'' - \frac{1}{2}\alpha_1' - \frac{1}{2}\alpha_1'' + \frac{1}{2}\beta_1' + \frac{1}{2}\beta_1'' + \frac{3}{4}\beta_2' + \frac{3}{4}\beta_2'' - \frac{3}{4}\alpha_2' - \frac{3}{4}\alpha_2'',\\
s.t.\quad &-2x_1' + 2x_1'' + 7x_2' - 7x_2'' = -17,\\
&x_1' + 2\beta_1' + 2x_1'' + 2\beta_1'' + x_2' + 7\alpha_2' + x_2'' + 7\alpha_2'' = 13,\\
&2x_1' + 2\alpha_1' + x_1'' + 2\alpha_1'' + x_2' + 7\beta_2' + x_2'' + 7\beta_2'' = 14,\\
&4x_1' - 4x_1'' - x_2' + x_2'' = 8,\\
&2x_1' + 4\alpha_1' + 2x_1'' + 4\alpha_1'' + x_2' + \beta_2' + 3x_2'' + \beta_2'' = 12,\\
&2x_1' + 4\beta_1' + 2x_1'' + 4\beta_1'' + 3x_2' + \alpha_2' + x_2'' + \alpha_2'' = 10,\\
&x_1' \geq \alpha_1', \quad x_1'' \geq \beta_1'', \quad x_2' \geq \alpha_2', \quad x_2'' \geq \beta_2'',\\
&x_1'x_1'' = 0, \quad x_1''\alpha_1' + x_1'\beta_1'' = 0; \quad x_1'\alpha_1'' + x_1''\beta_1' = 0,\\
&x_2'x_2'' = 0, \quad x_2''\alpha_2' + x_2'\beta_2'' = 0; \quad x_2'\alpha_2'' + x_2''\beta_2' = 0;
\end{aligned}$$

The optimal solution of the crisp programming problem is
$x_1' = 1.5,\ \alpha_1' = 0.46,\ \beta_1' = 0.98, \qquad x_1'' = 0,\ \alpha_1'' = 0,\ \beta_1'' = 0$
$x_2' = 0, \quad \alpha_2' = 0, \quad \beta_2' = 0, \qquad x_2'' = 2,\ \alpha_2'' = 1.077,\ \beta_2'' = 1.154$
Since $\tilde{x}_1 = \tilde{x}_1' + \tilde{x}_1''$ and $\tilde{x}_2 = \tilde{x}_2' + \tilde{x}_2''$, the fuzzy optimal solution is $\tilde{x}_1 = (1.5, 0.46, 0.98)$, $\tilde{x}_2 = (-2, 1.077, 1.154)$. By putting $\tilde{x}_1$ and $\tilde{x}_2$ in objective function, the fuzzy optimal value is $\tilde{z} = (-3, 7.771, 5.682)$.

Conclusion

In the past few years, a growing interest has been shown in fuzzy linear programming and currently there are several methods for solving FFLP. However, most of them has not discussed the solving FFLP with unrestricted fuzzy coefficients and fuzzy variables. In this paper a new efficient method for FFLP has been proposed, in order to obtain the fuzzy optimal solution with unrestricted variables and parameters. To show the efficiency of the proposed method a numerical example have been illustrated.

References

Dehghan M., Hashemi B. and Ghatee M. (2006), Computational methods for solving fully fuzzy linear systems, *Appl. Math. Comput.* 179, 328-343.

Ezzati R., Khorram E. and Enayati R. (2015), A new algorithm to solve fully fuzzy linear programming problems using the MOLP problem, *Appl. Math. Modell.* 39, 3183-3193.

Lotfi F.H., Allahviranloo T., Jondabeha M.A. and Alizadeh L. (2009), Solving a fully fuzzy linear programming using lexicography method and fuzzy approximate solution, *Appl. Math. Modell.* 33, 3151-3156.

Saberi Najafi H., Edalatpanah S.A. and Dutta H. (2016), A nonlinear model for fully fuzzy linear programming with fully unrestricted variables and parameters, *Alexandria Engineering Journal* 55, 2589-2595.

Sakawa M. (1993), *Fuzzy Sets and Interactive Multiobjective Optimization*, Plenum Press, New York.

Shamooshaki M.M., Hosseinzadeh A. and Edalatpanah S.A. (2015), A new method for solving fully fuzzy linear programming problems by using the lexicography method, *Appl. Comput. Math.* 4, 1-3.

9th National Seminar on
Fuzzy Statistics and Probability

Evaluation of incidence rate of cancer based on index of prevalence obesity by using fuzzy-multivariate adaptive regression splines

Mina Rezaei*[1], S. Mahmoud Taheri[2], Mahshid Namdari[3], Roshanak Alimohammadi[4]

[1]Department of Mathematics Sciences, Alzahra University

[2]School of Engineering Science, College of Engineering, University of Tehran

[3]Oral Health Department, School of Dentistry, Shahid Beheshti University of Medical Sciences

[4]Department of Mathematics Sciences, Alzahra University

Abstract

Using the Multivariate Adaptive Regression Splines (MARS) technique, a fuzzy regression model for imprecise reponse and crisp explanatory variables is studied. The investigated fuzzy regression model is applied to a certain health index based on some real data. The accuracy of the method is compared with two well-known fuzzy regression models, namely, the fuzzy least absolutes model and the fuzzy least squares model. The comparison results reveal that the MARS fuzzy model performs better than the other models. This comparison is done based on two goodness of fit criteria.

Keywords: Fuzzy regression, Similarity measure, Incidence rate of cancer, Prevalence obesity.

Mathematics Subject Classification (2010): 62A86.

1 Introduction

In real studies, especially in medical ones, we usually face to vague data. In such situations the data are measured non-crisp or they are reported non-crisp. In these cases, an alternative suitable method to analysis the data is to use fuzzy modeling (see e.g. Namdari et al. (2014)).

In the present work, the fuzzy MARS model is explained, and then the application of such a model to find the relationship between a certain medical characteristic and incidence rate of cancer is studied.

In Section 2, we expain why and when using MARS is effective. The investigated model is described in Section 3. In Section 4, the criterion that are used to evaluate the model are reviwed. The fuzzy MARS method is employed for modeling a relationship between index of prevalence obesity and incidence rate of cancer in Section 5. A comparison study is presented in Section 6. In Section 7, forecasting for a new case is explained.

*Speaker: mn.rezaee7@gmail.com

2 Why MARS?

The multivariate adaptive regression splines (MARS) is an adaptive non parametric piecewise regression approach that makes no assumptions about the functional relationship between the response variables and explanatory variables. It is an appropriate technique for modeling a data set, especially when the range of data is large and the relationship between variables is nonlinear, see Friedman (1991). MARS is a flexible regression technique that uses a modified recursive partitioning strategy to simplify high dimensional problems, into smaller yet highly accurate models. Chachi et al. (2016) introduced a novel approach which combine the MARS technique and fuzzy regression methods in order to analyse the fuzzy data for which a large data set or when relationship between the response variables (which is not crisp) and explanatory variables is nonlinear. As it is, the fuzzy least squares and fuzzy least absulotes approaches are two most studied paradigms in the parametric modelling of imprecise data well-known. While these methods are relatively easy to develpe and interpret, they have a limited flexibility and work well only when the true linear underlying relationship is close to the pre-specified approximated function in the model. In practical studies, however, these methods are not suitable and give misleading results while the true underlying relationship in nonlinear. To overcome the weaknesses of the parametric modelling approaches, non-parametric models are developed locally over specific subregions of the data, i.e. the dataset is searched for the optimum number of subregions and then a simple function is optimally fit to the realizations in each subregion. Such methods are able to approximate the underlying nonlinear relationship between a target variable and a set of explanatory variables as well as possible.

Unlike better known linear regression techniques, MARS does not assume coefficients are stable across the entire range of each variable and instead uses splines to fit piecewise continuous functions to model responses. This is very useful when it is suspected that model inputs have varying optima across different levels of the model inputs. MARS is highly sensitive to both sample size and design of experiment type, see Hastie et al. (2009).

The main objective of this paper is to employ the MARS technique to construct a fuzzy regression model for crisp input- fuzzy output data, and then to study such a model in a medical problem based on a real world dataset.

3 The investigated method

For estimating a fuzzy response variable $\tilde{y} = (y, \delta)$ using a set of explanatory variable x, the best model is the one with minimum total error in predicting the centre value and the spread value of fuzzy reponse variable. To find a relationship between a reponse variable, y, and an explanatory variable, x, it may be apparent that for different ranges of x, different linear relationships occure. In these cases, a single linear model may not provide an adequate description. The MARS technique is a form of regression that allows multiple linear (or nonlinear) models to be fitted to the data for different ranges of x, see Hastie et al. (2009). Therfore in this section, we present a new method to estimate the fuzzy response variable by mean of the idea introduced in Chachi et al. (2016):

a) modelling the centres of the response variable via a MARS model on the explanatory variables; and

b) simultaneously modelling the transformation of the spreads of the response throught another MARS model, as follows

$$y = \beta_0 B_0(x) + \sum_{m=1}^{M} \beta_m B_m(x) \tag{1}$$
$$G(\delta) = \gamma_0 B_0(x) + \sum_{m=1}^{M} \gamma_m B_m(x)$$

In which $G : (0, \infty) \to \Re$ is an invertible function.

In this method, we do not impose a non negativity condition to the estimation problem to avoid negative estimated spreads, but we use an alternative method proposed by Ferraro et al. (2010) to remove the non-negativity condition. In this method, we propose a transformation of the spread of the response through MARS model. Indeed, using this method, the spread of the response variable is obtained as:

$$\delta = G^{-1}(\gamma_0 B_0(x) + \sum_{m=1}^{M} \gamma_m B_m(x)) \tag{2}$$

which is always non-negative. A common approach consists in transforming the spread by mean of the natural logarithmic transformation. i.e. $G(t) = ln(t)$. We will use this approach in the numerical example to transform the spread into real variables without the restriction of non-negativity.

Now, by applying the *earth* package in software R, we can estimate the coefficient $\beta_0, \beta_1, ..., \beta_k$ and $\gamma_0, \gamma_1, ..., \gamma_k$.

4 Evaluation Methods

In this study, we use two well known criteria to evaluate the obtained fuzzy regression models. These criteria are the common indices for evaluating the gooness-of-fit of fuzzy regression models used by many authors (Chachi and Taheri , 2013; Hojati et al. , 2005; Kelkinnama and Taheri , 2012).

4.1 Mean of relative errors (MRE)

This criterion initially introduced by Kim and Bishu (1998), is defined as $MRE = \frac{1}{n} \sum_{i=1}^{n} E(\tilde{y}_i, \hat{\tilde{y}}_i)$, where

$$E(\tilde{y}_i, \hat{\tilde{y}}_i) = \int \frac{|\tilde{y}_i(x) - \hat{\tilde{y}}_i(x)|}{\int \tilde{y}_i(x)dx} dx \tag{3}$$

This index is the ratio of the total difference between the estimated and observed membership values of response variable to the total observed membership values of response variable.

4.2 Mean of similarity measures (MSM)

This index is defined based on the similarity of fuzzy numbers as $MSM = \frac{1}{n} \sum_{i=1}^{n} S(\tilde{y}_i, \hat{\tilde{y}}_i)$, in which

$$S(\tilde{y}_i, \hat{\tilde{y}}_i) = \frac{\int min\{\tilde{y}_i(x), \hat{\tilde{y}}_i(x)\}dx}{\int max\{\tilde{y}_i(x), \hat{\tilde{y}}_i(x)\}dx} \tag{4}$$

The MSM is between 0 and 1. More closer to 1, better model. The MRE could be any amount more than 0. When the amount is closer to 0, shows that the model estimates data well. For unifing ranges of criteria, we use a quantity $G = \frac{1}{1+MRE}$ instead of MRE.

5 Application in modeling a relationship between index of prevalence of obesity and incidence rate of cancer

Health is one of the most impotant issues that has been always discussed. As we know, health has lots of aspects so providing health is so hard and challenging for goverments. Certainly for a good planning, having reliable information is necessary. Due to importance of statistics and medical information, a priority of Ministry of Health and Medicine is improving statistics and informations, so Ministry of Health and Medicine published a report in 1390. The report is collections of data about illnesses and health in envoirement and their analysis. Main reason of collecting data is describing priorities indexes health in country about illnesses and envoiroment health which is partitioned provinces and medical sciences universities from 1385 to 1390. In this paper we want to describe the relationship between index of prevalence of obesity and incidence rate of cancer in 30 province of Iran. The methodology is based on ecological studies. The units of observation in an ecologic study are usually geographically defined populations (such as countries or regions within a country) or the same geographically defined population at different points in time, see Szklo and Nieto (2019). Values of index of prevalence obesity is crisp, but due to some limitations in experimental envoiroment values of incidence rate of cancer is considered triangular fuzzy numbers, shown in Table 1. We want to model the relationship between

Table 1: Crisp input values of index of prevalence obesity and fuzzy output values of incidence rate of cancer in each province in Iran (year 1388)

province	x_i	$_i = (y, \delta)_i$
Azarbayjane sharghi	15.63	$(295.22,71.12)_T$
Azarbayjane gharbi	18.09	$(222.34,59.67)_T$
Ardebil	18.59	$(223.21,63.76)_T$
Esfehan	14.68	$(313.60,77.54)_T$
⋮	⋮	⋮
Hormozgan	18.00	$(331.35,88.79)_T$
Hamedan	8.21	$(158.76,31.54)_T$
Yazd	14.31	$(177.14,43.30)_T$

incidence rate of cancer as the fuzzy dependent variable ($\tilde{y} = (y, \delta)_T$) and index of prevalence obesity as the crisp independent variable (x). The scatterplot of centres of incidence rate of cancer in terms of index of prevalence obesity and the estimated MARS model is shown in Figure 1. In this figure, a single linear model may not provide an adequate description. Therefore, we applied the MARS-fuzzy model to analyse the data. We also examine the fuzzy least absolutes and fuzzy least squares regression models for the data. The fuzzy regression models are obtained as follows.

Fuzzy least squares regression model Celmins (1987)

$$\hat{\tilde{y}} = (136.01 + 6.01x, 21.78 + 2.15x)_T \tag{5}$$

Fuzzy least absulotes regression model Kelkinnama and Taheri (2012)

$$\hat{\tilde{y}} = (127.79, 8.16)_T \oplus (6.31, 2.84)_T x \tag{6}$$

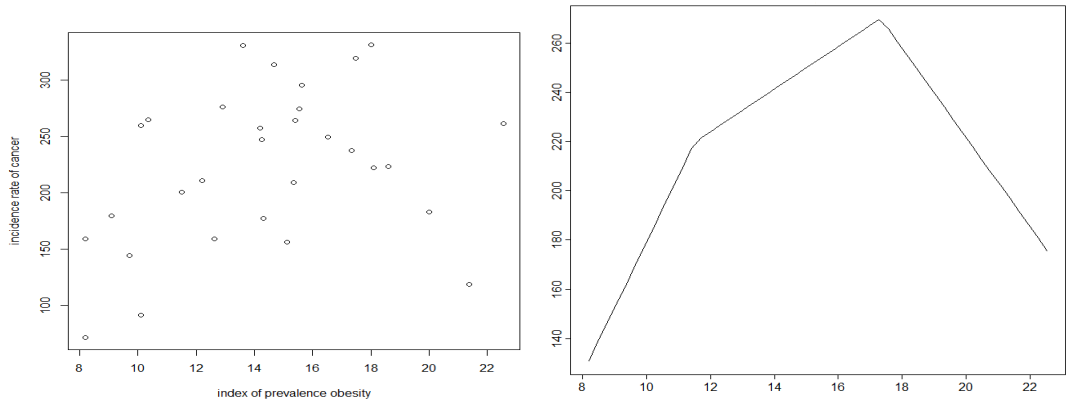


Figure 1: Scatter plot of centres of incidence rate of cancer in terms of index of prevalence obesity .

Fuzzy MARS model Chachi et al. (2016)

In the MARS method to estimate the fuzzy response variable there is no need to consider the non-negativity of the observed values of x. In this method, we transformed the spreads by means of the logarithmic transformation. Therefore, by applying the proposed procedure, the centres and spreads of fuzzy response variable are estimated as $\hat{\tilde{y}}_{MARS} = (\hat{y}, \hat{\delta})_T$, where:

$$\hat{y} = 375.70 - 18.12max(0, x - 11.5) - 26.78max(0, 17.33 - x) \tag{7}$$

$$\hat{\delta} = exp\{58.99 - 12.33max(0, 11.5 - x)\} \tag{8}$$

6 Comparison study

To compare the performances of the three fuzzy regression models, the indices MSM and G (MRE) are adopted to calculate the errors and similarities in estimating the the observed responses. The results are shown in Table 2. As a comparison between models, The MSM of MARS model is 0.21, which is bigger

Table 2: Comparison between fuzzy least-squares regression model, fuzzy least-absulotes model and fuzzy MARS.

Model	G	MSM
Fuzzy least-squares regression	0.01	0.11
Fuzzy least-absulotes regression	0.08	0.02
MARS-fuzzy regression	0.37	0.21

than those of 0.02 and 0.11, calculated from the least squares and least absulotes models, respectively. This means that the MARS model fits the majority of the data better than the other two models. Also, The G of the MARS model is 0.37, which in bigger than those of 0.08 and 0.01. Based on this criteria, the MARS model fits the data better than two other models.

7 Forecasting for a new case

The above models can also be adopted for forecasting the associated response for a specific value of a independent variable. For example, suppose that for a new case, we observe x=11.84 for the index of prevalence obesity, and we want to forecast the associated amount of incidence rate of cancer. By substituting this value into the models, the estimated responses based on the LS model, the LA models and the MARS model are obtained as follows:

$$\hat{\tilde{Y}}_{LS} = (241.51, 109.58)_T$$

$$\hat{\tilde{Y}}_{LA} = (385.42, 124.17)_T$$

$$\hat{\tilde{Y}}_{MARS} = (222.51, 104.83)_T$$

This means that, according to the LS model, the predicted value of incidence rate of cancer would be about 241.51 with a spread value of 109.58, according to the LA model, the predicted value of incidence rate of cancer would be about 385.42 with a spread value of 124.14, according to the MARS model, the predicted value of incidence rate of cancer would be about 222.51 with a spread value of 104.83.

Conclusion

A fuzzy regression method for modelling crisp input-fuzzy output data has been reviewed based on MARS technique. Such a model was applied to health index studies to estimate incidence rate of cancer based on index of prevalence obesity. The MARS fuzzy model seems to be able to provide suffieciently accurate results at least for for the solved numerical example in this paper. More numerical studies are needed to have a better evaluation of fuzzy MARS method.

References

Celmins A. (1987), Least squares model fitting to fuzzy vector data, *Fuzzy Sets and Systems* 23, 245-269.

Chachi J. and Taheri S.M. (2013), A least-absulotes regression model for imprecise response based on the generalized Hausdroff-metric, *Journal of Uncertain Systems* 7, 265-276.

Chachi J., Taheri, S.M. and Pazhand H.R. (2016), Suspended load estimation usind L1-fuzzy regression, L2-fuzzy regression and MARS-fuzzy regression models, *Hydrological Sciences Journal* 61, 1489-1502.

Ferraro M.B., Coppi R., González Rodríguez G. and Colubi A. (2010), A linear regression model for imprecise response, *International Journal of Approximate Reasoning* 51, 759-770.

Friedman J. (1991), Multivariate adaptive regression splines, *The Annals of Statistics* 19, 1-16.

Hastie T., Tibshirani R. and Friedman J. (2009), *The Elements of Statistical Learning*, Second Edition, Springer, New York.

Hojati M., Bector C.R. and Smimou K. (2005), A simple method for computation of fuzzy linear regression, *European Journal of Operational Research* 166, 172-184.

https://behdasht.gov.ir/uploads/291-1041-simayei-salamat.pdf.

Kelkinnama M. and Taheri S.M. (2012), Fuzzy least-absulotes regression using shape preserving operations, *Information Sciences* 214, 105-120.

Kim B. and Bishu R.R. (1998), Evaluation of fuzzy linear regression models by comparing membership functions, *Fuzzy Sets and Systems* 100, 343-352.

Namdari M., Abadi A., Taheri, S.M., Rezaei M., Kalantari N. and Omidvar N. (2014), Effect of folic acid on appetite in children: Ordinal logistic and fuzzy logistic regression, *Nutrition* 30, 274-278.

Szklo M. and Nieto J. (2019), *Epidemiology Beyond the Basics*, Fourth Edition, Jones & Bartlett Learning.

9th National Seminar on
Fuzzy Statistics and Probability

Bayesian confidence limits of a process capability index for discrete processes

M. Salehi Sarvestani*, A. Parchami, M. Mashinchi,
Department of Statistics, Faculty of Mathematics and Computer,
Shahid Bahonar University of Kerman, Kerman, Iran

Abstract

Process capability indices provide numerical measures to compare the output of a process to Clients expectations. However, most of the existing researches have used traditional distribution frequency method by using a single sample due to assess process capability. An alternative to this approach is to use the Bayesian method. In this paper, we utilize a Bayesian approach based on subsamples to check process capability via capability index C_{pc}. As a new suggestion, we used the informative gamma prior distribution and the characteristics of sufficient statistic of the parameter to drive the posterior distribution. Finally, we derive a Bayesian confidence limit for the index C_{pc}.

Keywords: Bayesian Confidence Bound, Posterior Probability, Process Capability Indices.
Mathematics Subject Classification (2010): 60E05, 62F15, 62N10.

1 Introduction

The vast majority of the process capability indices that have been considered are associated only with processes that can be described through some continuous and, in particular, normally distributed characteristics. The most widely used such indices C_p, C_{pk}, C_{pm} and C_{pmk} or their generalizations for non-normal processes, suggested by Clements Clements (1989), Pearn and Kotz Pearn and Kotz (1994), and Pearn and Chen Pearn and Chen (1995). Often, however, one is faced with processes described by a characteristic whose values are discrete. Therefore, in such cases none of these indices can be used. To our knowledge, the only indices suggested so far whose assessment is meaningful regardless of whether the studied process is discrete or continuous are those suggested by Yeh and Bhattacharya Yeh and Bhattacharya (1998), Borges and Ho Borges and Ho (2001), and Perakis and Xekalaki Perakis and Xekalaki (2002).

A very common example of a process that can be described through a discrete-valued characteristic is the number of defects per produced unit by an industry. In this case, it is obvious that small process values are desirable. In some other cases, however, large values may be desirable. In what follows, we consider only discrete processes with unilateral tolerances (i.e. processes connected with just one specification limit, lower (L) or upper (U)), since in practice, discrete processes with bilateral tolerances are only rarely considered. Of course, extending the results obtained below to the case of bilateral tolerances is possible but tedious and requires a much more complicated analysis.

*Speaker: salehimohammad991@gmail.com

As already mentioned, Perakis and Xekalaki (2002) proposed a new index, which can be used regardless of whether the examined process is discrete or continuous. This index is defined as

$$C_{pc} = \frac{1-p_0}{1-p},$$

where p and p_0 denote the proportion of conformance (yield) and the minimum allowable proportion of conformance of the examined process, respectively. As is well known, the term proportion of conformance refers to the probability of producing within the so-called specification area, i.e. the interval determined by L and U. If the tolerance are unilateral, then the value op p is given by $P(X > L)$, if only L has been set, and by $P(X < U)$, if only U has been assigned. According to Perakis and Xekalaki (2002), the value 0.9973 is a plausible choice for p_0, since this probability is usually regarded as sufficiently large in statistical process control. However, different choices of p_0 can be made, according to the nature of the process examined. In the sequel, the value 0.9973 is chosen. Nevertheless, the analysis given in the sequel can be readily modified for any other choice of p_0. The properties of C_{pc} in the case where the distribution of the process studied is normal or exponential are investigated thoroughly by Perakis and Xekalaki (2002).

2 The index C_{pc} for Poisson processes

As already mentioned, one of the advantages of the index C_{pc} is that its assessment is possible even if the examined process is discrete. In this section, the properties of C_{pc} are examined in the case where the studied process is described by a Poisson-distributed characteristic with some parameter $\theta > 0$. In order to avoid confusion, the following notation is adopted: if only the value of U has been set, the index is denoted by C_{pcu}, while if only the value of L has been set, the index is denoted by C_{pcl}. Thus, the index C_{pcu} is defined as

$$C_{pcu} = \frac{0.0027}{1-p},$$

where $p = P(X < U)$ and the index C_{pcl} is defined by

$$C_{pcl} = \frac{0.0027}{1-p},$$

with $p = P(X > L)$. Note that the values U and L are assumed to lie outside the specification area. The probability p that is involved in the denominator of the index C_{pcu} is equal to the sum

$$\sum_{x=0}^{U-1} \frac{e^{-\theta}\theta^x}{x!}$$

Note that p is the cumulative distribution function of the Poisson distribution with parameter $\theta > 0$ evaluated at $U-1$. According to page 196 Johnson et al. (2005) we have

$$\begin{aligned} p &= P(X \leq U-1) \\ &= P(\chi^2_{2U} > 2\theta) \end{aligned}$$

where χ^2_{2U} denotes the Chi-square distribution with $2U$ degrees of freedom. Using this property, the index C_{pcu}can be written as

$$C_{pcu} = \frac{0.0027}{P\left(\chi^2_{2U} < 2\theta\right)}.$$

Similarly, one may observe that the value of p that appears in the denominator of the index C_{pcl} can be written in the form

$$\begin{aligned} p &= 1 - P(X \leq L) \\ &= 1 - P(\chi^2_{2(L+1)} > 2\theta) \\ &= P(\chi^2_{2(L+1)} < 2\theta) \end{aligned}$$

and thus

$$C_{pcl} = \frac{0.0027}{P\left(\chi^2_{2(L+1)} > 2\theta\right)}.$$

It would be interesting to remark that by their definition, the indices, C_{pcu} and C_{pcl} are rerelated directly to the proportion of conformance of the process. This rather interesting property constitutes, undoubtedly, an appealing feature lacked by all the most frequently used process capability indices, such as C_p, C_{pk}, C_{pm}, and C_{pmk}.

3 Bayesian approach for capability estimation

We can say that defining or choosing the prior distribution for the parameters is the most challengeable step in Bayesian inference. Throughout the papers Pearn and Wu (2005), Wu and Pearn (2006), Wu and Pearn (2005) in order to utilize their newmethod, Pearn and Wu usually supposed that there is no information about the parameter(s) and used non-informative reference prior $h(\mu, \sigma) = 1/\sigma$ which minimize the effect of prior information on posterior distribution. However, in real world, all of those manufacturers who want to have an effective presence in the free market have to perform a quality control system on their production lines. Hence, the first step for them is monitoring process stability with regular and fixed sampling plan which costs considerable time and money. As a sensible act, one should try to use these prior samples information as much as possible in order to choose an informative prior distribution which improves the validity of inference.

3.1 Why informative Gamma prior?

Bayesian inference is a dynamic inferential process which updates our information about the under investigation parameter. Main point of Bayesian analysis and its most important difference with frequency approach is the selection of a prior model for behavior of the parameter based on prior information. Choosing an informative prior for parameter does not mean that we want to select a complete model to reflect the whole properties of the parameter because such a prior does not exist. We just try to use a valuable prior information about the parameter to choose a suitable model for it, one with the most similarity to our parameter which reflects the parameters characteristics as much as possible. Hence, as the first step in Bayesian inference, we should accept the challenge of finding a suitable model for our parameter. Actually, what we call non-informative priors were introduced by Jeffreys around 200 years after Bayes theory introduction in order to convince that group of criticizer who claimed that Bayesian approach is a Subjective method and should not be considered as a scientific one. Berger (1985), Robert (2007). Non-informative priors are made to minimize the effect of our prior information on our inference, and it is common to utilize them to introduce a new Bayesian approach (like something that Pearn and Wu did), but in real world, we are allowed to start our Bayesian inference with a non-informative prior just when there is not any reliable information about the parameter. This is a rare situation which happens at the start point of an industrial process only. Otherwise, using non-informative prior will be in opposition with the philosophy of Bayesian inference, and the final results of our inference will be a

frequency conclusion just under the name of Bayesian. We should try to utilize informative prior and analyze different aspects of this usage in order to bring our theoretical knowledge to practical cases. Therefore, in this paper, we try to bring the valuable results of Pearn and Wus paper to the conditions of real world.

Also, one of the most common procedures for choosing prior distribution is comparing the parameters attributes with known distributions (like normal, uniform, gamma and etc.) and choosing one of them which have the most compatibility Robert (2007). Moreover, mean of a manufacturing process has been regarded as one of the most important factor in quality control, which its real value is always unknown. Since unknown m is a location parameter and varies along real numbers axis, choosing gamma distribution as its model can seem logical.

Our under investigation parameter is the overall mean of the population. A major assumption is the poisson distribution for every member of this population in an industrial process. Hence, selecting gamma distribution as the prior for this overall mean makes sense, and that is another reason for why we use gamma prior in capability testing.

3.2 Posterior probability of process capability

Considering the above discussion in subsection 3.1, we will assume the following assumptions for the rest of this paper:

1. The process is statistically stable and under control.
2. The process is Poisson $P(\theta)$ with unknown mean.
3. Based on the prior information, the unknown parameter of the Poisson model has Gamma distribution $\Gamma(\alpha, \beta)$, where α and β are constants.

Theorem 3.1. *Suppose that the process is under statistical control and it is Poisson with unknown mean θ. Considering prior distribution $\Gamma(\alpha, \beta)$ for unknown parameter θ and squared error loss function, a $100(1-\alpha)\%$ Bayesian lower confidence limit for C_{pcu} is given by*

$$\frac{0.0027}{P\left(\chi^2_{2U} < B\chi^2_{2A(\underline{x}),1-\alpha}\right)},$$

and $100(1-\alpha)\%$ lower confidence limit for C_{pcl} is given by

$$\frac{0.0027}{1-P\left(\chi^2_{2(L+1)} < B\chi^2_{2A(\underline{x}),\alpha}\right)},$$

where $A = n\bar{x} + \alpha$, $B(\underline{x}) = \dfrac{1}{n + \dfrac{1}{\beta}}$ and $\chi^2_{\nu,1-\alpha}$ denotes the $1-\alpha$ quantile of the Chi-square distribution with ν degrees of freedom.

4 Simulation

In order to test the performance of the observed lower confidence limit for the index C_{pcu} and C_{pcl} a simulation study was conducted. In the conducted simulation study 50 random samples were generated from the Poisson distribution for various values of the parameter θ and we assumed that $L = 5$, $U = 20$ and $\theta \sim \Gamma(1, 1)$. The below table summarizes the obtained results.

θ	C_{pcu}	*Bayesian lower bound for* C_{pcu}	C_{pcl}	*Bayesian lower bound for* C_{pcl}
8	10.67449	2.21681	0.01411868	0.01268588
9	2.55693	1.194904	0.02333813	0.01589968
10	0.781625	0.3362839	0.04024687	0.02853576
11	0.290652	0.1195389	0.07196198	0.05457955
12	0.1268811	0.09578708	0.1327366	0.06448502

As we can see our approach works well for all the studied cases, especially for larger values of the parameter θ and it seems to be quite close to true value of the index C_{pcu} and C_{pcl}.

Conclusion

In this article, under squared error loss function, we found a Bayesian lower confidence bound for the index C_{pc} where the studied process is described by a Poisson distribution with some parameter θ and assuming prior distribution $\Gamma(\alpha, \beta)$ for θ. The obtained results offer a useful approach for measuring capabilities of the processes that are described by this assumptions. The study of C_{pc} under different distributional assumptions would be an interesting issue for further research.

References

Berger J.O. (1985), *Statistical Decision Theory and Bayesian Analysis*, Second Edition, Springer, New York.

Borges W. and Ho L.L. (2001), A fraction defective based capability index, *Quality and Reliability Engineering International* 17, 447-458.

Clements J.A. (1989), Process capability calculations for non-normal distributions, *Quality Progress* 22, 95-100.

Johnson N.L., Kotz S. and Kemp A.W. (2005), *Univariate Discrete Distributions*, Third Edition, Wiley, New Jersey.

Pearn W.L. and Chen K.S. (1995), Estimating process capability indices for non-normal pearsonian populations, *Quality and Reliability Engineering International* 11, 389-391.

Pearn W.L. and Kotz S. (1994), Application of Clements method for calculating second- and third-generation process capability indices for non-normal pearsonian populations, *Quality Engineering* 7, 139-145.

Pearn W.L. and Wu C.W. (2005), A Bayesian Approach for Assessing Process Precision Based on Multiple Samples, *European Journal of Operational Research* 165, 385-395.

Perakis M. and Xekalaki E. (2002), A process capability index that is based on the proportion of conformance, *Journal of Statistical Computation and Simulation* 72, 707-718.

Robert C.P. (2007), *The Bayesian Choice from Decision-Theoretic Foundations to Computational Implementation*, Second Edition, Springer, New York.

Wu C.W. and Pearn W.L. (2005), Capability Testing Based C_{pm} with Multiple Samples, *Quality and Reliability Engineering International* 21, 29-42.

Wu C.W. and Pearn W.L. (2006), Bayesian, Approach for Measuring EEPROM Process Capability Based on the One-sided Indices C_{pu} and C_{pl}, *The International Journal of Advanced Manufacturing Technology* 31, 135-144.

Yeh A.B. and Bhattacharya S. (1998), A robust process capability index, *Communications in Statistics Simulation Computation* 27, 565-589.

9th National Seminar on
Fuzzy Statistics and Probability

Minimum cost flow problem with triangular fuzzy demands/supplies

Javad Tayyebi*[1] , Elham Hosseinzadeh[2], Sakine Beigi[3]

[1]Department of Industrial Engineering, Birjand University of Technology, Birjand, Iran

[2]Department of Mathematics, Kosar University, Bojnord, Iran

[3]Department of Industrial Engineering, Kosar University, Bojnord, Iran

Abstract

This paper addresses a generalized form of minimum cost flow problems in which demands and supplies are triangular fuzzy numbers. This problem admits a fuzzy flow on arcs of the network due to fuzzy demands and supplies. Under a special lexicographic order, a successive shortest path algorithm is introduced to solve the problem.

Keywords: Minimum cost flow problem, fuzzy demands, fuzzy supplies, fuzzy flow.

1 Introduction

One of most well-known problems in the operation research field is minimum cost flow problem. It is to send a flow of minimum cost from sources to sinks in a given network. The problem contains two groups of constraints: balanced constraints and bound constraints. The balanced constraints guarantee that the entering flow to a node is equal to its leaving flow. The bound constraints state that flow along an arc cannot exceed the arc's capacity. The special versions of the minimum cost flow problem play a central role in the theory and applications of network flows. (see Ahuja et al. (1993) for more study).

In this paper, a generalized form of minimum cost flow problems is investigated in which nodes' demands/supplies are vague and ambiguous. The problem is interested by some authors (see Ghatee and Hashemi (2008, 2009)). The general idea is to convert the problem into a crisp one by using ranking functions and then to solve the crisp problem to obtain an optimal fuzzy flow. This paper presents an efficient algorithm to be capable of solving the problem without converting it into a crisp problem. This maintains the fuzzy essence of the problem.

It is assumed that they can be determined by triangular fuzzy numbers. Let s and t be respectively the source and sink nodes. Nodes not belonging to the set $\{s, t\}$ are called transshipment nodes. It is remarkable that there may be more than one source and one sink in practice. Fortunately, one can transform this situation to the case with only one source and one sink by introducing a super source and a super sink, see Ahuja et al. (1993). Without loss of the generality, we assume that the demand of t is equal to the supply of s. Obviously, if not, one can add a dummy node of zero costs to the network to satisfy the assumption. Suppose that $\tilde{b} = (b^1, b^2, b^3)$ is triangular fuzzy demand and supply of nodes

*Speaker: javadtayyebi@birjandut.ac.ir

s and t where $b^1 \leq b^2 \leq b^3$. Since $\tilde{b}$ is a fuzzy number, it follows that flows (decision variables) are also fuzzy (see Mahdavi-Amiri and Nasseri (2007); Nasseri et al. (2010) for more study on fuzzy variables).

Assume that $\tilde{x}_{ij} = (x^1_{ij}, x^2_{ij}, x^3_{ij})$ is triangular fuzzy flow along arc (i, j). One can simply formulate the problem as follows:

$$\min \quad \tilde{z} = \sum_{(i,j)\in A} c_{ij}\tilde{x}_{ij} \tag{1a}$$

$$\text{s.t.} \quad \sum_{j:(i,j)\in A} \tilde{x}_{ij} = \sum_{j:(j,i)\in A} \tilde{x}_{ji} + \tilde{b} \qquad i = s, \tag{1b}$$

$$\sum_{j:(i,j)\in A} \tilde{x}_{ij} + \tilde{b} = \sum_{j:(j,i)\in A} \tilde{x}_{ji} \qquad i = t, \tag{1c}$$

$$\sum_{j:(i,j)\in A} \tilde{x}_{ij} = \sum_{j:(j,i)\in A} \tilde{x}_{ji} \qquad \forall i \in V\backslash\{s, t\}, \tag{1d}$$

$$0 \leq \tilde{x}_{ij} \leq \tilde{u}_{ij}, \tag{1e}$$

in which $G(V, A)$ is the underlying graph, c_{ij} and $\tilde{u}_{ij} = (u^1_{ij}, u^2_{ij}, u^3_{ij})$ are respectively the cost and the capacity associated with arc (i, j). Since the problem (1) contains fuzzy values in its objective value and its constraints, it is required to define the notions of the equality and the inequality. Here, we mean that two triangular fuzzy numbers $\tilde{a} = (a^1, a^2, a^3)$ and $\tilde{b} = (b^1, b^2, b^3)$ are equal together if and only if $a^i = b^i$ for $i = 1, 2, 3$. We use a lexicographic order to compare fuzzy numbers $\tilde{a}$ and $\tilde{b}$ in the following manner:

$$\begin{aligned}&\tilde{a} \leq \tilde{b} \text{ iff } \tilde{a} = \tilde{b} \text{ or } a^1 < b^1 \text{ or } (a^1 = b^1 \text{ and } a^2 < b^2) \text{ or}\\ &(a^1 = b^1 \text{ and } a^2 = b^2 \text{ and } a^3 < b^3).\end{aligned} \tag{2}$$

2 Proposed algorithm

Our proposed algorithm is a generalization of the well-known successive shortest path algorithm.

At first, the algorithm sets $\tilde{x}_{ij} = (0, 0, 0)$ for every $(i, j) \in A$. Based on the lexicographic order, the priority of the first component of fuzzy numbers is greater than the others. So, the algorithm focuses sending flow with respect to first components in the first phase. For this reason, it first finds a shortest path from s to t and then sends a maximum possible flow along it. The maximum value sent along path P is (f_1, f_1, f_1) where $f_1 = \min\{b^1_s, \min_{(i,j)\in P} u^1_{ij}\}$. To construct the corresponding residual network, it changes capacity of every $(i, j) \in P$ to $(u^1_{ij} - f_1, u^2_{ij} - f_1, u^3_{ij} - f_1)$. Moreover, it adds an arc (j, i) with cost $-c_{ij}$ and capacity (f_1, f_1, f_1) for every $(i, j) \in P$. Finally, the supply of s (the demand of t) is updated to $(b^1 - f_1, b^2 - f_1, b^3 - f_1)$. The procedure is repeated until $b^1 = 0$. Then, the algorithm begins the second phase.

In the second phase, the algorithm corrects the second and third components of arc capacities in the following manner:

- if $x^1_{ij} = u^1_{ij}$, then u^2_{ij} and u^3_{ij} are not changed;
- if $x^1_{ij} < u^1_{ij}$, then $u^2_{ij} = u^3_{ij} = +\infty$.

These changes guarantee that bound constraints are satisfied in the lexicographic order. Then, the algorithm finds a shortest path P from s to t in the residual network. It sends $(0, f_2, f_2)$ units of flow along P and update the residual network where $f_2 = \min\{b^2, \min_{(i,j)\in P} u^2_{ij}\}$. Sending flow along shortest paths is repeated until $b^2 = 0$. Then, the third phase begins.

In the second phase, the algorithm corrects the third component of arc capacities in the following manner:

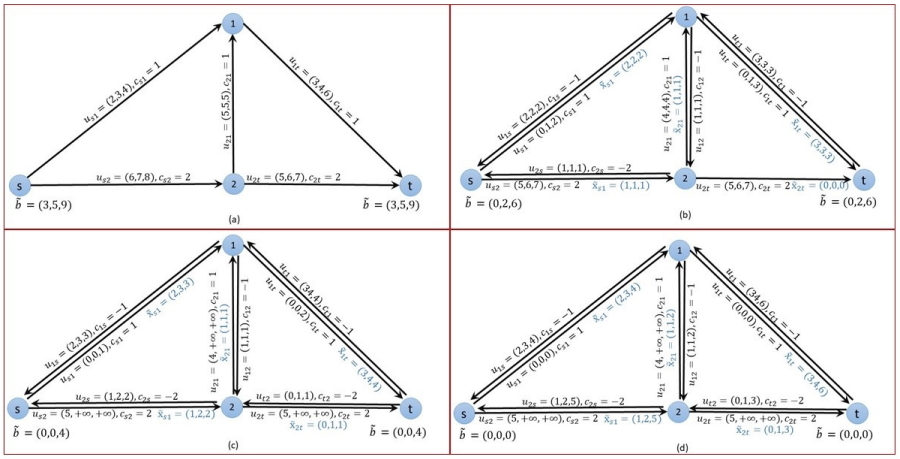


Figure 1: (a)An instance of the problem. (b) The first phase of the algorithm. (c) The second phase of the algorithm. (d) The third phase of the algorithm.

- if $x_{ij}^2 = u_{ij}^2$, then u_{ij}^3 are not changed;
- if $x_{ij}^2 < u_{ij}^2$, then $u_{ij}^3 = +\infty$.

Then, it finds a shortest path from s to t in the residual network and it sends $(0, 0, f_3)$ units of flow along it where $f_3 = \min\{b^3, \min_{(i,j)\in P} u_{ij}^3\}$. The algorithm repeats this procedure until $b^3 = 0$. It is notable that the algorithm may terminate in each phase since there is no path from s to t in the residual network for sending flow along it. In this situation, the problem is infeasible because it cannot satisfy the demand of t. Let us illustrate the algorithm by an example.

Example 2.1. *This example contains a source s and a sink t with fuzzy demand and supply $(3, 5, 9)$ (see Figure 1.a for observing complete data). Figure 1.b, 1.c and 1.d respectively show the first, second and third phases of the algorithm. Fuzzy flows are highlighted by blue color. The optimal value is* $\tilde{z} = 1 \times (2, 3, 4) + 2 \times (1, 2, 5) + 1 \times (1, 1, 2) + 1 \times (3, 4, 6) + 2 \times (0, 1, 3) = (8, 14, 28)$.

Let us now state the complexity of the algorithm. Since the algorithm sends at least one unit of flow from s to t in each iteration, it follows that the number of iterations is at most equal to b^3. On the other hand, the algorithm finds a shortest path with respect to nonnegative arc costs in each iteration. So its worst-case complexity is $O(b^3 S(|V|, |A|))$ in which $S(|V|, |A|)$ is the complexity of finding a shortest path in a network containing $|V|$ nodes as well as $|A|$ arcs. For instance, if one uses the well-known dijkstra algorithm which has the complexity $O(|V|^2)$, then our proposed algorithm runs in $O(b^3|V|^2)$ time.

References

Ahuja R.K., Magnanti T.L. and Orlin J.B. (1993), *Network Flows: Theory, Algorithms, and Applications*, Prentice Hall.

Ghatee, M. and Hashemi S.M. (2008), Generalized minimal cost flow problem in fuzzy nature: an application in bus network planning problem, *Applied Mathematical Modelling* 32, 2490-2508.

Ghatee M. and Hashemi S.M. (2009), Application of fuzzy minimum cost flow problems to network design under uncertainty, *Fuzzy Sets and Systems* 160, 3263-3289.

Mahdavi-Amiri N. and Nasseri S.H. (2007), Duality results and a dual simplex method for linear programming problems with trapezoidal fuzzy variables, *Fuzzy Sets and Systems* 158, 1961-1978.

Nasseri S.H., Ebrahimnejad A. and Mizuno S. (2010), Duality in fuzzy linear programming with symmetric trapezoidal numbers, *Applications and Applied Mathematics* 5, 1467-1482.

روش چند هدفه در برآورد ضرایب مدل رگرسیون خطی با ورودی- خروجی فازی LR

محمد رضا ربیعی[1]، نسرین فلاحتی نژاد، دلارا کرباسی

شاهرود، دانشگاه صنعتی شاهرود، دانشکده علوم ریاضی، گروه آمار

چکیده: مسأله‌ی تحلیل رگرسیون فازی در مجموعه‌های فازی مورد مطالعه قرار گرفته است. در این مقاله، مدل رگرسیون خطی برای مطالعه‌ی وابستگی متغیر پاسخ فازی LR بر روی یک مجموعه از متغیر های توضیحی فازی LR به کمک روش چند هدفه فازی بر روی پارامترها معرفی شده است. مزیت مهم این روش آن است که دو مدل دیگر رگرسیون فازی یعنی زمانی که فقط ضرایب فازی LR و یا زمانی که فقط متغیر توضیحی فازی LR باشند را در بر می‌گیرد. با معیار نیکویی برازش مناسب مدل فوق ارزیابی می‌شود. استراتژی فوق با جزئیات و با مراجعه به کاربرد داده‌های حقیقی گردآوری شده، در مراجع پیشین در ساختار یک مطالعه‌ی میدانی تشریح می‌شود.

واژه‌های کلیدی: رگرسیون خطی چندگانه، روش کمترین توان‌های دوم، متغیر پاسخ فازی LR، متغیر توضیحی فازی LR، نیکویی برازش.

[1] **نام ارائه دهنده‌ی مقاله :** rabiei1354@yahoo.com

تأثیر عملگرهای محاسباتی در محاسبه‌ی قابلیت اطمینان سیستم در شرایط نادقیق

رضا زارعی[1]

گروه آمار، دانشکده علوم ریاضی، دانشگاه گیلان

چکیده: تحلیل قابلیت اطمینان سیستم در شرایطی که احتمال کارکرد اجزای آن به‌صورت کمیتی نادقیق گزارش شده است، به عوامل مختلفی مانند صورت‌بندی ابهام موجود، فرم تابع عضویت و نوع عملگرهای محاسباتی بستگی دارد. در این مقاله به بررسی تأثیر این عوامل در تعیین قابلیت اطمینان سیستم‌ها در شرایط نادقیق می‌پردازیم. در این راستا، ابتدا تحلیل درخت عیب سیستم تحت بررسی ارائه می‌شود. سپس با استفاده از دو رویکرد محاسبات بازه‌های و استفاده از تی نرم دراستیک و در نظر گرفتن اعداد فازی و فازی شهودی برای صورت‌بندی احتمالات نادقیق اجزاء، قابلیت اطمینان سیستم محاسبه و تحلیل می‌گردد.

واژه‌های کلیدی: قابلیت اطمینان سیستم، محاسبات بازه‌ای، آلفا برش، تی نرم دراستیک، عدد فازی.

کد موضوع‌بندی ریاضی (۲۰۱۰): 68M15، 03E72.

[1] **نام ارائه دهنده‌ی مقاله:** rzarei@guilan.ac.ir

یک متر جدید برای محاسبه‌ی فاصله‌ی بین هر دو عدد فازی LR

مجتبی کاشانی[1]، محمد آرشی، محمد رضا ربیعی

دانشکده علوم ریاضی، دانشگاه صنعتی شاهرود، شاهرود

چکیده: براي سنجش فاصله‌ی بین دو عدد فازی مترهای مختلفی مطرح شده که تفاوت‌هایی از دیدگاه‌های مختلف به همراه دارد. در این مقاله برآنیم با یك دیدگاه متفاوت مبتنی بر αـ برش، یك متر جدید جهت اندازه گیری فاصله‌ی بین دو عدد فازی LR ارائه دهیم. خواص متر بودن را اثبات کرده، به عنوان کاربردی از آن در تحلیل رگرسیون فازی برای مقایسه‌ی فاصله‌ی بین خروجی‌های مشاهده شده و برآورد شده در کنار روش‌های پیشین تعیین فاصله خواهیم داشت.

واژه‌های کلیدی: اعداد فازی LR، αـ برش، فاصله، متر.

کد موضوع‌بندی ریاضی (۲۰۱۰): 28E10، 03E72.

[1] **نام ارائه دهنده‌ی مقاله :** kashani.mojtaba@yahoo.com

ارزیابی قابلیت سیستم فازی با استفاده از توزیع طول عمر رایلی فازی شهودی

محمد جواد نوراللهی قراخیلی[۱]

دانشکده علوم پایه، گروه آمار، دانشگاه آزاد اسلامی، واحد قائمشهر، ایران

چکیده: در این مقاله، قابلیت اطمیان فازی شهودی و میانگین زمان شکست فازی شهودی، تابع خطر فازی شهودی و αـ برش های آن‌ها را مورد بحث قرار می‌دهیم و این مشخصه‌ها را در توزیع رایلی فازی شهودی محاسبه می‌نمائیم. سپس قابلیت اطمینان سیستم‌های سری، موازی و k از m را ارزیابی می‌کنیم. نهایتاً توزیع رایلی فازی شهودی را با رویکرد جدید معرفی و به تبع آن مشخصه‌های قابلیت اطمینان (قابلیت اطمینان، نرخ خطر، میانگین زمان شکست) را مورد ارزیابی قرار می‌دهیم.

واژه‌های کلیدی: اعداد فازی شهودی، αـ برش، تابع قابلیت اطمینان فازی، تابع نرخ خطر فازی شهودی، توزیع رایلی، میانگین طور عمر شکست فازی شهودی.

[۱] **نام ارائه دهنده‌ی مقاله :** entesharatjavad94@gmail.com

Income inequality indices for fuzzy data

Zahra Behdani[1]

Department of Statistics Behbahan Khatam Alanbia University of Technology, Behbahan, Iran

Abstract: Income inequality indices are used by social scientists to measure the distribution of income and economic inequality among the participants in a particular economy, such as that of a specific country or of the world in general. But in some cases the concept of inequality is vague and thus cannot be measured as an exact concept. Therefore, fuzzy set theory provides naturally a useful tool for such positions. In this paper, we introduce the income inequality indices for fuzzy random variables. Some examples illustrating the computation of the fuzzy inequality index are also considered.

Keywords: Income inequality, Fuzzy number, Lorenz curve, Gini index.

Mathematics Subject Classification (2010): 62P20, 62E86.

[1]Speaker: zbehdani@yahoo.com

An overview of insurance risk model in random fuzzy environment

Sara Ghasemalipour[1], Behrouz Fathi-Vajargah

Department of Statistics, Faculty of Mathematical Sciences, University of Guilan, Iran

Abstract: In this paper, we introduce the insurance risk model in random fuzzy variable environment. We consider a new insurance risk model in which both the claim amount and premium are assumed to be random fuzzy variables. Some new theorems are stated and proved.

Keywords: Insurance risk model, Renewal process, Fuzzy variable.

[1]Speaker: s.ghasemalipour@gmail.com

Fuzzy linear regression models with fuzzy metric

R. Ghasemi[1], M.R. Rabiei, A. Nezakati

Faculty of Mathematical Sciences, Shahrood University of Technology, Shahrood, Iran

Abstract: In this paper, we introduce and study the concept of fuzzy metric space for fuzzy sets in the sense of George and Veeramani and define the fuzzy normed space by using the definition of Saadati and Vaezpour. Also, we obtain the estimation of the regression model parameters by using this metric (fuzzy metric for fuzzy set), while the dependent variable and coefficients are triangular fuzzy numbers. The method is constructed on the basis of maximizing the fuzzy metric between observed and estimated spread values.

Keywords: Fuzzy metric space, Fuzzy normed space, Regression model parameters.
Mathematics Subject Classification (2010): 54A40, 54D35, 54E50.

[1]Speaker: rezaghng@gmail.com

On the non-parametric multivariate control charts in fuzzy environment

B. Sadeghpour Gildeh, Z. Abbasi Ganji

Department of Statistics, Faculty of Mathematical Sciences, Ferdowsi University of Mashhad, Mashhad, Iran

Abstract: Multivariate control chats are generally used in situations where the simultaneous monitoring or control of two or more related quality characteristics is necessary. In most processes in the real world, distribution of the process characteristics are unknown or at least unnormal, so the non-parametric or distribution-free charts are desirable. Most non-parametric statistical process-control techniques depend on ranks. In this survey, we apply the fuzzy set theory to deal with the circumstances that the values of each characteristic are presented in linguistic form, so we propose non-parametric multivariate control charts based on sign and Wilcoxon signed-rank tests. The performance of the proposed charts is investigated in a simulation study. Numerical examples are used to demonstrate the effectiveness and performance of the proposed charts.

Keywords: Multivariate control charts, Non-parametric tests, Fuzzy logic.