دانشگاه تبریز

دانشکده علوم ریاضی

گروه آمار

پایاننامه

برای دریافت درجه کارشناسی ارشد در رشته‌ی

آمار، گرایش آمار ریاضی

عنوان

تابع چگالی با کمترین فاصله کای‌دو با داشتن تابع چگالی پیشین و گشتاورها

استاد راهنما

دکتر حسین بیورانی

استاد مشاور

دکتر غلامحسین غلامی

پژوهشگر

سولماز سیف الهی کوسه‌لر

۱۳۹۰


بسم الله الرحمن الرحیم

بسم الله الرّحمن الرّحیم

سپاس خدایی را که سخنوران در ستودن او بمانند و شمارگران شمردن نعمتهای او ندانند و کوشندگان، حقّ او را گزاردن نتوانند. خدایی که پای اندیشه ی تیزگام در راه شناسایی او لنگ است و سر فکرت ژرف رو به دریای معرفتش بر سنگ. صفت های او تعریف ناشدنی است و به وصف در نیامدنی و در وقت ناگنجیدنی و به زمانی مخصوص نابودنی. به قدرتش خلایق را بیافرید، و به رحمتش بادها را بپراکنید، و با خرسنگ ها لرزه ٔ زمین را در مهار کشید.

گواهی می دهم که خدا یکتاست، انبازی ندارد و بی همتاست. گواهی از روی اعتقاد و ایمان، بی آمیغ برآمده از امتحان؛ و گواهی می دهم که محمّد (ص) بنده ٔ او و پیامبر اوست. او را بفرستاد با دینی آشکار، و با نشانه هایی پدیدار، و قرآنی نبشته در علم پروردگار. که نوری است رخشان و چراغی است فروزان، و دستورهایش روشن و عیان. تا گرد دودلی از دلها بزداید، و با حجّت و دلیل ملزم فرماید.

پاک خدایا! چه بزرگ است آنچه می بینیم از خلقت تو؛ و چه خرد است، بزرگی آن در کنار قدرت تو؛ و چه با عظمت است آنچه می بینیم از ملکوت تو، و چه ناچیز است برابر آنچه بر ما نهان است از سلطنت تو، و چه فراگیر است نعمت تو در این جهان؛ و چه اندک است در کنار نعمتهای آن جهان.

خدایا! اگر در پرسش خود درمانم یا راه پرسیدن را ندانم، صلاح کارم را به من نما و دلم را بدانچه رستگاری من در آن است متوجه فرما! که چنین کار از راهنمایی های تو ناشناخته نیست و از کفایت های تو نه.

از فرمایشات حضرت علی (ع)

تقدیم به:

پدرم، مادرم،

برادرم، خواهرانم

و اسمای عزیزم

بنام خدا

و لَمْ يَشْكُر الْمَخلوق لَمْ يَشْكُر الْخالِقْ.

سپاس و ستایش پروردگار متعال را که به اینجانب توفیق تلاش در راه کسب علم و دانش را عطا فرمود. امیدوارم بتوانم آموخته‌هایم را در راه پیشرفت علمی وطن خویش مورد استفاده قرار دهم.
در آغاز وظیفه‌ی خود می‌دانم از زحمات بی‌دریغ استاد راهنمای خود، جناب آقای دکتر **حسین بیورانی** صمیمانه تشکر و قدردانی نمایم که قطعاً بدون راهنمایی‌های ارزنده‌ی ایشان این مجموعه به انجام نمی‌رسید.
از جناب آقای دکتر **غلامحسین غلامی** که زحمات مطالعه و مشاوره‌ی این رساله را تقبل فرمودند و در آماده‌سازی این پایان‌نامه به نحو احسن اینجانب را مورد راهنمایی قرار دادند، کمال امتنان را دارم.
از جناب آقای .... که داوری این رساله را با نهایت دقت و صرف وقت زیاد انجام دادند، تشکر می‌نمایم.

در پایان از کلیه اعضای خانواده‌ام که در تمامی مراحل همواره یار و مشوق من بوده‌اند، سپاسگزاری می‌کنم.

سولماز سیف‌الهی کوسه‌لر
۱۳۹۰

نام خانوادگی دانشجو: سیف الهی کوسه‌لر نام: سولماز

عنوان: تابع چگالی با کمترین فاصله کای‌دو با داشتن تابع چگالی پیشین و گشتاورها

استاد راهنما : دکتر حسین بیورانی

استاد مشاور : دکتر غلامحسین غلامی

مقطع تحصیلی: کارشناسی ارشد رشته: آمار گرایش: آمار ریاضی دانشگاه تبریز

دانشکده علوم ریاضی تاریخ فارغ التحصیلی: ۱۳۹۰ تعداد صفحات: ۱۲۶

کلید واژه‌ها: اطلاع کلبک-لایبلر، تابع چگالی احتمال، توزیع دیریکله، توزیع گاما، توزیع نمایی، توزیع وایبل، فاصله کای‌دو، گشتاور

چکیده

برآورد تابع چگالی احتمال از مباحث کاربردی در آمار می‌باشد، که در تاکنون روش‌های متفاوتی برای آن ارائه گردیده است. یک دیدگاه که در زمینه برآورد تابع چگالی وجود دارد، برآورد تابع چگالی با داشتن اطلاعات روی گشتاورهای توزیع می‌باشد. از جمله روش‌هایی که برای این منظور وجود دارد می‌توان به اصل ماکزیمم آنتروپی، اصل کمترین تفاوت اطلاعات و اصل کمترین فاصله کای‌دو اشاره کرد.

در این پایاننامه به بررسی اصل کمترین فاصله کای‌دو که توسط کومار و تانجا [۱۷، ۱۸] معرفی شده است می‌پردازیم. نخست به معرفی فاصله کای‌دو و اصل کمترین فاصله کای‌دو می‌پردازیم و سپس روش برآورد تابع چگالی را با داشتن تابع چگالی اولیه و اطلاعاتی در مورد تابع چگالی مجهول را بیان می‌کنیم. در ادامه برای زمانی که تابع چگالی اولیه یا مشاهده شده توزیع گاما [۱۹] و وایبل و همچنین اطلاعات گشتاوری توزیع، اطلاعاتی در مورد میانگین هندسی، حسابی و یا واریانس می‌باشد تابع چگالی با کمترین فاصله کای‌دو را ارائه می‌دهیم و با مثال‌های عددی نتایج بدست آمده را تشریح

خواهیم نمود.
در پایان به تعمیم روش کومار و تانجا برای برآورد تابع چگالی احتمال توأم با کمترین فاصله کای‌دو خواهیم پرداخت. همانند حالت تک‌متغیره، ضمن تعریف فاصله کای‌دو و اصل کمترین فاصله کای‌دو در حالت دومتغیره، روش برآورد تابع چگالی احتمال توأم را بیان و سپس با در نظرگرفتن توزیع دیریکله و حاصلضرب دو توزیع نمایی به عنوان توزیع اولیه و اطلاعات روگشتاور حاصلضرب یعنی $E(XY)$ تابع چگالی را بدست می‌آوریم.

فهرست مطالب

فهرست اشکال

فهرست جداول

مقدمه

از مباحث کاربردی در آمار، برآورد تابع چگالی است که تاکنون کارهای زیادی روی این موضوع انجام گرفته است. از جمله‌ی این کارها می‌توان به روش کرنل [۱] و بوت‌استرپ[۲] برای برآورد تابع چگالی اشاره نمود.

در بررسی یا آزمایش‌های پی در پی، اطلاعاتی در باره توزیع احتمال از طریق نمونه‌گیری و برآورد بر حسب میانگین و یا واریانس متغیر تصادفی معمولاً در دسترس است. لذا دیدگاهی که می‌تواند در زمینه برآورد تابع چگالی ارائه شود این است که چگونه می‌توان تابع چگالی را پیدا نمود که در این اطلاعات یا محدودیت‌ها صدق نماید. در روش‌های بیان شده، یعنی روش کرنل و بوت‌استرپ نمی‌توان این اطلاعات را روی تابع چگالی برآورد شده اعمال کرد. بنابراین لازم است تا روش‌های دیگری برای برآورد تابع چگالی جستجو کرد.

از جمله روش‌هایی که در این زمینه استفاده می‌شود می‌توان به روش ماکزیمم درستنمایی و روش کمترین مربعات اشاره نمود. علاوه بر این روش‌ها جی‌آسو[۳] [۵] به تشریح بحث انحرافات وزنی با داشتن مقدار میانگین متغیر تصادفی و تعیین بهترین توزیع احتمالی با در نظر گرفتن انحرافاتی مانند کای‌دوی پیرسن [۲۰]، کای‌دوی کاهشی نیمن [۴]، فاصله کلبک–لایبلر [۱۵] و شاخص کلموگروف [۳] پرداخته است. در این میان اصل ماکزیمم آنتروپی (MEP) [۴] و اصل کمترین تفاوت اطلاعات (MDIP) [۵] یا اصل کمترین آنتروپی ضربی به عنوان روش‌های شناخته شده برای تعیین و مشخص کردن اکثر توزیع‌های احتمالی نااریب تک متغیره و چند متغیره می‌باشند.

---

[۱] Kernel

[۲] Bootstrap

[۳] Guiasu

[۴] Maximum Entropy Principle

[۵] Minimum Discrimination Information Prenciple

اصل ماکزیمم آنتروپی که توسط جاینز [6] [۷] مطرح شده است، برای پیدا کردن تابع چگالی احتمال علاوه بر داشتن اطلاعات روی گشتاورها از آنتروپی شانون [7] به عنوان معیاری برای شناسایی و مشخص نمودن تابع چگالی به کار برده است. روش کار به این ترتیب است که بین تمام توابع چگالی که در اطلاعات داده شده به صورت گشتاوری، صدق می‌کنند تابع چگالی انتخاب می‌شود که آنتروپی شانون را ماکزیمم کند.

از جمله کسانی که روی این اصل بررسی‌هایی انجام داده‌اند می‌توان به شور و جانسون[8] [۲۱]، کاوامورا و ایواز[9] [۱۳]، کاپور[10] [۱۱، ۱۲]، کاگن و همکاران[11] [۱۰]، جاینز[12] [۸]، جاستیک[13] [۹]، گخل[14] [۴] و ... اشاره نمود. تنها ایرادی که این روش داشت این بود که غالباً توزیع یکنواخت را به عنوان توزیع مناسب برآورد می‌کرد و در آن نمی‌توان از توزیع اولیه استفاده کرد. اما در روش دوم یعنی اصل کمترین آنتروپی ضربی که کلبک [۱۵] ارائه کرده و در آن از فاصله کلبک–لایبلر [۱۶] برای برآورد تابع چگالی استفاده می‌شود، می‌توان از توزیع اولیه و هم از اطلاعات به شکل گشتاوری استفاده نمود.

کامپبل[15] [۳] طی بررسی‌های خود دریافت که تابع چگالی بدست آمده از طریق اصل گاوس با تابع چگالی که از اصل کمترین آنتروپی ضربی بدست می‌آید معادل و همچنین کمترین آنتروپی ضربی با ماکزیمم تابع درستنمایی معادل می‌باشند. همچنین کساوان و کاپور[16] [۱۴] به تعمیم روش ماکزیمم آنتروپی و کمترین تفاوت اطلاعاتی پرداخته است.

علاوه بر این دو اصل کومار و تانجا[17] [۱۷] با در نظر گرفتن اصل کمترین فاصله کای‌دو، با داشتن توزیع اولیه و اطلاعاتی در باره گشتاورهای متغیر تصادفی به تعیین بهترین توزیع احتمال پرداخته و تابع چگالی را با استفاده از توزیع‌های گسسته‌ای همچون توزیع هندسی، پواسون و دوجمله‌ای برآورد

[6] Jaynes

[7] Shannon entropy

[8] Shore and Jonhson

[9] Kawamura and Iwase

[10] Kapur

[11] Kagan et al

[12] Jaynes

[13] Justice

[14] Gokhle

[15] Campbell

[16] Kesavan and Kapur

[17] Kumar and Taneja

کردند.

در این پایاننامه نحوه برآورد تابع چگالی با استفاده از اصل کمترین فاصله کای‌دو را بطور اجمالی مورد بررسی قرار می‌دهیم. برای این منظور در فصل اول به بیان تعاریف و مفاهیم اولیه مورد نیاز در فصل‌های بعدی می‌پردازیم و در فصل دوم مقاله کومار و تانجا [۱۸] بررسی خواهیم کرد. ابتدا اصل کمترین فاصله کای‌دو را تعریف کرده و سپس تابع چگالی با کمترین فاصله کای‌دو را زمانی که اطلاعات داده شده یا محدودیت‌های روی تابع چگالی، شرایطی روی میانگین هندسی، حسابی و (یا) واریانس می‌باشد بررسی نموده و در ادامه نتایج بدست آمده را با در نظر گرفتن توزیع یکنواخت به عنوان توزیع اولیه تشریح می‌کنیم.

در فصل سوم بر روی مقاله کومار [۱۹] متمرکز می‌شویم، فرض می‌کنیم که توزیع مشاهده شده، توزیع گاما می‌باشد. نخست فرض می‌کنیم که اطلاع در مورد گشتاور، مربوط به گشتاور مرتبه n-ام تابع چگالی $f(x)$ می‌باشد، تابع چگالی با کمترین فاصله کای‌دو را برای این حالت بررسی می‌نماییم و سپس حالت خاصی از آن را مانند داشتن میانگین حسابی و گشتاور مرتبه دوم بیان می‌کنیم. در ادامه به نحوه پیدا کردن تابع چگالی با کمترین فاصله کای‌دو را برای زمانی که میانگین هندسی و میانگین حسابی همراه با گشتاور مرتبه دوم (یا واریانس) مشخص باشند، می‌پردازیم. در پایان با مثال‌های عددی برآوردهای بدست آمده را تشریح و با استفاده از نمودار، تفاوت تابع چگالی احتمال برآورد شده را با داشتن اطلاعات متفاوت بطور واضح نمایش خواهیم داد. در این مقاله اشکالاتی یافت شده است که تصحیح آنرا به همراه شماره عبارات در ضمیمه پایاننامه عنوان می‌کنیم.

به دلیل استفاده‌ای که توزیع وایبل در اغلب مباحث کاربردی مانند کنترل کیفیت آماری [۱۸]، قابلیت اعتماد [۱۹] و قابلیت دسترس‌پذیری [۲۰] دارد لذا در فصل چهارم به برآورد تابع چگالی احتمال با استفاده از اصول کمترین فاصله‌ی کای‌دو در حالتی که توزیع احتمال مشاهده شده، توزیع وایبل می‌باشد و اطلاعاتی نیز در مورد گشتاورهای متغیر تصادفی در اختیار داریم، می‌پردازیم. مانند فصل سوم نخست در حالت کلی یافتن تابع چگالی احتمال با کمترین فاصله کای‌دو را برای موردی مطرح می‌کنیم که گشتاور مرتبه n-ام مشخص باشد و سپس برای حالت خاص مثل داشتن میانگین حسابی و (یا) گشتاور مرتبه دوم می‌پردازیم. در نهایت مطابق فصل‌های قبل سعی می‌کنیم با مثال‌های عددی برآوردهای بدست آمده را تشریح و سپس نمودار هر کدام را برای نشان دادن تفاوتشان رسم خواهیم نمود.

در فصل آخر به تعمیم روش کومار و تانجا برای ساخت تابع چگالی توأم با کمترین فاصله کای‌دو

[۱۸] Statistical quality control

[۱۹] Reliability

[۲۰] Availability

به کمک تابع چگالی اولیه و گشتاورها خواهیم پرداخت. بدین منظور لازم است فاصله کایدو را برای حالت دومتغیره تعریف کنیم. برای برآورد تابع چگالی توأم ابتدا اصول کمترین فاصله کایدو بین دو تابع چگالی توأم را تعریف کرده و سپس به تشریح روش پیدا کردن تابع چگالی توأم با کمترین فاصله کایدو خواهیم پرداخت. مانند فصل‌های قبل با در نظر گرفتن دو توزیع به عنوان توزیع مشاهده شده به تشریح برآورد بدست آمده می‌پردازیم و در نهایت با مثال‌های عددی نتایج بدست آمده را بررسی می‌نماییم.

خاطر نشان می‌شویم که کلیه شبیه‌سازی این پایاننامه با استفاده از **نرم‌افزار R** انجام گرفته است.

فصل ۱

تعاریف و مفاهیم مقدماتی

در این فصل به تعریف مفاهیم اولیه که طی فصل‌های بعدی برای تعیین تابع چگالی با کمترین فاصله کای‌دو و بیان ویژگی‌های آن مورد نیاز است، می‌پردازیم.

تعریف ۱.۱ تابع چگالی احتمال [۱]
اگر X یک متغیر تصادفی پیوسته باشد، تابع چگالی $f(x)$ را تابع چگالی احتمال X می‌نامند هرگاه
۱) به ازای هر $x \in \mathbb{R}_X$; $f(x) > ۰$
۲) برای هر $A \subseteq \mathbb{R}_X$; $p(A) = \int_A f(x)\, dx$
۳) $\int_{\mathbb{R}_X} f(x)\, dx = ۱$
که در آن $\mathbb{R}_X$ تکیه‌گاه متغیر تصادفی X می‌باشد.

تعریف ۲.۱ توأم تابع چگالی احتمال [۲]
اگر (X,Y) متغیرهای تصادفی توأماً پیوسته باشند، تابع چگالی $f(x,y)$ را تابع چگالی احتمال (X,Y) می‌نامیم هر گاه:
۱) $f(x,y) \geq ۰$
۲) $\int_X \int_Y f(x,y)\, dxdy = ۱$.

تعریف ۳.۱ فاصله کای‌دو[۳]
اگر $f(x)$ و $g(x)$ دو تابع چگالی احتمال باشند، فاصله کای‌دو بین این دو تابع چگالی احتمال به صورت زیر تعریف می‌شود:

$$\chi^۲(f,g) = \int_X \frac{f^۲(x)}{g(x)}\, dx - ۱.$$

تعریف ۴.۱ اطلاع کلبک-لایبلر [۴]
اگر $f(x)$ و $g(x)$ دو تابع چگالی احتمال باشند، اطلاع کلبک-لایبلر بین این دو تابع چگالی احتمال به صورت زیر تعریف می‌شود:

[۱] Probability density function
[۲] Joint probability density function
[۳] Chi-square divergency
[۴] Kullback-leibler information

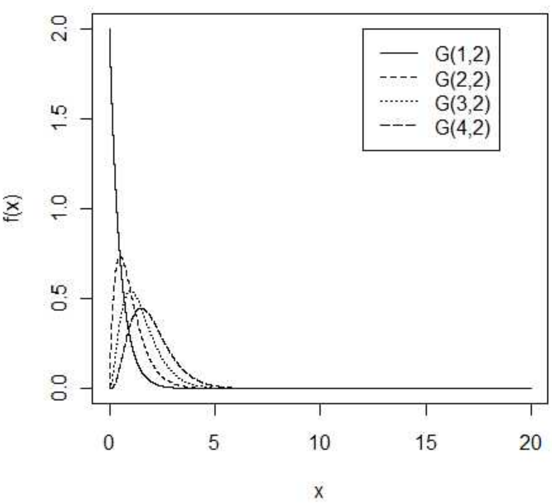
$$K(f,g)=\int_X f(x)ln\frac{f(x)}{g(x)}\,dx.$$

تعریف ۵.۱ توزیع گاما [۵]

توزیع گاما به عنوان توزیع طول عمر یا زمان انتظار مطرح می‌شود. تابع چگالی احتمال این توزیع به صورت زیر است:

$$f(x)=\frac{۱}{\Gamma(\alpha)\beta^{\alpha}}x^{\alpha-۱}e^{-\frac{x}{\beta}};\quad x>۰$$

که در آن $\alpha>۰$ و $\beta>۰$ است و $\Gamma(.)$ تابع گاما نامیده می‌شود. در زیر نمودار تابع چگالی گاما با پارامترهای متفاوت برای مقایسه رسم شده‌اند.



شکل ۱.۱: چگالی‌های توزیع گاما با $\beta=۲$ و α مختلف

[۵] Gamma distribution

تعریف ۶.۱ توزیع نمایی [۶]
اگر متغیر تصادفی X دارای تابع چگالی زیر باشد:

$$f(x) = \frac{1}{\lambda}e^{-x/\lambda}; \quad x > 0$$

که در آن $\lambda > 0$ است، در آن صورت گوییم X دارای توزیع نمایی است. توزیع نمایی حالت خاصی از توزیع گاماست که در آن $\alpha = 1$ و $\beta = \lambda$ می‌باشد.

تعریف ۷.۱ توزیع وایبل [۷]
توزیع وایبل در اغلب مباحث کاربردی مانند کنترل کیفیت آماری[۸]، قابلیت اعتماد [۹] و قابلیت دسترس پذیری [۱۰] مورد استفاده قرار می‌گیرد.
گوییم X دارای توزیع وایبل است، اگر تابع چگالی احتمال آن به فرم زیر باشد:

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

که در آن $\alpha > 0$ و $\beta > 0$ مقادیر ثابتی هستند. در زیر نمودار تابع چگالی وایبل با پارامترهای متفاوت برای مقایسه رسم شده‌اند.

[۶] Exponential distribution
[۷] Weibull distribution
[۸] Statistical quality control
[۹] Reliability
[۱۰] Availability