بسم الله الرحمن الرحیم

عنوان پایان نامه

برآوردهای بیز، شبه‌بیز و بیز تجربی
در
جدولهای توافقی

نام مؤلف:

حسنعلی سعادت

پایان نامه کارشناسی ارشد
آمار ریاضی

تیرماه ۱۳۷۳

دانشکده علوم

دانشگاه فردوسی مشهد

شماره ....

تاریخ ....

پیوست ....

صورتجلسه دفاع فوق لیسانس آمار ریاضی

در تاریخ ۷۳/۴/۲۳ ~~خانم~~ آقای حسنعلی سعادت دانشجوی کارشناسی ارشد

رشته آمار دانشکده علوم دانشگاه فردوسی "مشهد" از رساله خود تحت عنوان :

" برآوردهای بیز، شبه بیز وبیزتجربی درجدولهای توافقی "

باارائه خلاصه ای ازکارانجام شده و پاسخ به سئوالات داوران دفاع نمودند .

این پایان نامه با نمره ۱۹/ معادل عالی قبول ~~رد~~ گردید .

استاد راهنما : دکترعلی مشکانی

استاد مشاور : دکتر ناصررضاارقامی

اعضاء هیأت داوران :

۱ - دکترابوالقاسم بزرگنیا

۲ -

علی مشکانی

مدیرگروه آمار

نشانی : مشهد - گروه آمار دانشکده علوم دانشگاه فردوسی "مشهد"   تلفن : ۱۲-۷۱۵۰۱۱

تقدیم به :

همه کسانی که بنحوی بر من حقی داشته و دارند،

بویژه پدر و مادرم.

سپاسگزاری

بدین وسیله از آقای دکتر علی مشکانی استاد راهنمای رساله که همواره مشوق و راهنمای پرحوصله‌ای در طول تدوین این رساله بوده‌اند و آقای دکتر ناصر رضا ارقامی استاد مشاور رساله بخاطر راهنمایی‌هایشان تشکر و قدردانی می‌کنم.

مراتب امتنان خود را از آقای دکتر ابوالقاسم بزرگ نیا داور رساله که در تصحیح آن مرا راهنمائی کرده اند،ابراز می‌دارم.

از آقایان علیرضا وطن‌دوست و مجید سرمد مسئولین کامپیوتر گروه آمار دانشگاه فردوسی بواسطه همکاری صمیمانه شان سپاسگزاری می‌کنم.

از آقای رسول اتحاد مسئول کتابخانه گروه ریاضی دانشگاه فردوسی بواسطه همکاری ایشان در تهیه بعضی منابع تشکر می‌کنم.

بی‌تردید رساله حاضر، دارای کاستی‌هایی است ازکسانی که این موارد رایادآوری کنند سپاسگزار خواهم بود.

حسنعلی سعادت

چکیده :

در یک جدول توافقی با ابعاد زیاد، تعداد خانه ها اغلب بسیار بالا است در حالی که متوسط تعداد مشاهدات در هر خانه در بسیاری از مسائل عملی کوچک است، و بنابراین بسیاری از خانه ها می‌توانند درایه های صفر داشته باشند. مایلیم که احتمالهای مربوط به این خانه ها را برآورد کنیم. اطلاعات اضافی در بارهٔ این احتمالها ممکن است از توزیع عمومی شمارشها یا ازکناره‌ها در دسترس باشد. رهیافتهای بیزی روشهایی را برای برآورد کردن این احتمالها عرضه می‌کنند، اما مسئله انتخاب پارامتر توزیع پیشین را به عهدهٔ استفاده کننده می‌گذارد، کاری که ممکن است او برای انجام آن مجهز نباشد. تحقیقات نظری، با نشان دادن بعضی روشها به خواننده که وی را از خطاهای بد محفوظ خواهند داشت، می‌تواند خواننده را در انتخاب پارامترها یاری کند. دراین رساله برآوردگرهای بیز و شبه بیز و بیزتجربی رابرای احتمالهای خانه‌ای درجدولهای توافقی مطرح کرده واین برآوردگرها را باهم مقایسه می کنیم.

فهرست

عنوان — صفحه

فصل اول: روشهای برآورد بیز و بیز تجربی

فصل دوم: برآورد شبه بیز در جدولهای توافقی

فصل سوم: برآورد بیز تجربی در جدولهای توافقی

پیوست

مقدمه :

در فصل اول این رساله ابتدا به معرفی آمار بیز می‌پردازیم . سپس برآورد بیزی پارامترهای یک خانواده از توابع چگالی احتمال را در حالتهای مختلف مطرح نموده و یک روش کلی برای برآورد بیزی پارامتر یک خانواده نمایی یک پارامتری وقتی تابع زیان توان دوم خطا باشد را مورد بررسی قرار می‌دهیم.

بیز تجربی و برآورد به روش بیز تجربی را در قسمت بعد بطورگسترده‌ای مورد بحث و بررسی قرار می دهیم و در انتها روشهای برآورد تابع چگالی را عنوان می‌کنیم.

چون در یک جدول توافقی با ابعاد زیاد، تعداد خانه ها اغلب بسیار بالا است در حالی که متوسط تعداد مشاهدات در هر خانه در بسیاری از مسائل عملی کوچک است، و بنابراین بسیاری از خانه ها می‌توانند درایه های صفر داشته باشند. مایلیم که احتمالهای مربوط به این خانه ها را برآورد کنیم. اطلاعات اضافی دربارهٔ این احتمالها ممکن است از توزیع عمومی شمارشها یا ازکناره‌ها در دسترس باشد. رهیافتهای بیزی روشهایی را برای برآورد کردن این احتمالها عرضه می‌کنند، اما مسئله انتخاب پارامتر توزیع پیشین را به عهدهٔ استفاده کننده می‌گذارد، کاری که ممکن است او برای انجام آن مجهز نباشد. تحقیقات نظری، با نشان دادن بعضی روشها به خواننده که وی را از خطاهای بد محفوظ خواهند داشت، می‌تواند خواننده را در انتخاب پارامترها یاری کند. دراین فصل دوم ابتدا به معرفی توزیع دریکله و خواص آن می پردازیم . برآوردگرهای بیز و شبه بیز را برای احتمالهای خانه‌ای در جدولهای توافقی مطرح کرده و این برآوردگرها را با هم مقایسه می کنیم .

در بخش (۳-۱) دو خانواده پیشین راکه لئونارد[1](۱۹۷۵) و لرد[2] (۱۹۷۸) وآلبرت[3]و کوپتا[4] (۱۹۸۳ و ۱۹۸۲) ارائه دادند، معرفی می‌کنیم و در بخش (۳-۲) بیز تجربی جدید برآوردگرهای بیزی را با استفاده از کلاس توزیعهای مخلوط دریکله گسترش می‌دهیم. نشان می‌دهیم این برآوردگرها

1-Leonard
2-Laird
3-Albert
4- Gupta

رفتار مشابهی نسبت به برآوردگرهای پیش آزمون دارند و برای هر دو نوع برآوردگرها اندازه فشردگی به سمت برآوردهای استقلال بستگی به توافق داده‌ها با الگوی استقلال دارد. در قسمت (۳-۳) رفتار برآوردهای جدید را بطور عددی برای بسیاری از جدولهای ۲×۲ تحلیل کرده و این برآورد را با برآوردهای بیز تجربی لرد(۱۹۷۸) که با استفاده از کلاسی توزیعهای نرمال لگ - خطی[1] گسترش یافته، مقایسه می‌کنیم. یک ارزیابی مختصر از متوسط عملکرد برآوردگرهای بیز تجربی و برآوردگرهای بیزی در آزمایشهای طولانی را در قسمت (۳-۴) مورد بررسی قرار می‌دهیم. این فصل را در قسمت (۳-۵) با نشان دادن اینکه چگونه توزیع پیشین مخلوط دریکله را می‌توان برای مدل بندی مفهوم استقلال شرطی در یک جدول سه طرفه بکار برد،ادامه می‌دهیم و در قسمت (۳-۶) که مقایسه‌ای بین انواع برآوردگرهای مطرح شده است، پایان می‌دهیم.

1-linear- log

فصل اول:

روشهای برآورد بیز

و

بیز تجربی

مقدمه:

در این فصل ابتدا به معرفی آمار بیز می‌پردازیم. سپس برآورد بیزی پارامترهای یک خانواده از توابع چگالی احتمال را در حالتهای مختلف مطرح نموده و یک روش کلی برای برآورد بیزی پارامتر یک خانواده نمایی یک پارامتری وقتی تابع زیان توان دوم خطا باشد را مورد بررسی قرار می‌دهیم. بیز تجربی و برآورد به روش بیز تجربی را در قسمت بعد بطورگسترده‌ای مورد بحث و بررسی قرار می دهیم و در انتها روشهای برآورد تابع چگالی را عنوان می‌کنیم.

۱ـ۱: معرفی آمار بیز

فرض کنید متغیر تصادفی X دارای تابع چگالی $f(x;\theta)$ باشد پارامتر θ را می توان اغلب بعنوان مقدار مشاهده شده یک متغیر تصادفی در نظر گرفت. بعنوان مثال در آزمون ورودی دانشگاه معیار اجازه ورود به دانشگاه نمره‌ای است که شخص درکنکورکسب می‌کند. اما منطقی بنظر می رسد که فرض کنیم هر شخص دارای یک توانائی واقعی به اندازه θ است که نمی‌توانیم آنرا مستقیماً اندازه‌گیری نمائیم و X نمره کنکور هر شخص، متغیری تصادفی با میانگین θ و واریانس

σ^2 است. با در نظرگرفتن اینکه توانائی اشخاص متفاوت است می‌توان θ را بـــــصورت یک متغیر تصادفی در نظر گرفت. بعنوان مثال دیگر فرض کنید توده‌ای به حجم معین از مـــحصولات یک خط تولید را در یک مقطع زمانی در نظرگرفته و نمونه‌ای به حجم N (بدون جـایگذاری) از آن استخراج نمائیم. همچنین فرض کنیذ θ نسبت اقلام معیوب در این خط تـولید و مـتغیر تـصادفی X نشان دهنده تعداد اقلام معیوب در نمونه n تائی از توده فوق‌الذکر باشد واضح است که تابع احتمال متغیر تصادفی X یک تابع احتمال فوق هندسی بصورت زیر می‌باشد.

$$P(X=x)=\frac{\binom{N\theta}{x}\binom{N-N\theta}{n-x}}{\binom{N}{n}} \qquad \max[n-N(1-\theta),0]\leqslant x\leqslant \min(N\theta,n)$$

طبیعی است در اثر فرسوده شدن ماشینهای خط تولید و عوامل دیگر نسبت اقلام معیوب (θ) نوسان دارد و نمی توان همیشه آنرا یک مقدار ثابت تلقی کرد.

آماری راکه با این دید پایه گذاری می‌شود امار بیز می‌نامند.

مجدداً متغیر تصادفی X با تابع چگالی احتمال $f(x;\theta)$ را در نـــظر مـــی‌گیریم. چـون می‌خواهیم θ را بعنوان مقدار مشاهده شده متغیر تصادفی Θ در نظر بگیریم لذا این چگـالی را بصورت چگالی شرطی $f(x/\theta)$ نشان می‌دهیم. برای اینکه نقش θ را بیاد داشته باشیم در این فصل همواره به جای نماد $f(x;\theta)$ نماد شرطی $f(x/\theta)$ را بکار خواهیم برد.

تعریف ۱-۱-۱: فرض کنید پارامتر θ، مقدار مشاهده شده متغیر تصادفی Θ باشد. چگالی این متغیر را می‌توان به یاری اطلاعات گذشته یا از راه تجربه یا بر پایه سلیقه و عقیده شخصی تعیین نمود. این چگالی را با $\pi(\theta)$ نشان می‌دهیم و آنرا چگالی پیشین [1] می‌نامیم.

تعریف ۱-۱-۲: چگالی شرطی θ رابا معلوم بودن $\underset{\sim}{X}=\underset{\sim}{x}$ بصورت زیر:

$$h(\theta/\underset{\sim}{x})=\frac{f(\underset{\sim}{x}/\theta)\,\pi(\theta)}{\int f(\underset{\sim}{x}/\theta)\,\pi(\theta)\,d\theta}$$

1-prior density

نشان می‌دهیم و آنرا چگالی پسین [1] θ می‌نامیم.

برای محاسبه چگالی پسین در حالت پیوسته از قضیه کلی زیر استفاده می‌کنیم.

قضیه ۱-۱-۱: فرض کنید $X_1,\ldots\ldots,X_n$ یک نمونه تصادفی به حجم n از چگالی $f(x/\theta)$ باشد. اگر $\pi(\theta)$، $h(\theta/\underset{\sim}{x})$ و $L(\theta)$ به ترتیب چگالی پیشین، چگالی پسین و تابع درستنمایی باشد آنگاه داریم:

$$h(\theta/\underset{\sim}{x}) \propto \pi(\theta)L(\theta)$$

اگر ضریب تناسب راکه تنها تابعی از $\underset{\sim}{x}$ است با c نشان دهیم داریم:

$$h(\theta/\underset{\sim}{x}) = c\,\pi(\theta)L(\theta)$$

اثبات: چگالی توام X و θ را با $f(x,\theta)$ نشان می‌دهیم. در این چگالی $n+1$ متغیر θ و $x_1,\ldots\ldots,x_n$ داریم. می‌دانیم‌که $f(\underset{\sim}{x},\theta)=\pi(\theta)\cdot f(\underset{\sim}{x}/\theta)$ و چگالی شرطی $f(\underset{\sim}{x}/\theta)$ برای $\underset{\sim}{x}$ مشاهده شده تابعی از θ است که آن را با

$$f(\underset{\sim}{x}/\theta) = L(\theta)$$

نمایش می‌دهیم. بنابراین داریم:

$$h(\theta/\underset{\sim}{x}) = \frac{f(\underset{\sim}{x},\theta)}{f(\underset{\sim}{x})} = \frac{\pi(\theta)\,f(\underset{\sim}{x}/\theta)}{f(\underset{\sim}{x})} = \frac{\pi(\theta)L(\theta)}{f(\underset{\sim}{x})}$$

که در آن:

$$f(\underset{\sim}{x}) = \int f(\underset{\sim}{x}/\theta)\,\pi(\theta)\,d\theta$$

چگالی کناری به θ بستگی ندارد قضیه ثابت است.

اثبات قضیه فوق در حالتی که متغیر X یا θ گسسته باشد مشابه است.این قضیه نه تنها از نظر محاسبه چگالی پسین دارای اهمیت می‌باشد بلکه صریحاً نشان می‌دهد که چگالی پسین شامل اطلاعاتی در باره پارامتر θ است که در چگالی پیشین و تابع درستنمائی نهفته است.

1-posteriotr density

محاسبه چگالی پسین در بعضی موارد بسیار دشوار می‌باشد. اغلب ممکن است بخواهیم چگالی پسین را بعد از مدتی با توجه به داده‌های جدید، بعنوان چگالی پیشین بکار ببریم و یک چگالی پسین جدید پیدا کنیم . انجام اینکار برای خانواده چگالی پیشین مزدوج[1] آسان می‌باشد.

تعریف ۱-۱-۳: فرض کنید $f(x/\theta)$ چگالی شرطی متغیر تصادفی X ، متعلق به خانواده F^* و $\pi(\theta)$ چگالی پیشین θ متعلق به خانواده P^* باشد، P^* را یک خانواده مزدوج برای F^* می‌نامند هرگاه:

$$f(x/\theta)\in F^* \;,\; \pi(\theta)\in P^* \Longrightarrow h(\theta/\underset{\sim}{x})\in P^*$$

بنابراین P^* را یک خانواده مزدوج برای f می‌نامیم هرگاه چگالی پسین هم متعلق به خانواده چگالی پیشین می‌باشد. هر اندازه بتوانیم اطلاعات شخصی را بیشتر مورد استفاده قرار دهیم، این چگالی بیشتر در چگالی پسین تاثیر دارد یا به تعبیر دیگر آگاهی‌بخش می‌باشد چرا که عرض نقطه ماکزیمم منحنی اینگونه چگالی پیشین زیاد می‌باشد و از این رو می‌گویند چگالی پیشین آگاهی‌بخش[2] مرتفع می‌باشد.

اگر چگالی پیشین در همه نقاط به محور طولها بسیار نزدیک باشد در حقیقت وضعی بوجود می آید که حاکی از بی‌اطلاعی و عدم اظهار عقیده در باره پارامتر θ می‌باشد این نوع چگالی را چگالی آگاهی نابخش[3] و در نتیجه مسطح گوئیم.

مثال ۱-۱-۱: فرض کنید $X_1,\dots\dots,X_n$ نمونه تصادفی به حجم n از توزیع $N(\mu,\sigma^2)$ باشد که در آن σ^2 معلوم و μ مقدار مشاهده شده یک متغیر تصادفی است که دارای چگالی پیشین $N(\mu_0,\alpha^2)$ است که در آن μ_0 و α^2 مقادیر معلوم می‌باشد چگالی پسین μ بصورت زیر قابل محاسبه است.

$$X_i \sim N(\mu,\sigma^2)$$
$$\pi(\mu)=\frac{1}{\sqrt{2\pi\alpha^2}}\exp\left\{\frac{-1}{2\alpha^2}(\mu-\mu_0)^2\right\}$$

1–prior conjugate family　　2- infromative

3-non-infromative

$$L(\mu)=\left(\frac{1}{2\pi\sigma^2}\right)^{n/2} exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

با استفاده از قضیه (۱-۱-۱) و اتحاد زیر:

$$\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}(x_i-\bar{x})^2+n(\mu-\bar{x})^2$$

و حذف ضرایبی که به μ بستگی ندارند می‌توان نوشت:

$$h(\mu/\underset{\sim}{x}) \propto exp\left\{\frac{-1}{2\alpha^2}(\mu-\mu_0)^2-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$\propto exp\left\{\frac{-1}{2\alpha^2}(\mu^2-2\mu\mu_0+\mu_0^2)-\frac{1}{2\sigma^2}n(\mu-\bar{x})^2\right\}$$

$$\propto exp\left\{\frac{-1}{2\alpha^2}(\mu^2-2\mu\mu_0+\mu_0^2-\frac{n}{2\sigma^2}(\bar{x}^2-2\mu\bar{x}+\mu^2)\right\}$$

$$\propto exp\left\{\frac{-\mu^2}{2}\left(\frac{1}{\alpha^2}+\frac{n}{\sigma^2}\right)+\mu\left(\frac{\mu_0}{\alpha^2}+\frac{n\bar{x}}{\sigma^2}\right)\right\}$$

با تعریف μ_1 و σ_1^2 بوسیله روابط زیر:

$$\frac{1}{\sigma_1^2}=\frac{1}{\alpha^2}+\frac{n}{\sigma^2}$$

$$\frac{\mu_1}{\sigma_1^2}=\frac{\mu_0}{\alpha^2}+\frac{n\bar{x}}{\sigma^2}$$

داریم:

$$h(\mu/\underset{\sim}{x}) \propto exp\left\{\frac{-\mu^2}{2\sigma_1^2}+\frac{\mu\mu_1}{\sigma_1^2}\right\}$$

$$\propto exp\left\{\frac{-1}{2\sigma_1^2}(\mu-\mu_1)^2\right\}$$

بنابراین چگالی پسین μ عبارت از:

$$h(\mu/\underset{\sim}{x})=\frac{1}{\sqrt{2\pi\sigma_1^2}}\ exp\left\{\frac{-1}{2\sigma_1^2}(\mu-\mu_1)^2\right\} \sim N(\mu_1,\sigma_1^2)$$