دانشکده علوم ریاضی

پایان‌نامه

جهت اخذ درجه کارشناسی ارشد آمار ریاضی

عنوان

استنباط آماری بر اساس داده‌های سانسور شده پیش رونده نوع دو تطبیقی در توزیع نمایی

استاد راهنما:

دکتر آرزو حبیبی راد

استاد مشاور:

دکتر مهدی دوست پرست

نگارنده:

هادی سنجی

خرداد ۱۳۹۰

فهرست مندرجات

چکیده

در این پایان‌نامه آمیخته‌ای از طرحهای سانسور نوع یک و سانسور پیش رونده نوع دو بنام طرح سانسور پیش رونده نوع دو تطبیقی که درآزمونهای طول عمر و قابلیت اعتماد مرسوم می‌باشد، معرفی و به بررسی استنباطی و اثبات برخی ویژگیها در این سانسور می‌پردازیم. در این سانسور، اندازه نمونه موثر (m) از قبل ثابت در نظر گرفته می‌شود و تعداد واحدهای خارج شده از آزمایش (طرح سانسور) در طول آزمایش به صورت زیر تغییر می‌کند.

۱ – اگر m امین شکست قبل از زمان از پیش تعیین شده T رخ دهد، آزمایش را در زمان m امین شکست پایان می‌دهیم و طرح سانسور، مشابه طرح سانسور پیش رونده نوع دو است.

۲ –اگر m امین شکست بعد از زمان T رخ دهد و تنها J شکست تا قبل از زمان T رخ داده باشد، آنگاه عمل حذف مولفه‌ها را برای شکست‌های بعد از زمان T تا لحظه رسیدن به m امین شکست متوقف کرده و در زمان m امین شکست، تمام واحدهای باقیمانده ($R_m = n - m - \sum_{i=1}^{j} R_i$) را از آزمایش خارج می‌کنیم.

از روش برآوردگر ماکسیمم درستنمایی برای برآورد پارامتر مجهول تحت توزیع نمایی استفاده کرده و سپس فواصل اطمینان مختلف را محاسبه کرده و با استفاده از شبیه سازی، نتایج حاصل را با یکدیگر مقایسه می کنیم. در فصل ۴ برنامه‌های رایانه‌ای نوشته شده توسط نرم افزار R قرار داده شده است.

واژه‌های کلیدی: آزمون طول عمر، برآوردگر ماکسیمم درستنمایی، سانسور پیش رونده نوع دو تطبیقی، روشهای مجانبی، توزیع نمایی، احتمال پوشش.

پیشگفتار

بسیاری از اوقات در آزمون‌های طول عمر، آزمایش‌های کلینیکی، مطالعات تأثیر دوز (Dose) سم‌ها، تحقیقات زیست شناسی و دیگر زمینه‌های علم آمار، این امکان وجود دارد که زمان شکست کامل بعضی از واحدها مشاهده نشود، به این معنی که در برخی وضعیت‌ها، برای برخی از واحدها در طول آزمایش شکست اتفاق نمی‌افتد و یا از ادامه آزمایش باز نمی‌مانند و به جای دانستن زمان شکست، تمام آنچه می‌دانیم این است که این واحدها طول عمری متجاوز از مقداری مانند y دارند. این محدودیت‌های پیش آمده در نمونه‌ها را سانسور گوییم. در مطالعات پزشکی و یا تحلیل های قابلیت اعتماد، بسیار معمول است که شکست هر واحد به بیش از یک علت مربوط باشد. به علاوه، داده های مشاهده شده اغلب سانسور شده اند.
این پایان نامه مشتمل بر چهار فصل می‌باشد.

- در فصل اول، به معرفی مفاهیم و مقدماتی از جمله روش ماکسیمم درستنمایی، داده‌های سانسور شده، آماربیز، روشهای بوت استرپ، روش دلتا و مفاهیم دیگری که در فصلهای بعد مورد استفاده قرار می‌گیرد می‌پردازیم.

- در فصل دوم، به معرفی انواع سانسورها، پرداخته و به امتیازات و معایب برخی از آنها نیز اشاره کرده‌ایم و بعد از آن سانسور پیش رونده نوع دو تطبیقی را که هدف اصلی این پایان‌نامه است را مورد بررسی قرار می‌دهیم.

- در فصل سوم، مجدداً سانسور پیش رونده نوع دو تطبیقی را معرفی و سپس برآوردگر ماکسیمم درستنمایی پارامتر مجهول و فواصل اطمینان مختلف را بر اساس این سانسور در توزیع نمایی به دست می‌آوریم و با استفاده از شبیه سازی به مقایسه روشهای مختلف ساخت فاصله اطمینان می‌پردازیم.

- در فصل چهارم برنامه‌های رایانه‌ای مربوط به فصل سوم که در محیط نرم افزار R نوشته شده را قرار می‌دهیم.

در پایان لازم می‌دانم تشکر خود را نسبت به استاد محترم، سرکار خانم دکتر حبیبی راد که زحمت راهنمایی این پایان نامه را بر عهده داشتند و همچنین جناب آقای دکتر دوست پرست که سمت مشاوره این پایان نامه را عهده دار بودند، اعلام نمایم. از سرکار خانم دکتر یوسف زاده که در امر برنامه نویسی در محیط نرم افزار R اینجانب را یاری و راهنمایی کردند، کمال تشکر و قدردانی را دارم. همچنین شایسته است مراتب قدردانی خود را از پدر و مادرم که همواره مشوق و همراه من بودند را اعلام نمایم. در انتها از تمام اعضای هیات علمی گروه آمار، مدیریت دانشکده، مسئولین و کارکنان آزمایشگاه رایانه، کتابخانه، آموزش، دوستان و همه کسانی که طی این دوره مرا یاری رسانده‌اند، تشکر نمایم.

هادی سنجی

خرداد ۱۳۹۰

فصل ۱

مفاهیم و مقدمات

۱.۱ مقدمه

در این فصل به تعریف مفاهیم و اصطلاحات مورد نیاز در فصل‌های بعدی می‌پردازیم. مفاهیمی نظیر داده‌های سانسور شده، روش ماکسیمم درستنمایی، فاصله اطمینان، آمار بیز و مفاهیمی دیگر که در فصل‌های بعدی مورد استفاده قرار خواهند گرفت.

۲.۱ مقدمه‌ای در مفاهیم بقا

تحلیل بقا[۱] مجموعه‌ای شامل روش‌های آماری متنوع برای تجزیه و تحلیل متغیرهای تصادفی است که دارای مقادیر مثبت می‌باشند. مقدار این متغیر تصادفی، زمان شکست یک مولفه فیزیکی (مکانیکی یا الکترونیکی) یا زمان مرگ یک واحد زنده (سلول، بیمار، حیوان و غیره) است، که گاهی متغیر تصادفی مورد نظر مربوط به زمان نمی‌باشد. برای مثال، متغیر مربوط به هزینه پرداختی یک شرکت بیمه به بیمه شدگان در یک وضعیت خاص باشد. به طور مثال در برخی موارد، یک بیمار بهبود یافته و مبلغ کل پرداختی بیمه او معلوم است. در مواردی دیگر، بیماری هنوز ادامه دارد و تنها مبلغ پرداختی تا آن زمان، معلوم است.

برای تحلیل بقا نامهای گوناگونی در نظر می‌گیرند، مثل تحلیل داده های طول عمر، تحلیل قابلیت اعتماد، تحلیل زمان انتظار پیشامدها و سرانجام تحلیل سیر پیشامدها که بسته به نوع کاربرد تحلیل بقا، ممکن است از هر کدام از آن‌ها استفاده شود.

برای تعیین زمان بقا، دو نقطه زمانی باید معین شود. زمان مبدا یعنی زمانی که در آن پیشامد آغازین مانند تولد رخ می‌دهد و نیز زمان به اصطلاح شکست، یعنی زمانی که در آن پیشامدی نهایی مثل مرگ رخ می‌دهد.

مشخصه بارز داده‌های بقا که آن را از سایر شاخه‌های آمار متمایز می‌کند این است که این داده ها به طور معمول سانسور می‌شوند. سانسور شدن زمانی اتفاق می‌افتد که اطلاعات ناقصی درباره زمان بقای عناصر در دست باشد.

[۱] Survival analysis

۳.۱ داده‌های سانسور شده

در بسیاری از آزمون‌های طول عمر، تحقیقات زیست شناسی و دیگر زمینه‌های علم آمار با نمونه‌هایی سروکار داریم که همه مشاهدات نمونه ثبت نشده و یا همه واحدهای نمونه به نتیجه نرسیده‌اند. به عنوان مثال در یک نمونه به حجم n ممکن است تنها اطلاعات مربوط به n_1 واحد نمونه به طور کامل در دسترس باشد و لذا برای $n - n_1$ واحد نمونه اطلاعات کاملی وجود ندارد. این محدودیت ممکن است به صورت اختیاری و یا از قبل توسط آزمایشگر اعمال شود. به عنوان مثال، اگر یکی از واحدهای آزمایشی به طور تصادفی خراب شود یا شخص مورد مطالعه از ادامه همکاری کناره‌گیری کند در این صورت حذف از نوع غیر عمدی رخ داده است، ولی اگر به علت برخی شرایط ویژه مانند تمام شدن بودجه، طولانی شدن مدت آزمایش، فرصت کم برای اعلام نتایج، آزمایشگر آزمایش را قبل از حصول زمان خرابی تمام واحدها متوقف کند، حذف از نوع عمدی اتفاق افتاده است.

به داده هایی که قبل از مشاهده زمان شکست از آزمایش خارج می‌شوند، داده سانسور شده می‌گویند. بدیهی است با سانسور کردن مشاهدات، مقداری از اطلاعات موجود در نمونه را از دست خواهیم داد. با این وجود گاهی اوقات مایل یا مجبور به سانسور کردن هستیم.

سانسورها به صورت های مختلفی اعمال می شوند. بعضی از انواع سانسورها عبارتند از: سانسور نوع اول، سانسور نوع دوم، سانسور تصادفی، سانسور دوطرفه، سانسور هیبرید و دیگر سانسورها که برای جزییات بیشتر درباره انواع سانسورها و کاربرد آنها می توان به کتاب های آرنولد[۲] و همکاران (۱۹۹۲)، لاولس[۳] (۲۰۰۳ ، ۱۹۸۲) و نلسن[۴](۲۰۰۴) مراجعه کرد.

یکی از اولین تلاشها در جهت تحلیل مسائل آماری که شامل داده‌های سانسور شده بود توسط دانیل برنولی[۵] (۱۷۶۶) انجام گرفت که در تحلیل شیوع بیماری آبله و اطلاعات مربوط به مرگ و میر در اثر این بیماری بود.

---

[۲] Arnold

[۳] Lawless

[۴] Nelson

[۵] Daniel Bernoulli

۴.۱ روش ماکسیمم درستنمایی

روش ماکسیمم درستنمایی[۶] (ML) یکی از قدیمی‌ترین و پر اهمیت‌ترین روش‌ها در نظریه برآوردهاست. برآورد ML که متداول‌ترین روش در بین استفاده کنندگان کاربردی از علم آمار است، اولین بار توسط گوس[۷](۱۸۲۱) به کار گرفته شد و پس از آن به صورت گسترده‌ای توسط فیشر[۸] (۱۹۲۵) در انتقاد از روش "برآورد گشتاوری" مورد استفاده قرار گرفت.
روش MLE عبارت است از دستورالعملی برای به دست آوردن برآوردگری به نام "برآوردگر ماکسیمم درستنمایی" و مبتنی بر تابع آماری به نام تابع درستنمایی است.

۱.۴.۱ تابع درستنمایی

نمونه تصادفی $X_1, X_2, ..., X_n$ را از متغیر تصادفی X با تابع چگالی احتمال $f(x;\theta)$ در نظر می‌گیریم. فرض می‌کنیم مقدار واقعی پارامترθ به مجموعه Θ، به نام فضای پارامتر، تعلق داشته باشد. یافته‌های این نمونه تصادفی را با $x_1, x_2, ..., x_n$ نشان می‌دهیم. می دانیم X_iها $(i = 1, 2, ..., n)$ مستقل و همتوزیع با متغیر تصادفی X هستند. بنابراین تابع چگالی احتمال هر X_i برابر است با $f(x_i;\theta)$ و تابع چگالی احتمال توام X_iها با حاصلضرب توابع چگالی احتمال آن‌ها برابر است. این تابع چگالی احتمال توام را با فرض این که x_i ها ثابت و θ متغیر باشد با $L(\theta)$ نشان می‌دهیم و آن را تابع درستنمایی[۹] برای یافته‌های نمونه تصادفی می‌نامند. به طور

[۶] Maximum likelihood
[۷] Gauss
[۸] Fisher
[۹] Likelihood function

خلاصه داریم

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta).$$

در حالت گسسته $L(\theta)$ با احتمال مشاهده یافته‌های بالا بر حسب پارامتر مجهول θ برابر است. چه در حالت گسسته و چه در حالت پیوسته، هر اندازه $L(\theta)$ بیشتر باشد شانس مشاهده $x = (x_1, x_2, ..., x_n)$ بیشتر است. از این رو مقداری از $\theta \in \Theta$ را به عنوان برآورد θ بر می‌گزینیم که $L(\theta)$ را ماکسیمم کند.

لازم به ذکر است که

الف) تابع درستنمایی $L(\theta)$ لزوماً نسبت به θ مشتق پذیر نیست.

ب) تابع درستنمایی $L(\theta)$ لزوماً یک تابع چگالی احتمال نیست.

۲.۴.۱ برآورد ماکسیمم درستنمایی

مقداری از θ که $L(\theta)$ را ماکسیمم کند برآوردگر ماکسیمم درستنمایی (MLE) نامیده و آن را با $\hat{\theta}$ نشان می‌دهیم. به عبارت دیگر

$$L(\hat{\theta}) \geq L(\theta), \quad \forall \theta \in A.$$

در اغلب موارد به خصوص برای مشتق گیری آسانتر است که به جای $L(\theta)$ از $l(\theta) = \ln L(\theta)$ استفاده کنیم.

۵.۱ فاصله اطمینان

فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی از تابع چگالی $f(.;\theta)$ باشد. اگر $L = l(X_1, X_2, ..., X_n)$ و $U = u(X_1, X_2, ..., X_n)$ دو آماره باشند به طوریکه $L \leq U$ و

$P_\theta(L < \theta < U) = ۱ - \alpha$، در این صورت بازه تصادفی (L, U) را یک خانواده از فواصل اطمینان $۱۰۰(۱-\alpha)$درصدی برای θ در اندازه $(۱-\alpha)$ می‌گویند. یک مقدار تصادفی (l, u) از بازه تصادفی (L, U) را نیز یک بازه اطمینان $۱۰۰(۱-\alpha)$درصدی برای θ می‌نامند. توجه داشته باشید L و U به ترتیب حدود اطمینان پایینی و بالایی برای θ و $(۱-\alpha)$ ضریب اطمینان نامیده می‌شود.

تعریف ۱.۵.۱ فاصله اطمینان مجانبی: بازه (L_n, U_n) را فاصله اطمینان مجانبی گویند هر گاه گزاره

$$P(L_n < \theta < U_n) = ۱ - \alpha,$$

به صورت حدی وقتی حجم نمونه (n) به سمت بینهایت میل کند، درست باشد.

۶.۱ آزمون نسبت درستنمایی

آزمون نسبت درستنمایی، [۱۰] که به اختصار با LRT نمایش می‌دهیم یکی از قدیمی‌ترین و متداول‌ترین روشها در نظریه آزمون فرضیه‌هاست. روش LRT در سال ۱۹۲۸ توسط نیمن و پیرسون [۱۱] مطرح شد. نقشی که این روش در نظریه آزمون فرض‌ها داراست، همانند روش ماکسیمم درستنمایی در نظریه برآوردهاست. روش LRT با وجود این که به طور شهودی قابل توجه است، همواره بهترین روش نیست و نباید همیشه مورد استفاده قرار گیرد. در حقیقت قضیه‌ای که بیانگر این باشد که روش LRT جز در موارد مجانبی دارای خواص خوبی باشد، وجود ندارد.

[۱۰] Likelihood Ratio Test

[۱۱] Neyman and Pearson

۱.۶.۱ آزمون نسبت درستنمایی برای فرضیه‌های ساده

فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی n تایی از توزیعی با تابع چگالی احتمال $f_\theta(x)$ باشد و بخواهیم آزمون $H_0 : \theta = \theta_0$ در مقابل $H_1 : \theta = \theta_1$ انجام دهیم. آماره آزمون برابر است با

$$\Lambda = \frac{f_{\theta_0}(\underline{x})}{f_{\theta_1}(\underline{x})},$$

به طوریکه در این روش سعی در ماکسیمم کردن $L(\theta) = f_\theta(\underline{x})$ است. بنابراین اگر $f_{\theta_0}(\underline{x}) < f_{\theta_1}(\underline{x})$ آنگاه ماکسیمم تابع درستنمایی در θ_1 اتفاق می افتد، بنابراین باید فرضیه H_0 را رد کرد. به بیان دیگر اگر $\Lambda \leq k$ باشد فرضیه H_0 را نمی‌پذیریم و اگر $\Lambda > k$ آنگاه فرضیه H_1 را نمی‌پذیریم. مقدار بحرانی k غالباً برای یک سطح معنی داری α انتخاب می‌شود، به طوری که

$$P_{\theta_0}(\Lambda > k) = 1 - \alpha.$$

۲.۶.۱ آزمون نسبت درستنمایی برای فرضیه‌های مرکب

فرض کنید $X_1, X_2, ..., X_n$ یک نمونه تصادفی n تایی از توزیعی با تابع چگالی احتمال $f_\theta(x)$ باشد که در آن $\theta \in \Theta$. همچنین فرض کنید Θ_0 و Θ_1 زیر مجموعه هایی از Θ باشند به طوری که

$$\Theta_0 \cup \Theta_1 = \Theta \quad , \Theta_0 \cap \Theta_1 = \emptyset$$

علاقه‌مند به آزمون $H_0 : \theta \in \Theta_0$ در مقابل $H_1 : \theta \in \Theta_1$ هستیم. می دانیم اصل ماکسیمم درستنمایی توجه خود را به مقداری از θ، یعنی $\hat{\theta}$، که تابع درستنمایی $L(\theta)$ را ماکسیمم کند، معطوف می‌کند. یعنی

$$L(\hat{\theta}) = \sup_{\theta} L(\theta).$$

در حقیقت چنین مقداری از $\hat{\theta}$، مشاهدات $\underline{x} = (x_1, x_2, ..., x_n)$ را بهتر بیان می‌کند. بنابراین در انتخاب H_0 و H_1، طبیعی به نظر می رسد که مقایسه بین بهترین بیان کننده موجود در H_0 و بهترین بیان کننده در H_1 انجام پذیرد. پس وقتی $\boldsymbol{X} = \boldsymbol{x}$ مشاهده شود، فرضیه H_0 را رد می‌کنیم، اگر و فقط اگر

$$\sup_{\Theta_0} L(\theta) < \sup_{\Theta_1} L(\theta),$$

تعمیم ایده فوق برای دستیابی به انعطاف بیشتر در مورد سطح آزمون، ما را بر آن می‌دارد که ناحیه رد آزمون را به شکل

$$\lambda^\star(\underline{x}) = \frac{\sup_{\Theta_0} L(\theta)}{\sup_{\Theta_1} L(\theta)} < \lambda^\star_0,$$

داشته باشیم، که در آن $\lambda^\star_0$ عددی ثابت است. در بسیاری از موارد راحت تر است که به جای مقایسه ماکسیمم درستنمایی تحت Θ_0 و Θ_1 این مقایسه را تحت Θ_0 و کل فضای پارامتر یعنی Θ انجام دهیم که این منجر به تشکیل ناحیه رد زیر می‌شود که تعمیمی از روش ناحیه رد روش درستنمایی است.

$$\lambda(\underline{x}) = \frac{\sup_{\Theta_0} L(\theta)}{\sup_{\Theta} L(\theta)} < \lambda_0, \qquad \lambda_0 < 1.$$

بنابراین با توجه به مطالب بیان شده می‌توان برای انجام آزمون فرضیه H_0 در مقابل H_1 از ناحیه رد $\lambda(\underline{x}) < \lambda_0 < 1$ استفاده کرد. بنابراین تابع آزمون

$$\phi(\underline{x}) = \begin{cases} 1 & \lambda(\underline{x}) < \lambda_0 \\ \gamma & \lambda(\underline{x}) = \lambda_0 \\ 0 & \lambda(\underline{x}) > \lambda_0 \end{cases}$$

آزمون نسبت درستنمایی است، که در آن λ_0 و γ به گونه‌ای تعیین می‌شوند که اندازه آزمون α شود، یعنی

$$\alpha = \sup_{\theta \in \Theta_0} E_\theta[\phi(X)].$$

نکته ۱.۶.۱ : در آزمون نسبت درستنمایی مواردی وجود دارند که در آنها قادر به تعیین توزیع آماره $\lambda(.)$ یا آماره معادل آن نیستیم. در چنین مواردی باید از توزیع مجانبی استفاده کنیم. وقتی اندازه نمونه بزرگ باشد آماره $-2ln\lambda(\underline{X})$ دارای توزیع تقریبی χ^2 با r درجه آزادی است، که در آن r برابر اختلاف تعداد پارامترها تحت فرضیه H_0 و H_1 می‌باشد.

۷.۱ آمار بیز

پارامتر θ را می توان اغلب به صورت یک متغیر تصادفی از فضای پارامتر Θ در نظر گرفت. به عنوان مثال فرض کنید قطعات تولیدی یک کارخانه دارای توزیع نمایی با میانگین θ باشد. طبیعی است به علت فرسودگی دستگاهها میانگین θ در سال تغییر کند و آن را نمی توان یک مقدار همیشه ثابت در نظر گرفت. آماری که با این دید پایه گذاری شد به آمار بیز [12] شهرت یافت که امروزه مورد توجه آماردانان قرار گرفته و از رشته‌های مهم علم آمار می‌باشد.

۱.۷.۱ تابع چگالی پیشین

فرض کنید پارامتر θ به عنوان مقدار یک متغیر تصادفی W در نظر گرفته می‌شود که مقادیر ممکن آن از فضای پارامترΘ است و دارای تابع توزیع $G(\theta)$ یا تابع چگالی احتمال $g(\theta)$ است که از آن به عنوان توزیع پیشین یا تابع احتمال پیشین یاد می‌کنیم. در واقع، قبل از مشاهده و جمع آوری هر گونه داده‌ای، اطلاعات و دانسته‌های قبلی آمارگر، وی را متقاعد می‌کند بر این باور باشد که شانس قرار گرفتن مقادیر θ در فضای Θ چگونه است و ما فرض می‌کنیم که چنین باورهایی را می‌توان در قالب یک تابع بیان کرد. البته لزومی ندارد که چگالی پیشین خواص تابع توزیع را داشته باشد فقط کافیست مثبت و انتگرال آن روی فضای پارامتر متناهی باشد.

[12] Bayes Statistics

۲.۷.۱ تابع چگالی پسین

به توزیع شرطی W به شرط $X = x$ توزیع پسین W می‌گوییم. توزیع پسین W دارای چگالی پسینی [۱۳] به صورت

$$\pi_{W|X=x}(\theta) = \frac{f_{X,W}(x;\theta)}{f_x(X)},$$

است، که در آن $f_{X,W}(.,.)$ تابع چگالی احتمال توام X و W و $f_X(x)$ تابع چگالی احتمال حاشیه‌ای X است. حال برای محاسبه توزیع پسین با فرض آن که W گسسته یا پیوسته باشد داریم

$$\pi(\theta|x) = \begin{cases} \frac{g(\theta)f(\theta|x)}{\sum_{\Theta} g(\theta)f(\theta|x)}, & \text{در حالت گسسته} \\ \frac{g(\theta)f(\theta|x)}{\int_{\Theta} g(\theta)f(\theta|x)\, d\theta}. & \text{در حالت پیوسته} \end{cases} \qquad (۱)$$

۳.۷.۱ فاصله اطمینان بیزی

اگر $f(.|\theta)$ تابع چگالی یک نمونه تصادفی $\underline{X} = (X_۱, X_۲, ..., X_n)$ و $\pi(\theta)$ چگالی پیشین θ باشند، آنگاه چگالی پسین θ با فرض معلوم بودن $\underline{X} = \underline{x}$ عبارتست از

$$\pi(\theta|\underline{x}) = \frac{f(\underline{x}|\theta)\pi(\theta)}{\int_{\Theta} f(\underline{x}|\theta)\pi(\theta)d\theta}.$$

بر این اساس $(l(\underline{x}), u(\underline{x}))$ را یک فاصله اطمینان بیزی در اندازه $(۱ - \alpha)$ برای θ گوییم هر گاه

$$\int_{l(\underline{x})}^{u(\underline{x})} \pi(\theta|\underline{x})\, d\theta = ۱ - \alpha.$$

[۱۳] Posterior density

در عمل می‌توان مقادیری از $l(\underline{x})$ و $u(\underline{x})$ را که در رابطه بالا صدق می‌کنند پیدا کرد که برای آن $u(\underline{x}) - l(\underline{x})$ کمترین باشد.
به عنوان مثال، در یک آزمایش هنگامیکه گفته می‌شود یک فاصله باور ۹۵ درصد برای پارامتر θ، (۳۰،۴۰) می‌باشد، به این معناست که احتمال پسین این که θ در بازه $(30, 40)$ قرار گیرد ۹۵ درصد می‌باشد.

۸.۱ شرایط نظم

فرض کنید X یک متغیر تصادفی حقیقی مقدار باشد که مطلقاً پیوسته و دارای تابع توزیع تجمعی $F_\theta \in \Omega_\theta = F_\theta : \theta \in \Theta$ باشد. در این پایان‌نامه خانواده توزیع‌هایی را مورد بررسی قرار می‌دهیم که دارای شرایط نظم به شرح زیر باشند:
(۱) Θ یک زیر بازه ناتباهیده از اعداد حقیقی بوده و حوزه تغییرات X به θ بستگی نداشته باشد.
(۲) به ازای هر $\theta \in \Theta$، $f(x;\theta)$، نسبت به m اندازه پذیر باشد و به ازای هر $x \in X$، $f(x;\theta)$ نسبت به θ مشتق پذیر باشد.
(۳) اگر $A \subset \chi$ یک مجموعه اندازه پذیر نسبت به m باشد، آنگاه :

$$\frac{\partial}{\partial\theta}\int_A f(x;\theta)dx = \int_A \frac{\partial}{\partial\theta} f(x;\theta)dx.$$

(۴) مشتق دوم $F_\theta \in \Omega$ نسبت به x وجود داشته باشد و رابطه زیر برقرار باشد:

$$\frac{\partial^2}{\partial\theta\partial x}F(x;\theta) = \frac{\partial^2}{\partial x\partial\theta}F(x;\theta) = \frac{\partial}{\partial\theta}f(x;\theta).$$

(۵) برای هر برآوردگر T پارامتر θ، رابطه زیر برقرار باشد:

$$\frac{\partial}{\partial\theta}\int t(x)f(x;\theta)dx = \int \frac{\partial}{\partial\theta}t(x)f(x;\theta)dx.$$