دانشگاه پیام نور

دانشکده علوم پایه

گروه آمار

عنوان پایان نامه :

محاسبه اندازه نمونه به روش بیز

پایان نامه : برای دریافت درجه کارشناسی ارشد

در رشته آمار (آمار ریاضی)

استاد راهنما:

دکتر علی شادرخ

استاد مشاور:

دکتر مسعود یار محمدی

مولف:

قاسم گنجی

شهریور ۹۱

جمهوری اسلامی ایران
وزارت علوم، تحقیقات و فناوری

مرکز تهران شرق

دانشگاه پیام نور

دانشگاه پیام نور استان تهران

والعلم کل نواحیه الفرح والعافیة والنصر

صورتجلسه دفاع از پایان نامه کارشناسی ارشد

جلسه دفاع از پایان نامه کارشناسی ارشد آقای قاسم گنجی نودهی

دانشجوی رشته آمار ریاضی به شماره دانشجویی: ۸۸۰۲۷۲۵۶۳

<u>تحت عنوان:</u>

محاسبه اندازه نمونه به روش بیز

جلسه دفاع با حضور داوران نامبرده ذیل در روز دوشنبه مورخ: ۹۱/۰۶/۲۷ ساعت: ۱۰-۹ درمحل تهران شرق برگزار شد، و پس از بررسی پایان‌نامه مذکور با نمره به عدد ۱۸/۵ به حروف ......هجده و نیم...... و با درجه ارزشیابی......عالی......مورد قبول واقع شد ☑ نشد ☐

ردیف	نام و نام خانوادگی	هیات داوران	مرتبه دانشگاهی	دانشگاه/ موسسه	امضاء
۱	دکتر علی شادرخ	استاد راهنما	استادیار	پیام نور	
۲	دکتر مسعود یارمحمدی	استاد مشاور	دانشیار	پیام نور	
۳	دکتر عادل محمد پور	استاد داور	دانشیار	دانشگاه صنعتی امیر کبیر	
۴	دکتر خدیجه احمدی آملی	نماینده علمی گروه/ نماینده تحصیلات تکمیلی	استادیار	پیام نور	

... خیابان کریمخان
... خیابان استاد نجات
... خیابان شهید فلاح
... پلاک ۲۷ مرکز
... ن شرق
۸۸۹۱۳۴۷۵
...نگار: ۸۸۹۴۸۹۸۴
Tshargh.Tpnu.ac
Tshargh@Tpnu.ac

فهرست مطالب

عنوان **صفحه**

فصل سوم :اندازه‌ی نمونه با استفاده از تابع مطلوبیت

فصل چهارم:روش‌های ارائه شده برای توزیع نرمال

چکیده

در این پایان‌نامه به مرور برخی از روش‌های اصلی محاسبه اندازه نمونه از دیدگاه کلاسیک و دیدگاه بیزی می‌پردازیم. ممکن است در ذهن خواننده ایجاد شود که چرا به سراغ روش بیز می‌رویم و از روش کلاسیک استفاده نمی‌کنیم. با فرض اینکه نمونه ای تصادفی دارای توزیع نرمال باشد به این سؤال پاسخ خواهیم داد. از طرفی یک نمونه با اندازه‌ی بهینه هزینه‌ی زیادی ندارد و دقت کافی را فراهم می‌کند. چون توزیع نمونه به چند پارامتر بستگی دارد، محاسبه‌ی نمونه با اندازه بهینه در روش کلاسیک همیشه آسان نیست. تلاش شده است که هیچ یک از روش‌های اصلی و کلیدی از قلم نیفتد. دو زمینه‌ی اصلی این دیدگاه عبارتند از زمینه بیز استنباطی و زمینه کاملاً بیزی (نظریه تصمیم). در زمینه بیز استنباطی، غالباً درگیر استنباط راجع به پارامتر مجهول و مورد علاقه‌مان در جامعه هستیم و اندازه‌ی نمونه را با استفاده از پارامترهای چگالی پسین بدست می‌آوریم.

حال آنکه در زمینه‌ی کاملاً بیزی (نظریه تصمیم) با بهینه کردن تابع هدف اندازه نمونه را بدست می‌آوریم.

واژه‌های کلیدی: اندازه‌ی نمونه، روش بیز استنباطی، روش کاملاً بیزی (نظریه تصمیم)، روش درستنمایی، عامل بیز، روش مبتنی بر فاصله گشتاورها.

پیشگفتار:

آمار بیز که در چند دهه اخیر مورد بحث و بررسی بوده است، امروز جای خود را در میان آمار دانان باز کرده است و از رشته های مهم آمار می‌باشد. اساس این نوع روش، تعبیر احتمال به طریق شخصی می‌باشد نه به طریق فراوانی نسبی. آمار دان بیز می‌گوید پارامتر (θ) را می‌توان اغلب به چشم یافته متغیر تصادفی (θ) نگاه کرد در حالی که آمار دان بیز چنین اعتقادی را ندارد. مثلاً فرض کنید θنسبت لامپهای ناسالم به تعداد تمام لامپهایی باشد که یک کارخانه تولید می‌کند. طبیعی است که در اثر فرسوده شدن ماشین‌های کارخانه ▫ به تدریج افزایش می‌یابد و نمی‌توان همیشه آن را یک مقدار ثابت تلقی کرد. در واقع آمار دان بیزی به روش دنباله ای علاقه‌مند است که از یک مرحله به بعد توزیع پسین را مانند توزیع پیشین مورد استفاده قرار دهد تا به نتیجه مورد نظرش برسد. در این رساله در فصل اول نگاهی اجمالی به تعیین اندازه نمونه به روش کلاسیک خواهیم داشت. در فصل دوم و سوم روش‌های تعیین اندازه نمونه بیزی را بررسی می‌کنیم و در فصل چهارم روش‌های ارائه شده را برای توزیع نرمال مورد بحث قرار می‌دهیم و در پایان شبیه سازی و نتیجه گیری را خواهیم داشت.

فصل اول

مقدمه

۱-۱- مقدمه

در طرح‌های آماری از آنجایی که هزینه‌ی طرح بیشتر مورد توجه است باید بین هزینه و دقت تعادل ایجاد شود بنابراین در مسائل مختلف آماری که نیاز به نمونه‌گیری مشاهده می‌شود، معمولاً این تعادل چنین برقرار می‌شود «انتخاب کوچک‌ترین اندازه‌ی نمونه جهت رسیدن به دقت مورد نظر». از آنجایی که در عبارت بالا اشاره‌ای به هزینه‌ی طرح نشده است ولی در این تحقیق حداقل مقدار ممکن اندازه نمونه(n)، جهت رسیدن به دقت مورد نظر و در ضمن آن، حداقل هزینه در نظر گرفته شده است.

در تعریف احتمال یک مناقشه‌ی قدیمی میان آ ما رد انان وجود دارد. گروهی از آ ما رد انان که به گروه کلاسیک[1] مشهور هستند، احتمال را به صورت «فراوانی مورد انتظار از پیشامد مورد نظر در آزمایش‌های تکرارپذیر در درازمدت و در شرایط یکسان» تعریف می‌کنند، از این رو احتمال پیشامد A برابر با $P(A)=\frac{n}{N}$ تعریف می‌شود که در آن n تعداد دفعات رخداد پیشامد A در N آزمایش با شرایط یکسان است. این در حالی است که از دیدگاه گروه دوّم یعنی آ ما ردان های بیزی، احتمال وابسته به درجه‌ی اعتقاد[2] آمار دان است.

در استنباط آماری هر یک از این دو گروه دیدگاه خاص خود را دارند. استنباط آماری از دیدگاه کلاسیک، استنباط آماری کلاسیک نامیده می‌شود و استنباط آماری از دیدگاه آ ما رد ان‌های بیزی، استنباط آماری بیزی نامیده می‌شود.

---

[1] Classic

[2] Degree

در استنباط آماری به شیوه کلاسیک، فرض می‌شود داده‌ها از یک توزیع معلوم ولی با پارامترهای نامعلوم بدست آمده‌اند، این نوع استنباط وابسته به فرض تصادفی بودن داده‌هاست و در استنباط آماری به روش بیزی از دانسته‌های پیشین استفاده می‌شود، به این معنی که یک توزیع پیشین برای پارامترها نامعلوم در نظر گرفته می‌شود. از توزیع پیش بینی کننده[1](چگالی کناری) برای بروز کردن دانسته‌های پیشین و بدست آوردن دانش پسین استفاده می‌شود و دانش پسین برای توزیع پیش بینی کننده[2](چگالی کناری)بعدی یک دانسته‌ی پیشین به شمار می‌آید. [۱۷]

برخلاف روش کلاسیک که مسئله تعیین اندازه نمونه را در آزمایش‌های مختلف از جمله آزمایش‌های کلینیکی متوالیاً مورد توجه قرار می‌دهد در روش بیزی این مسئله کمتر مورد توجه و بحث و بررسی قرار می‌گیرد. در سال ۱۹۵۲ فام- گیا[3][۵] دلایل مختلفی را برای این موضوع برشمرد از جمله این که یک آمار دان بیزی بیشتر به روش دنباله‌ای علاقمند است که از یک مرحله به بعد توزیع پسین راهمانند توزیع پیشین مورد استفاده قرارمی دهد تا به نتیجه‌ی مورد نظرش برسد بنابراین به نظر نمی‌رسد نیازی به تعیین اندازه نمونه ثابت وجود داشته باشد در نتیجه اندازه‌ی نمونه ثابت دارای اهمیت کمتری است.

دلیل دیگر می‌تواند این باشد که مشکلات محاسباتی مانع استفاده از شرایط مرتبط با «دقت» یک برآورد یاب می‌گردد این گونه محاسبات مکرراً در روش کلاسیک مورد استفاده قرار می‌گیرند. به

---

[1] $p(x) = \int p(x|\theta)p(\theta)dx$

[2] $p(x) = \int p(x|\theta)p(\theta)dx$

[3] Pham- Gia

عنوان مثال، محاسبه فاصله‌های اطمینان پسین با استفاده از فاصله اطمینان چگالی پسین، نیازمند جداول

خاص مشابه جداولی که در سال ۱۹۷٤ ایزاکس و همکاران[1] [٦] معرفی کردند می‌باشد یا نیاز به نرم افزارهای خاص دارد. توجه کنید که مسئله محاسبه‌ی اندازه نمونه در این حالت با توجه به شرایطی که روی طول این فاصله‌ها قرار داده می‌شود، مورد بررسی و تجزیه و تحلیل قرار می‌گیرد. مثلاً اندازه‌ی نمونه را به نحوی تعیین می‌کنیم که طول این فاصله‌ها کمتر از یک کمیت از قبل تعیین شده باشد که این روش، مسئله مشخص کردن اندازه نمونه را به یک فرآیند بازگشتی پیچیده تبدیل می-کند.

تعدادی از محققین روش‌های بیز استنباطی را برای مسئله محاسبه اندازه نمونه مورد بررسی قرار دادند. برای میانگین توزیع نرمال، در سال ۱۹۸۸ آدکودک[2] [۲۳] فرمول‌های فشرده اندازه نمونه را برای هردو حالتی که واریانس جمعیت شناخته شده و یا ناشناخته باشد، ارائه داد. او از میانگین گیری پوشش فاصله‌های اطمینان پسین با طول ثابت روی توزیع پیش‌بینی کننده[3] داده‌ها استفاده نمود. در سال ۱۹۷۷جوزف و ربلیسل[4] [۱۲] از ایده‌ای مشابه استفاده نمودند و فرمول‌هایی برای حالت میانگین طول‌های فاصله‌های اطمینان پسین با پوشش ثابت ارائه دادند. این روش‌ها به اختصار در بخش‌های

[1] Isaacs et al

[2] Adcock

[3] Predictive Distribution

[4] Joseph and Belisle

بعدی مورد بررسی قرار می‌گیرند. به دلیل اینکه در این پایان‌نامه «اندازه‌ی نمونـه» را از دیـدگاه بیـزی بررسی می‌کنیم ولی ضروری است مطالب کلی راجع به محاسبه‌ی اندازه‌ی نمونه به روش کلاسیک و نیز در مورد روش بیز ارائه شود.

۱-۲- اندازه نمونه به روش کلاسیک

در فرمول‌های خطای معیار و فاصله‌های اطمینـان انـدازه نمونـه دارای عامـل $\sqrt{n}$ در مخـرج کسرها هستند بنابراین با افزایش اندازه نمونه خطای معیار کاهش می‌یابد و فاصـله‌هـای اطمینـان نیـز کوتاه‌تر می‌شوند. از طرف دیگر، هم در عمل نمونه‌گیری و هم در پردازش داده‌هـا افـزایش انـدازه‌ی نمونه باعث افزایش مخارج و صرف وقت بیشتر خواهد شـد. در مرحلـه‌ی برنامـه‌ریـزی تحقیـق، در مورد اندازه‌ای که باید نمونه داشته باشد تا دقت مطلوب در روش برآورد حاصل گردد، تصمیم‌گیـری می‌شود؛ این تصمیم از لحاظ عملی اهمیت دارد.

ابتدا نحوه‌ی گزینش اندازه‌ی نمونه برای برآورد میانگین جامعه را بررسی می‌کنیم. بـا توجـه به هدف برآورد، خطای قابل اغماض معین و بعلاوه میزان احتمالی مشخص می‌شود که با آن احتمال، این کران خطا را می‌توان ثابت نگاه داشت فرض کنید که انحـراف معیـار جامعـه (σ) معلـوم باشـد. فرمول کران بالای خطای $(1-\alpha)$ ۱۰۰٪ در برآورد μ به وسیله‌ی $\overline{X}$، به صورت زیر است:

$$Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

برای اینکه $(1-\alpha)$ ۱۰۰٪ مطمئن باشیم که خطا بیشتر از مقدار d نمی‌شود، بایستی داشته باشیم $Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}=d$ این رابطه، معادله‌ای است که در آن n مجهول است. این معادله را نسبت به n حل می‌کنیم و اندازه نمونه‌ی لازم را بدست می‌آوریم:

$$n=\left[\frac{Z_{\frac{\alpha}{2}}\sigma}{d}\right]^2 \qquad (۱-۱)$$

البته، چون اندازه نمونه نمی‌تواند کسری باشد جواب را به اولین عدد صحیح بزرگ‌تر گرد می‌کنیم.

ممکن است محقق بخواهد اندازه نمونه لازم را طوری تعیین کند که فاصله‌ی اطمینان $(1-\alpha)$ ۱۰۰٪ برای μ، دارای طول معین، باشد. چون طول فاصله‌ی اطمینان

$$(\bar{X}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\bar{X}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}})$$

برابر با

$$2Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

است، اندازه نمونه n از رابطه‌ی زیر بدست می‌آید:

$$۲Z_{\frac{\alpha}{۲}}\frac{\sigma}{\sqrt{n}}=c \ , \ Z_{\frac{\alpha}{۲}}\frac{\sigma}{\sqrt{n}} : \qquad (۲-۱)$$

اگر $\frac{c}{2}$ را با d (کران خطای معیار) یکی بگیریم، می‌بینیم که این دو نوع فرمول بندی هم ارزند.

در این راه حل، از توزیع نرمال برای $\bar{x}$ استفاده کردیم لذا، این روش وقتی اعتبار دارد که توزیع جامعه نرمال و یا n بزرگ باشد. در غیر این موارد می‌توان به کمک نابرابری چبیشف به راه حلی محافظه کارانه مطابق فرمول ذیل دست یافت.

$$p[\left|\overline{X}-\mu\right| \le d] \ge 1-\frac{V(\overline{X})}{d^2} = 1-\frac{\sigma^2}{nd^2}$$

($1-\frac{\sigma^2}{nd^2}$) را با سطح اطمینان مطلوب ($1-\alpha$) مساوی قرار می‌دهیم و اندازه نمونه به صـورت زیـر بدست می‌آید:

$$n = \frac{\sigma^2}{\alpha d^2}$$

بنابراین ($1-\alpha$) ۱۰۰٪ مطمئن باشیم که خطای $\left|\bar{x}-\mu\right|$ از d بزرگ‌تر نیست، اندازه نمونه لازم، بایـد برابر باشد با

$$n = \left[\frac{Z_{\frac{\alpha}{2}}\sigma}{d}\right]^2 \qquad (۱-۳)$$

اگر n کوچک و جامعه غیر نرمال باشد

$$\frac{\sigma^2}{\alpha d^2} \qquad (۱-٤)$$

را به عنوان کران بالایی اختیار می‌کنیم.

قضیه بیز ۱-۱- (احتمالات پیشین و پسین):

ممکن است شما احتمال وقوع پیشامدی را در شرایط معمولی بدانید ولی اگر اطلاعات جدیدی بدست آورید در احتمال وقوع پیشامد اولیه تجدید نظر کنید. به احتمال وقوع پیشامدی قبل از کسب اطلاعات جدید «احتمال پیشین» و به احتمال وقوع آن پیشامد بعد از کسب اطلاعات جدید «احتمال پسین» می‌گوییم این گونه احتمال در نظریه تصمیم‌گیری کاربرد فراوان دارد و قضیه بیز، را در تجدید نظر احتمالات، در صورت دسترسی به اطلاعات جدید، کمک می‌کند.

با استفاده از تعریف احتمال شرطی و فرمول تفکیک احتمال داریم

$$P(A_j|B)=\frac{P(A_j\cap B)}{P(B)}=\frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n}p(B|A_i)P(A_i)} \qquad (۵-۱)$$

این فرمول را یک کشیش انگلیسی به نام توماس بیز در سال ۱۷۶۳ میلادی پیدا کرد. معمولاً $P(A_j)$

را احتمال پیشین و را احتمال پسین پیشامد A_j می‌نامند.

تعریف ۱-۱. چگالی پیشین: چگالی A را با $\pi(A)$ نشان می‌دهیم و آن را چگالی پیشین[1] می‌نامیم.

تعریف ۱-۲. چگالی پسین: چگالی شرطی A را به شرط داشتن x با $\pi(A|X)$ نشان می‌دهیم و آن‌را چگالی پسین[1] Aمی‌نامیم.

[1] Prior Density

قضیه‌ی ۱-۲. چگالی پسین: فرض کنید $X=(X_1,\ldots,X_n)$ با یافته $x=(x_1,\ldots,x_n)$ یک نمونه تصادفی از X با چگالی $\theta \in A$ ،$f(x|\theta)$ باشد اگر $L(\theta),\pi(\theta|X),\pi(\theta)$ به ترتیب چگالی پیشین و چگالی پسین و تابع درستنمایی باشند، آنگاه داریم

$$\pi(\theta|X) \propto \pi(\theta)L(\theta) \qquad (۱-۶)$$

این رابطه می‌گوید که چگالی پسین متناسب است با حاصل ضرب چگالی پیشین و تابع درستنمایی.

مثال ۱-۱: فرض کنید n نمونه تصادفی $x_1 \ldots\ x_n$ مستقل و هم توزیع با $X \sim N(\mu,\sigma^2)$ باشند. نیز فرض کنید σ^2 معلوم و μ دارای چگالی پیشین $N\ (\mu_0,{\sigma_0}^2)$ باشد پس داریم

$$\pi(\mu)=(2\pi\sigma_0^{\ 2})^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\frac{(\mu-\mu_0)^2}{\sigma_0}\right\}$$

$$L(\mu|X)=(2\pi\sigma^2)^{-\frac{n}{2}}\exp\left\{-\frac{\sum[(x_i-\bar{x})^2+n(\bar{x}-\mu)]^2}{2\sigma^2}\right\}$$

طبق قضیه‌ی ۱-۱ و رابطه‌ای (۱-۶) داریم

$$\pi(\mu|x)=\exp\left\{-\frac{1}{2}\left(\frac{(\mu-\mu_0)^2}{\sigma_0}+\frac{(\mu-\bar{x})^2}{\sigma^2}\right)\right\}$$

$$\pi(\mu|x)=\exp\left\{-\frac{1}{2\sigma_p^2}(\mu-\mu_p)^2\right\}$$

---

[1] Posterior Density

در نتیجه

$$\sigma_p^2 = \frac{\sigma^2 \sigma_0^2}{\sigma^2 + n\sigma_0^2} \quad \text{و} \quad \mu_p = \frac{\mu_0 \sigma^2 + n\bar{x}\sigma_0^2}{\sigma^2 + n\sigma_0^2}$$

مثال۱-۲ چگالی پسین برای پارامتر توزیع دوجمله ای: نسبت P را به عنوان یک متغیر با توزیع بتا با پارامترهای β, α در نظر بگیرید. جهت کسب اطلاعات بیشتر از p ، n نمونه‌ی مستقل از جامعه‌ای با توزیع برنولی با پارامتر p انتخاب کرده و فرض کنید کهx برآمد مطلوب از n آزمایش حاصل شده است به این ترتیب می‌توان نوشت

$$\pi(p|x)\alpha\, p^x(1-p)^{n-x}p^{\alpha-1}(1-p)^{\beta-1} \propto p^{\alpha+x-1}(1-p)^{n+\beta-x-1}$$
$$0 \le p \le 1 \quad , \alpha, \beta \ge 0$$

تعریف ۱-۳. خانواده چگالی مزدوج: فرض کنید f کلاس توابع چگالی $f(x|\theta)$ باشد. کلاس P از چگالی‌های پیشین را خانواده مزدوج برای f می‌گوییم هر گاه چگالی پسین $P(\theta|x)$ به ازای هر $f \in F$ و هر چگالی پیشین در p، در کلاس Pباشد.

۱-۳- روش بیز

به توجه به رابطه (۱-۱) و (۱-۲) داریم:

$$n = \left[\frac{2Z_{\frac{\alpha}{2}}\sigma}{C}\right]^2 \qquad (۷-۱)$$

که محدودیت‌هایی را در بردارد

– اندازه‌ی نمونه‌ای که از رابطه (۱–۷) بدست می‌آید به σ^2 بستگی دارد که مقدار σ قبل از انجام آزمایش نامعلوم است.

–صرف نظر از مقدار σنتایج نهایی در این زمینه بر اساس داده‌های مشاهده شده انجام و محاسبه شده‌اند(برآوردها)

روش بیز محدودیت‌های مذکور را ندارد و با در نظر گرفتن اطلاعات پیشین تعیین اندازه نمونه را ممکن می‌سازد.

فصل دوم

تعیین اندازه‌ی نمونه به روش بیز

۲-۱. مقدمه و مفاهیم اولیه

در این فصل نیز بحث تعیین اندازه‌ی نمونه را دنبال می‌کنیم با این تفاوت که در اینجا از دیدگاه فاصله اطمینان چگالی پسین به موضوع می‌پردازیم روش‌های ارائه شده در این فصل با اعمال شرایطی بر فاصله اطمینان چگالی پسین ملاک‌هایی را برای تعیین اندازه نمونه بیان می‌کنند. البته این روش‌ها کافی هستند چرا که فاصله اطمینان چگالی پسین برای احتمال پوشش معین شده‌ای به کوتاه‌ترین فواصل منجر می‌شوند و خلاصه‌ی مناسبی را از اطلاعات پسین راجع به θ فراهم می‌کنند.

روش‌های معرفی شده در این فصل را می‌توان به طور خلاصه به صورت زیر بیان کرد:

۱- با در نظر گرفتن طول ثابتی برای فاصله اطمینان چگالی پسین کمترین مقدار n را برای داشتن حداقل احتمال پوشش (α-۱) بدست می‌آوریم.

۲- با در نظر گرفتن احتمال پوشش ثابتی برای فاصله اطمینان چگالی پسین کمترین مقدار n را به گونه ای بدست می‌آوریم که طول فاصله حداکثر L باشد.

۳- در این روش محافظه کارانه، n را به گونه ای انتخاب می‌کنیم که برای مجموعه‌ای از محتمل‌ترین داده‌های توزیع پیشین به طور همزمان فاصله به اندازه‌ی کافی کوچک و احتمال پوشش به اندازه‌ی کافی بزرگ باشد.

با توجه به مطلب بالا و بستگی روش‌های بیان شده به فاصله اطمینان چگالی پسین لازم است که با این نواحی و برخی خواص آن‌ها مختصر آشنا شویم.