





دانشگاه محقق اردبیلی
دانشکده‌ی علوم ریاضی
گروه آمار و کاربردها

پایان‌نامه برای دریافت درجه کارشناسی ارشد
در رشته‌ی آمار ریاضی

عنوان:

برآورد بیز و بیز تجربی پارامتر n در توزیع‌های دوجمله‌ای و دوجمله‌ای منفی

استاد راهنما:

دکتر مسعود گنجی

استاد مشاور:

دکتر نسرین اقبالی

پژوهشگر:

مسعود عظیمیان

بهمن ماه ۱۳۹۲

تقدیم بہ پدر و مادر مہربانم

و کسانی کہ دوستان می دارم

خدایا...^۱

به من زیستنی عطا کن که در لحظه مرگ، بر بی‌ثمری لحظه‌ای که برای زیستن گذشته است، حسرت نخورم و مُردنی عطا کن که بر بیهودگیش، سوگوار نباشم. بگذار تا آن را، خود انتخاب کنم، اما آنچنان که تو دوست می‌داری.

تو می‌دانی و همه می‌دانند که شکنجه دیدن بخاطر تو، زندانی کشیدن بخاطر تو و رنج بردن به پای تو تنها لذت بزرگ زندگی من است، از شادی توست که من در دل می‌خندم، از امید رهایی توست که برق امید در چشمان خسته‌ام می‌درخشد و از خوشبختی توست که هوای پاک سعادت را در ریه‌هایم احساس می‌کنم. نمی‌توانم خوب حرف بزنم. نیروی شگفتی را که در زیر کلمات ساده و جمله‌های ضعیف و افتاده، پنهان کرده‌ام دریاب، دریاب.

تو می‌دانی و همه می‌دانند که زندگی از تحمیل لبخندی بر لبان من، از آوردن برق امیدی در نگاه من، از برانگیختن موج شعفی در دل من، عاجز است.

تو، چگونه زیستن را به من بیاموز، چگونه مردن را خود خواهم آموخت. به من توفیق تلاش در شکست، صبر در نومیدی، رفتن بی‌همراه، جهاد بی‌سلاح، کار بی‌پاداش، فداکاری در سکوت، دین بی‌دنیا، مذهب بی‌عوام، عظمت بی‌نام، خدمت بی‌نان، ایمان بی‌ریا، خوبی بی‌نمود، گستاخی بی‌خامی، قناعت بی‌غرور، عشق بی‌هوس، تنهایی در انبوه جمعیت، و دوست داشتن بی‌توقع، روزی کن.

اگر تنها ترین تنها شوم، باز خدا هست

او جانشین همه نداشتن هست...

^۱ مناجاتی از دکتر علی شریعتی.

سپاس‌گزاری...

سپاس خداوندگار حکیم را که با لطف بی‌کران خود، آدمی را زیور عقل آراست. در آغاز وظیفه خود می‌دانم از زحمات بی‌دریغ استاد راهنمای فرزانه‌ی خود، جناب آقای دکتر مسعود گنجی صمیمانه تشکر و قدردانی کنم که قطعاً بدون راهنمایی‌های ارزنده ایشان، این مجموعه به انجام نمی‌رسید.

از سرکارخانم دکتر نسرین اقبالی که زحمت مطالعه و مشاوره این رساله را تقبل فرمودند و در آماده‌سازی این پایان‌نامه، به نحو احسن اینجانب را مورد راهنمایی قرار دادند، کمال تشکر را دارم. از استاد گرانقدر جناب آقای دکتر ولی زاردشت که زحمت مطالعه و داوری این پایان‌نامه را تقبل فرمودند، سپاسگذاری می‌کنم.

از آقای دکتر صفر پارسی نیز که همواره از هیچ کمکی دریغ ننموده‌اند کمال تشکر را دارم. و در پایان، بوسه می‌زنم بر دستان خداوندگاران مهر و مهربانی، پدر و مادر عزیزم و بعد از خدا، ستایش می‌کنم وجود مقدس‌شان را و تشکر می‌کنم به پاس عاطفه سرشار و گرمای امیدبخش وجودشان، که در این سردترین روزگاران، بهترین پشتیبان من بودند. همچنین از تمامی دوستان دوران تحصیل که همواره یار و همدمم بوده‌اند کمال تشکر را دارم.

خدایا چنان کن سرانجام کار تو خوشنود باشی و ما رستگار.

مسعود عظیمیان

زمستان ۱۳۹۲

نام خانوادگی: عظیمیان

نام: مسعود

عنوان پایان نامه:

برآورد بیز و بیز تجربی پارامتر n در توزیع های دوجمله ای و دوجمله ای منفی

استاد راهنما: دکتر مسعود گنجی

استاد مشاور: دکتر نسرین اقبالی

مقطع تحصیلی: کارشناسی ارشد

رشته: آمار ریاضی

گرایش: محض

دانشگاه: محقق اردبیلی

دانشکده: علوم ریاضی

تاریخ دفاع: ۹۲/۱۱/۰۵

تعداد صفحات: ۱۳۶

چکیده

در این پایان نامه مسئله برآورد پارامتر n در توزیع های دوجمله ای و دوجمله ای منفی از دیدگاه بیزی در دو حالت معلوم و نامعلوم بودن پارامتر θ (احتمال موفقیت) بررسی گردیده است. پارامتر n در توزیع دوجمله ای تعداد آزمایش ها و در توزیع دوجمله ای منفی تعداد موفقیت ها را در یک دنباله از آزمایش های برنولی نشان می دهد. برآوردگریز و بیز تجربی پارامتر n ، در هر توزیع با در نظر گرفتن یک توزیع پیشین از چپ بریده شده برای n و یک توزیع بتا برای θ تحت تابع زیان مربع خطا بدست آورده شده اند. روش های دیگر برآورد، از جمله روش گشتاوری و ماکسیمم درستنمایی و غیره که توسط محققان دیگر بکار برده شده اند نیز معرفی گردیده است. در پایان برآوردگرهای بیز و بیز تجربی بدست آمده با برآوردگرهای قبلی مقایسه می گردد.

کلیدواژه ها: برآورد بیزی، برآورد بیز تجربی، توزیع دوجمله ای، توزیع دوجمله ای منفی

فهرست مطالب

ج	لیست جداول	۱
د	لیست تصاویر	۱
و	پیشگفتار	۱
۱	تعاریف و مفاهیم اولیه	۱
۲	۱.۱ مقدمه	۱
۲	۲.۱ روش‌های برآوردیابی	۱
۵	۳.۱ قضایا و نامساوی‌های معروف	۱
۵	۴.۱ توزیع‌های آماری	۱
۶	۵.۱ روش‌های محاسبه‌ی عددی انتگرال	۱
۷	۶.۱ استنباط بیزی	۱
۸	۱.۶.۱ مخاطره و برآورد بیزی	۱
۹	۲.۶.۱ روش بدست آوردن برآوردگر بیزی	۱
۱۳	۲ روش‌های بیز تجربی	۱
۱۴	۱.۲ مقدمه	۱
۱۶	۲.۲ برآورد بیز تجربی	۱
۲۴	۳ برآورد بیز و بیز تجربی پارامتر n در توزیع دوجمله‌ای	۱
۲۵	۱.۳ مقدمه	۱
۲۸	۲.۳ برآوردهای کلاسیک پارامتر n	۱
۲۸	۱.۲.۳ روش گشتاوری	۱
۲۹	۲.۲.۳ روش ماکسیمم درست‌نمایی	۱
۳۳	۳.۲.۳ برآوردگر کارول و لومبارد	۱
۳۵	۴.۲.۳ مثال عددی	۱

۳۶	برآوردهای بیزی پارامتر n	۳.۳
۳۶	برآوردگر اردوغان و دانیل	۱.۳.۳
۴۳	مثال عددی	۲.۳.۳
۴۴	برآورد بیز و بیز تجربی پارامتر n با انتخاب پیشین پواسن	۳.۳.۳
۵۹	مثال عددی	۴.۳.۳
۶۳	شبیه‌سازی	۴.۳
۷۱	نتیجه‌گیری	۵.۳
۷۲	برآورد بیز و بیز تجربی پارامتر n در توزیع دوجمله‌ای منفی	۴
۷۳	مقدمه	۱.۴
۷۵	روش‌های برآورد پارامتر n	۲.۴
۷۵	برآوردگر روش گشتاوری	۱.۲.۴
۷۶	برآوردگر روش ماکسیمم درست‌نمایی	۲.۲.۴
۷۸	برآوردگر بیز و برآوردگر بیز تجربی	۳.۲.۴
۸۳	برآوردگر نقطه‌ای بیز تجربی پارامتر n	۴.۲.۴
۹۴	شبیه‌سازی	۳.۴
۱۰۰	نتیجه‌گیری	۴.۴
۱۰۱	مراجع	
۱۰۴	پیوست الف	
۱۰۶	پیوست ب	
۱۳۵	واژه‌نامه فارسی به انگلیسی	

لیست جداول

۱۷	۱.۲	تعداد بطریه‌های شامل باکتری در هر آزمایش که توسط وون مایس (۱۹۴۳) جمع‌آوری شده‌اند.
۳۵	۱.۳	برآوردها برای نمونه‌های اصلی و اغتشاش یافته برای بررسی برآوردگر کارول لومبارد
۴۴	۲.۳	برآوردها برای نمونه‌های اصلی و اغتشاش یافته برای بررسی برآوردگر اردوغان و دانیل
۶۲	۳.۳	مقادیر برآوردهای \hat{n}_1 ، CLE و $EDBE$ برای مقادیر مختلف α و β در مثال داده‌شده
۶۴	۴.۳	مقادیر میانگین برآوردهای \hat{n}_1 ، $\hat{n}_1(emp, \alpha)$ ، CLE ، $EDBE$ و مقدار MSE آنها در داخل پرانتز برای حالتی که θ نامعلوم است:
۶۶	۵.۳	مقادیر میانگین برآوردهای \hat{n}_2 ، MME ، MLE و مقدار MSE آنها برای حالتی که θ معلوم است:
۹۰	۱.۴	مقادیر توزیع پسین برای حالتی که θ نامعلوم است
۹۲	۲.۴	مقادیر $h_N(n)$ برای حالتی که θ معلوم است
۹۸	۳.۴	مقادیر میانگین برآوردهای \hat{n}_1 ، MM ، ML و مقدار MSE آنها در داخل پرانتز برای حالتی که θ نامعلوم است
۹۹	۴.۴	مقادیر میانگین برآورد \hat{n}_1 و مقدار MSE آن در داخل پرانتز برای مقادیر متفاوت α
۹۹	۵.۴	مقادیر میانگین برآوردهای \hat{n}_2 ، MME ، MLE و مقدار MSE آنها در داخل پرانتز برای حالتی که θ معلوم است

لیست تصاویر

۳۱	نمودار تابع $D(n)$ براساس نمونه بدست آمده از توزیع دوجمله‌ای با $\theta = 0/45$	۱.۳
۶۰	نمودار $S(n)$ در برابر مقادیر n	۲.۳
۶۷	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/2$ نامعلوم برای نمونه‌های با اندازه ۳۰	۳.۳
۶۷	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/5$ نامعلوم برای نمونه‌های با اندازه ۳۰	۴.۳
۶۸	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/8$ نامعلوم برای نمونه‌های با اندازه ۳۰	۵.۳
۶۸	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/2$ معلوم برای نمونه‌های با اندازه ۳۰	۶.۳
۶۸	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/5$ معلوم برای نمونه‌های با اندازه ۳۰	۷.۳
۶۹	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/8$ معلوم برای نمونه‌های با اندازه ۳۰	۸.۳
۷۰	اندازه ۳۰	
۸۹	نمودار $S(n)$ در برابر مقادیر n	۱.۴
۹۰	نمودار مقادیر $h_N(n)$ برای حالتی که θ نامعلوم است	۲.۴
۹۲	نمودار مقادیر $Y(n)$ در برابر n	۳.۴
۹۳	مقادیر $h_N(n)$ برای حالتی که θ معلوم است	۴.۴
۹۵	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/2$ نامعلوم برای نمونه‌های با اندازه ۳۰	۵.۴
۹۵	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/5$ نامعلوم برای نمونه‌های با اندازه ۳۰	۶.۴
۹۵	نمودار مقایسه MSE برآوردهای n برای $\theta = 0/8$ نامعلوم برای نمونه‌های با اندازه ۳۰	۷.۴
۹۶	اندازه ۳۰	

۸.۴	نمودار مقایسه MSE برآوردهای n برای $\theta = ۰/۲$ معلوم برای نمونه‌های با
۹۶	اندازه ۱۰
۹.۴	نمودار مقایسه MSE برآوردهای n برای $\theta = ۰/۵$ معلوم برای نمونه‌های با
۹۷	اندازه ۳۰
۱۰.۴	نمودار مقایسه MSE برآوردهای n برای $\theta = ۰/۸$ معلوم برای نمونه‌های با
۹۷	اندازه ۱۰

پیشگفتار

برآورد پارامتر n در توزیع دوجمله‌ای به عنوان یک مسئله، بیش از نیم قرن است که توجه محققین را به خود جلب کرده است. درباره‌ی برآورد پارامتر n وقتی θ (احتمال موفقیت) معلوم یا نامعلوم است مقالات بسیاری به چاپ رسیده است. اما این مسئله هنوز هم حداقل به چهار دلیل زیر حائز اهمیت است:

- ۱- کم برآورد^۱ شدید n که یک مشکل اساسی در کاربردهای عملی است.
- ۲- عدم وجود راه حلی با محاسبات ساده و مقایسه‌ی برآوردهای معرفی شده و انتخاب یکی به عنوان برآورد بهینه.
- ۳- ناپایداری برآوردهای معرفی شده (به گونه‌ای که با تغییر کوچکی در یک یا دو مشاهده از نمونه شاهد نوسانات بسیار زیادی در برآورد پارامتر n هستیم).
- ۴- کاربردهای بسیار این مدل احتمال در زیست‌شناسی و هر زمینه‌ی دیگری که در آن تعداد کل یک جمعیت متناهی مجهول بوده و بخواهیم بر اساس دنباله‌ای از مشاهدات آن را برآورد کنیم. برخی از آماردانان این مسئله را از دیدگاه کلاسیک و برخی دیگر آن را از دیدگاه بیزی مورد بررسی قرار داده‌اند. در این پایان‌نامه این برآوردها را معرفی و به مقایسه آن‌ها از لحاظ معیار میانگین مربع خطا می‌پردازیم. این پایان‌نامه بر اساس مقاله‌ی بایوود^۲ (۲۰۱۱) تدوین شده است. مسئله‌ی دیگر که در اینجا مورد بررسی قرار گرفته است برآورد پارامتر n (تعداد موفقیت‌ها) در

^۱ underestimate

^۲ Bayoud

توزیع دوجمله‌ای منفی است که با روش بیز و روش بیز تجربی به دست آورده می‌شود. این کار در فصل چهارم از این پایان‌نامه ارائه گردیده که حاصل آن یک مقاله می‌باشد (گنجی و همکاران^۱ (۲۰۱۳)). این پایان‌نامه شامل چهار فصل می‌باشد: در فصل‌های اول و دوم تعاریف و موضوعات لازم که برای فصل‌های بعدی مورد نیاز است، بیان می‌گردد. در فصل سوم برآوردهای پارامتر n در توزیع دوجمله‌ای و مقایسه آنها بیان می‌گردد و در فصل چهارم همانطور که گفتیم برآوردهای پارامتر n در توزیع دوجمله‌ای منفی معرفی شده است. در این پایان‌نامه از نرم‌افزار متمتیکا^۲ برای محاسبات عددی و شبیه‌سازی استفاده شده است.

^۱Ganji et al.

^۲Mathematica

فصل ۱

تعاريف و مفاهيم اوليه

۱.۱ مقدمه

در این فصل مفاهیم و تعاریف اولیه و همچنین موضوعات و قضایای لازم که در فصل‌های بعدی لازم می‌باشد را بیان می‌کنیم. در واقع این فصل شامل بخش‌های روش‌های برآوردیابی، قضایا و نامساوی‌های معروف، توزیع‌های آماری، روش‌های محاسبه‌ی عددی انتگرال و استنباط بیزی می‌باشد. در هر کدام از این بخش‌ها، تعاریف و قضایای مربوط به هر بخش بیان شده‌است.

۲.۱ روش‌های برآوردیابی

تعریف ۱.۲.۱. تابع درست‌نمایی^۱: فرض کنید $\theta \in \Theta \subseteq \mathbb{R}^k$ فضای پارامترهای نامعلوم است) و $\mathbf{X} = (X_1, \dots, X_n)$ بردار n متغیر تصادفی با تابع چگالی احتمال توأم $f_\theta(\mathbf{x})$ باشد. برای هر مقدار داده شده‌ی $\mathbf{X} = \mathbf{x}$ ، تابع درست‌نمایی \mathbf{x} را تابع چگالی احتمال توأم \mathbf{X} ، یعنی $f_\theta(\mathbf{x})$ ، تعریف می‌کنیم که به صورت تابعی از θ در نظر گرفته می‌شود و آن را با نماد $L(\theta)$ نمایش می‌دهیم، به عبارتی $L(\theta) = f_\theta(\mathbf{x})$.

تعریف ۲.۲.۱. برآورد ماکسیمم درست‌نمایی: اگر $\delta(\mathbf{X})$ برآوردگری برای θ باشد به طوری که

$$(1) \text{ برای هر } \theta \in \Theta, P_\theta(\delta(\mathbf{X}) \in \Theta) = 1,$$

$$(2) \text{ برای هر } \theta \in \Theta, L(\delta(\mathbf{X})) \geq L(\theta),$$

آنگاه $\delta(\mathbf{X})$ به عنوان برآوردگر ماکسیمم درست‌نمایی^۲ (MLE) θ تعریف می‌شود.

در واقع روش ماکسیمم درست‌نمایی عبارت است از ماکسیمم کردن تابع درست‌نمایی نسبت به θ ، و آن مقدار θ که تابع درست‌نمایی را ماکسیمم می‌کند، برآورد ماکسیمم درست‌نمایی θ می‌نامیم.

^۱Likelihood function

^۲Maximum Likelihood estimation

تعریف ۳.۲.۱. برآورد گشتاوری: فرض کنید $\mathbf{X} = (X_1, \dots, X_n)$ یک نمونه‌ی تصادفی n تایی از توزیع F_θ باشد به طوری که $\theta = (\theta_1, \dots, \theta_k) \in \Theta \subseteq \mathbb{R}^k$. همچنین فرض کنید k گشتاور اول این توزیع که به صورت توابعی از θ هستند، وجود داشته باشد. می‌دانیم که گشتاور r -ام، در صورت وجود به صورت زیر تعریف می‌شود:

$$\begin{aligned}\mu_r &= \mu_r(\theta) \\ &= E_\theta(X_1^r) \quad r = 1, \dots, k\end{aligned}$$

اگر

$$M_r = \frac{1}{n} \sum_{i=1}^n X_i^r \quad r = 1, \dots, k;$$

نمایانگر r -امین گشتاور نمونه‌ای بر پایه‌ی نمونه‌ی تصادفی داده شده باشد، آنگاه برآوردهای گشتاوری پارامترهای مجهول $\theta = (\theta_1, \dots, \theta_k)$ از حل k معادله‌ی زیر حاصل می‌شوند:

$$\mu_r = M_r \quad r = 1, \dots, k;$$

معمولاً برآورد گشتاوری^۱ (MME) θ را با $\tilde{\theta}$ نمایش می‌دهند و لزومی ندارد که منحصر به فرد باشد.

تعریف ۴.۲.۱. فضای برآوردگرها: مجموعه کلیه‌ی برآوردگرهای ممکن پارامتر را فضای برآوردگرها (کلاس برآوردگرها) گوئیم و با D نشان می‌دهیم.

تعریف ۵.۲.۱. برآوردگر پایدار^۲ و برآوردگر ناپایدار^۳: یک برآوردگر را که براساس نمونه تصادفی $\mathbf{X} = (X_1, \dots, X_n)$ بدست آمده است را ناپایدار گوئید، اگر نسبت به تغییرات کوچک در مقادیر نمونه بسیار حساس باشد و با تغییر نمونه دچار تغییرات بسیار زیادی شود و در غیر اینصورت آن را پایدار گوئید.

^۱Method of moment estimation

^۲stable

^۳unstable

تعریف ۶.۲.۱. اگر $\mathbf{X} = (X_1, X_2, \dots, X_k)$ یک نمونه تصادفی باشد، آنگاه آماره‌های ترتیبی متناظر را با $X_{(1)}, X_{(2)}, \dots, X_{(k)}$ نشان می‌دهیم که در آن $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(k)}$.

تعریف ۷.۲.۱. تابع زیان^۱: فرض کنید در برآورد پارامتر $\theta \in \Theta$ برآوردگر $\delta(\mathbf{X}) \in D$ را در نظر بگیریم. مهمترین خصوصیت یک برآوردگر خوب آن است که برآوردی را برای θ بدست دهد که تا حد ممکن به مقدار واقعی θ نزدیک باشد. میزان این نزدیکی را با تابع زیان اندازه گیری می‌کنیم که با تابع $L(\theta, \delta(\mathbf{X})) : \Theta \times D \rightarrow [0, +\infty)$ نمایش داده می‌شود که برای هر θ و $\delta(\mathbf{X})$ ، $L(\theta, \delta(\mathbf{X})) \geq 0$ و اگر $\theta = \delta(\mathbf{X})$ آنگاه $L(\theta, \delta(\mathbf{X})) = 0$. چند تابع زیان مشهور عبارتند از:

$$L(\theta, \delta(\mathbf{X})) = (\delta(\mathbf{X}) - \theta)^2 \quad \text{۱- تابع زیان مربع خطا}$$

$$L(\theta, \delta(\mathbf{X})) = |\delta(\mathbf{X}) - \theta| \quad \text{۲- تابع زیان قدرمطلق خطا}$$

$$L(\theta, \delta(\mathbf{X})) = b(e^{a(\delta(\mathbf{X})-\theta)} - a(\delta(\mathbf{X})-\theta) - 1) \quad \text{۳- تابع زیان لاینکس^۲}$$

چون تابع زیان بستگی به $\delta(\mathbf{X})$ دارد پس مقدار آن ثابت نیست و یک متغیر تصادفی است. برای مقایسه دو برآوردگر، امید ریاضی تابع زیان این دو برآوردگر را که تابعی از θ می‌باشد با یکدیگر مقایسه می‌کنند که به آن تابع مخاطره^۳ گویند که به صورت زیر تعریف می‌شود:

$$R(\theta, \delta) = E_{\theta}[L(\theta, \delta(\mathbf{X}))].$$

تعریف ۸.۲.۱. تابع توزیع تجربی^۴: فرض کنید X_1, X_2, \dots, X_n متغیرهای تصادفی مستقل و هم توزیع با تابع توزیع تجمعی $F(t)$ باشند. آنگاه تابع توزیع تجربی بصورت زیر تعریف می‌شود:

$$\hat{F}_n(t) = \frac{\#(X_i \leq t)}{n} = \frac{1}{n} \sum_{i=1}^n I_{(-\infty, t]}(X_i)$$

که در آن $\#(X_i \leq t)$ برابر تعداد x های کوچکتر یا مساوی t است و $I_A(x)$ یک تابع نشانگر برای پیشامد A است که اگر $x \in A$ برابر یک و در غیر اینصورت برابر صفر است. برای یک t ثابت، تابع نشانگر $I_{(-\infty, t]}(X_i)$ ، یک متغیر تصادفی برنولی با پارامتر $p = F(t)$ است. بنابراین $n\hat{F}_n(t)$ یک

^۱ Loss Function

^۲ Linex Function

^۳ Risk Function

^۴ Empirical Distribution Function

متغیر تصادفی دوجمله‌ای با میانگین $nF(t)$ و واریانس $nF(t)(1-F(t))$ است که در واقع بیانگر این است که، $\hat{F}_n(t)$ یک برآوردگر نااریب برای $F(t)$ است. $\hat{F}_n(t)$ یک تابع پله‌ای است که در هر مقدار مشاهده شده به میزان $\frac{1}{n}$ پرش دارد.

۳.۱ قضایا و نامساوی‌های معروف

قضیه ۱.۳.۱. قانون قوی اعداد بزرگ: فرض کنید X_1, X_2, \dots یک دنباله از متغیرهای تصادفی مستقل و هم توزیع روی فضای احتمال (Ω, \mathbb{F}, P) باشند. قرار دهید $S_n = \sum_{i=1}^n X_i$ و $\bar{X}_n = \frac{S_n}{n}$ و $E(X_n) = \mu$ ، آنگاه

$$P[\lim_{n \rightarrow \infty} \bar{X}_n = \mu] = 1$$

یا به طور معادل وقتی $n \rightarrow \infty$ ، تقریباً همه جا $\bar{X}_n \rightarrow \mu$ به عبارتی می‌توان گفت میانگین نمونه در احتمال تقریباً همه جا به میانگین جامعه میل می‌کند.

تعریف ۱.۳.۱. تقریب استرلینگ: به ازای n های بزرگ: $n! \approx \sqrt{2\pi n} n^{n+\frac{1}{2}} e^{-n}$.

۴.۱ توزیع‌های آماری

تعریف ۱.۴.۱. توزیع بتا: اگر متغیر تصادفی X دارای تابع چگالی

$$f_X(x; \alpha, \beta) = \frac{1}{Beta(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1, \alpha > 0, \beta > 0;$$

باشد، آنگاه گوئیم X دارای توزیع بتا است و با نماد $X \sim Beta(\alpha, \beta)$ نشان می‌دهیم. میانگین و واریانس این توزیع بترتیب برابر است با:

$$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}, \quad E(X) = \frac{\alpha}{\alpha+\beta}.$$

تابع توزیع تجمعی متغیر تصادفی توزیع بتا عبارت است از:

$$F_X(x; \alpha, \beta) = I_{(0,1)}(x) \int_0^x \frac{1}{B(\alpha, \beta)} u^{\alpha-1} (1-u)^{\beta-1} du + I_{(1,\infty)}(x).$$

که $\int_0^x u^{\alpha-1} (1-u)^{\beta-1} du$ را بتای ناقص می‌نامند.

تعریف ۲.۴.۱. توزیع بتا-دوجمله‌ای: اگر متغیر تصادفی گسسته‌ی X دارای تابع جرم احتمال

$$f_X(x; n, \alpha, \beta) = \binom{n}{x} \frac{\Gamma(\alpha + \beta) \Gamma(x + \alpha) \Gamma(n + \beta - x)}{\Gamma(\alpha) \Gamma(\beta) \Gamma(n + \alpha + \beta)} I_{\{0, 1, \dots, n\}}(x),$$

باشد که در آن $\alpha > 0, \beta > 0$ و n یک عدد صحیح نامنفی است. گوییم که X دارای توزیع بتا-دوجمله‌ای با پارامترهای n, α و β است و آن را با نماد $BB(n, \alpha, \beta) \sim X$ نمایش می‌دهیم. میانگین و واریانس این توزیع به ترتیب برابر است با:

$$Var(X) = \frac{n\alpha\beta(n + \alpha + \beta)}{(\alpha + \beta)^2(\alpha + \beta + 1)}, \quad E(X) = \frac{n\alpha}{\alpha + \beta}.$$

تعریف ۳.۴.۱. توزیع دوجمله‌ای منفی: اگر متغیر تصادفی گسسته‌ی X دارای تابع جرم احتمال

$$f_X(x; n, \theta) = \binom{n + x - 1}{x} \theta^n (1 - \theta)^x, \quad x = 0, 1, 2, \dots$$

باشد که در آن $n = 1, 2, 3, \dots$ و $0 < \theta < 1$. گوییم که X دارای توزیع دوجمله‌ای منفی است و آن را با نماد $NB(n, \theta) \sim X$ نمایش می‌دهیم. میانگین و واریانس این توزیع به ترتیب برابر است با:

$$Var(X) = \frac{n(1 - \theta)}{(\theta)^2}, \quad E(X) = \frac{n(1 - \theta)}{\theta}.$$

رابطه‌ی بین توزیع‌های دوجمله‌ای و پواسن: اگر X و Y دو متغیر تصادفی مستقل و دارای توزیع پواسن به ترتیب با پارامترهای λ و λ' باشند آنگاه $Z = X + Y$ دارای توزیع پواسن با پارامتر $\lambda + \lambda'$ است و $X|Z = z$ نیز دارای توزیع دوجمله‌ای با پارامترهای z و $\frac{\lambda}{\lambda + \lambda'}$ است.

$$Z = X + Y \sim Poisson(\lambda + \lambda'), \quad X|Z = z \sim Bin(z, \frac{\lambda}{\lambda + \lambda'}).$$

۵.۱ روش‌های محاسبه‌ی عددی انتگرال

در حل مسائل اقتصادی و همچنین در زمینه‌های دیگر معمولاً به انتگرال‌هایی برخورد می‌کنیم که به شکل $\int_{-\infty}^{\infty} e^{-x^2} f(x) dx$ یا $\int_0^{\infty} e^{-x} f(x) dx$ هستند. برای محاسبه‌ی (عددی) این چنین انتگرال‌هایی به

ترتیب از دو روش گاوس-هرمیت^۱ و روش مربع سازی گاوس-لاگر^۲ استفاده می شود. ما در اینجا روش مربع سازی گاوس-لاگر را توضیح می دهیم.

در این روش برای محاسبه ی انتگرال $\int_0^{\infty} e^{-x} f(x) dx$ ، این انتگرال را با مجموع $\sum_{j=1}^m w_j f(x_j)$ تقریب می زنیم، یعنی $\int_0^{\infty} e^{-x} f(x) dx \simeq \sum_{j=1}^m w_j f(x_j)$ که در آن x_j مقادیر صفرهای چند جمله ای لاگر هستند و w_j ها وزن هایی هستند که به صورت $w_j = \frac{(n!)^2}{x_i (L'_m(x_i))^2}$ تعریف می شوند. در ریاضیات چند جمله ای های لاگر عبارتند از حل معادله ی لاگر $(xy'' + (1-x)y' + my = 0)$ ، که معمولاً با دنباله چند جمله ای های L_0, L_1, \dots نشان داده می شوند و جملات این دنباله با فرمول زیر مشخص می شوند:

$$L_m(x) = \frac{e^x}{m!} \frac{d^m}{dx^m} (e^{-x} x^m).$$

به عنوان مثال برای محاسبه $\int_0^1 e^{-x} \sin x dx$ که دارای مقدار ۰٫۵ است با این روش داریم:

$$\int_0^1 e^{-x} \sin x dx \approx \sum_{i=1}^3 w_i(x) \sin x_i = ۰٫۴۹۸۹$$

البته با افزایش m مقدار تقریب نیز دقیق تر می شود. آبراموویتز و استگان^۳ (۱۹۷۰) مقادیر x_j ها و w_j ها را برای مقادیر مختلف m بدست آورده و در یک جدول مرتب کرده اند که برای $m = ۸$ و $m = ۱۵$ جدول مقادیر این بردارها و همچنین برنامه متمتیکای مربوط به تولید آنها را در پیوست الف ضمیمه کرده ایم که برای هر m دلخواه می توان این بردارها را تولید کرد.

۶.۱ استنباط بیزی

همانطور که می دانیم در روش های کلاسیک پارامتر θ را یک مقدار ثابت نامعلوم در نظری می گیریم و یک نمونه ی تصادفی X_1, X_2, \dots, X_n از جامعه ای که دارای توزیع $F_{\theta}(x)$ بود جمع آوری کرده و براساس آن در مورد θ استنباط می کنیم. در روش بیزی θ را یک کمیتی در نظر می گیریم که خود

^۱ Gauss-Hermit

^۲ Gauss-Laguerre quadrature

^۳ Abramowitz and Stegun