دانشگاه پیام نور

دانشکده علوم پایه

پایان نامه

برای دریافت مدرک کارشناسی ارشد

رشته آمار ریاضی

گروه آمار

عنوان پایان نامه :

توزیع مجانبی مجموع وزن‌دار متغیرهای تصادفی نمایی

اسماعیل رضاپوریان

استاد راهنما :

دکتر نرگس عباسی

استاد مشاور :

دکتر محبوبه حسین‌یزدی

شهریور ۱۳۹۱

بسم الله الرحمن الرحیم

دانشگاه پیام نور

دانشکده علوم پایه

مرکز شیراز

پایان نامه

برای دریافت مدرک کارشناسی ارشد

رشته آمار ریاضی

گروه آمار

عنوان پایان نامه :

توزیع مجانبی مجموع وزن‌دار متغیرهای تصادفی نمایی

اسماعیل رضاپوریان

استاد راهنما :

دکتر نرگس عباسی

استاد مشاور :

دکتر محبوبه حسین‌یزدی

شهریور ۱۳۹۱

تاریخ : ۹۱/۰۶/۲۷
شماره : ۰۵/۱۶۲۷۶
پیوست :

(الله)
جمهوری اسلامی ایران
وزارت علوم، تحقیقات و فناوری

دانشگاه پیام نور استان فارس

باسمه تعالی

صورتجلسه دفاع از پایان‌نامه دوره کارشناسی ارشد

جلسه دفاع از پایان‌نامه دوره کارشناسی ارشد **آقای اسماعیل رضاپوریان** دانشجوی رشته **آمار ریاضی** به شماره دانشجویی **۸۹۰۰۶۷۷۲۵** با عنوان:

" توزیع مجانبی مجموع وزن‌دار متغییرهای تصادفی نمایی"

با حضور هیأت داوران در روز **دوشنبه** مورخ **۱۳۹۱/۰۶/۲۷** ساعت **۹** در محل ساختمان **غدیر دانشگاه پیام نور شیراز** برگزار شد و هیأت داوران پس از بررسی، پایان‌نامه مذکور را شایسته نمره به عدد ۱۹٫۲۵ به حروف نوزده و بیست و پنج با درجه عالی تشخیص داد.

ردیف	نام و نام خانوادگی	هیات داوران	مرتبه دانشگاهی	دانشگاه	امضاء
۱	دکتر نرگس عباسی	راهنما	دانشیار	پیام نور شیراز	
۲	دکتر محبوبه حسین‌یزدی	مشاور	استادیار	پیام نور شیراز	
۳	آقای عبدالرضا بازرگان لاری	داور	استادیار	شیراز	
۴	دکتر حسین تولّلی	نماینده تحصیلات تکمیلی	دانشیار	پیام نور شیراز	

رئیس اداره تحصیلات تکمیلی

دانشگاه پیام نور مرکز شیراز
معاونت آموزشی و تحصیلات تکمیلی

شیراز- شهرک گلستان، بلوار دهخدا
قبل از نمایشگاه بین المللی
تلفن : ۳-۰۷۱۱-۶۲۲۲۲۴۰
دورنگار : ۰۷۱۱-۶۲۲۲۲۴۹
صندوق پستی :۱۳۶۸ - ۷۱۹۵۵
www.spnu.ac.ir
Email : admin@spnu.ac.ir

اینجانب اسماعیل رضاپوریان دانشجوی ورودی سال ۸۹ مقطع کارشناسی ارشد رشته‌ی آمار ریاضی گواهی می نمایم چنانچه در پایان نامه خود از فکر ، ایده و نوشته دیگری بهره گرفته ام با نقل قول مستقیم یا غیر مستقیم منبع و ماخذ آن را نیز در جای مناسب ذکر کرده ام . بدیهی است مسئولیت تمامی مطالبی که نقل قول دیگران نباشد بر عهده خویش می دانم و جوابگوی آن خواهم بود .

دانشجو تائید می نماید که مطالب مندرج در این پایان نامه نتیجه تحقیقات خودش می باشد و در صورت استفاده از نتایج دیگران مرجع آن را ذکر نموده است .

اسماعیل رضاپوریان

تاریخ و امضاء

اسماعیل رضاپوریان

۹۱/۶/۳۰

اینجانب اسماعیل رضاپوریان دانشجوی ورودی سال ۸۹ مقطع کارشناسی ارشد رشته‌ی آمار ریاضی گواهی می نمایم چنانچه بر اساس مطالب پایان نامه خود اقدام به انتشار مقاله ، کتاب ، و ... نمایم ضمن مطلع نمودن استاد راهنما ، با نظر ایشان نسبت به نشر مقاله ، کتاب ، و ... و به صورت مشترک و با ذکر نام استاد راهنما مبادرت نمایم .

اسماعیل رضاپوریان

تاریخ و امضاء

اسماعیل رضاپوریان ۹۱/۶/۳۰

کلیه حقوق مادی مترتب از نتایج مطالعات ، آزمایشات و نوآوری ناشی از تحقیق موضوع این پایان نامه متعلق به دانشگاه پیام نور می باشد .

شهریور ۹۱

چکیده

یک سری مجانبی یکنواخت برای مجموع تعداد زیادی از متغیرهای تصادفی مستقل و هم‌توزیع بدست آورده‌ایم. بر خلاف معمول این سری یکنواخت، دقت خوبی را در سراسر دامنه می‌دهد. برای حالت خاص، متغیر تصادفی،

$$Z_{n,\alpha} = Y_1 + 2^{\alpha}Y_2 + \cdots + n^{\alpha}Y_n،$$

با $\alpha \in R$ و $Y_1, Y_2, \ldots$ متغیرهای تصادفی مستقل با توزیع نمایی و میانگین ۱ را در نظر می‌گیریم. تابع توزیع $Z_{n,\alpha}$ یک سری با علائم متناوب است که باعث مشکلات بزرگ عددی می‌شود. با استفاده از نسخه‌های توسعه یافته از روش نقطه‌ی زینی، یک بسط مجانبی یکنواخت برای $\mathbb{P}\left(Z_{n,\alpha} < x\right)$ می‌یابیم. برای مقادیر مختلف پارامتر توزیع، α، توزیع‌های معروف آماری نمایان می‌گردد.

واژگان کلیدی: بسط مجانبی، تقریب نقطه‌ی زینی، توزیع نمایی.

فهرست مطالب

پیشگفتار

روش نقطه‌ی زینی یک تعمیم از روش لاپلاس[1] است ، برای مقدار تقریبی انتگرال، که در آن، تغییر شکل انتگرال روی منحنی ساده بسته، در صفحه مختلط در نزدیکی نقطه‌ی زینی صورت می‌گیرد. تقریب‌های نقطه‌ی زینی ابزارهای توانایی برای بدست آوردن عبارات دقیق برای تابع توزیع‌ها هستند که برای بدست آوردن تقریب چگالی حاشیه‌ای و چگالی میانگین نمونه‌ای کاربرد دارند.

دیگر کاربرد روش نقطه‌ی زینی در بدست آوردن توزیع‌های مجانبی برای مجموع متغیرهای تصادفی مستقل و هم‌توزیع می‌باشد. دراین پایان‌نامه برای احتمال $\varphi_N(y)$ که در آن y برابر است با

$$y = \nu_1 + \nu_2 + \cdots + \nu_N$$

و $\nu_N \ldots \nu_1$ متغیر تصادفی مستقل و هم‌توزیع باشند یک توزیع مجانبی بدست آورده‌ایم.

در حالت خاص تابع توزیع متغیر تصادفی $Z_{n,\alpha}$، که ترکیبی خطی از متغیرهای تصادفی مستقل با توزیع نمایی و میانگین یک است را در نظر می‌گیریم. تابع توزیع $\mathbb{P}\left(Z_{n,\alpha} < x\right)$ یک سری متناوب نسبتاً ساده است. اما این سری به یک سری عددی مشکل افزایش می‌یابد. در اینجا یک توزیع تقریباً دقیق و یکنواخت برای همه‌ی $x \geq 0$ با استفاده از نسخه توسعه یافته‌ی روش نقطه‌ی زینی بدست می‌آوریم. به همین منظور ابتدا، $\mathbb{P}\left(Z_{n,\alpha} < x\right)$ را مانند یک انتگرال روی منحنی ساده‌ی بسته می‌نویسیم. که با استفاده از وارون تبدیل لاپلاس برای $Z_{n,\alpha}$ است. پس از تبدیل

[1] Laplace

انتگرال از روش نقطه‌ی زینی استفاده می‌کنیم.

به غیر از مورد $\alpha = 0$ ، $Z_{n,\alpha}$ شامل متغیرهای تصادفی با توزیع‌های مختلف است. انگیزه‌ی اولیه برای انجام این پایان‌نامه برای نشان دادن این نکته است که، چگونه روش نقطه‌ی زینی، همراه با برخی از روش‌های مجانبی دیگر منجر به بسط یکنواخت برای $\mathbb{P}\left(Z_{n,\alpha} < x\right)$ می‌شود که حتی برای مقادیر کوچک n و x دور از میانگین دقیق است.

در واقع برای همه $\alpha \geq -\frac{1}{2}$ متغیر تصادفی $Z_{n,\alpha}$ از قضیه‌ی حد مرکزی پیروی می‌کند و به متغیر تصادفی نرمال استاندارد همگرا می‌شود. نتیجه، تقریب نرمال $P\left(Z_{n,\alpha} < x\right)$ برای n های بزرگ و xهای نزدیک به میانگین $Z_{n,\alpha}$ مفید است. اما برای افزایش مقدار α یا برای مقادیری از x که از میانگین دور هستند قضیه‌ی حد مرکزی نتیجه کمتری می‌دهد.بسط (توزیع) مجانبی بدست آمده این اثر را تصحیح می‌کند.

بسط برای $\alpha < -\frac{1}{2}$ معتبر باقی می‌ماند. برای $\alpha = -1$ متغیر تصادفی $Z_{n,\alpha}$ از n مستقل و هم‌توزیع بیشترین شرح را می‌دهد که متغیر تصادفی به صورت نمایی توزیع شده است. این یک مثال از تئوری مقدار بی‌نهایت است، که وقتی $n \to \infty$، $Z_{n,\alpha}$ توزیع گامبل[1] را نتیجه می‌دهد. برای $\alpha = -2$، $Z_{n,\alpha}$ توزیع کلموگروف[2] را نتیجه می‌دهد. دیگر توزیع‌های کلاسیک که در این مطالعه نقش دارند، حرکت براوان[3] و ارتباط با تابع زتای ریمان[4] است. (بیانی[5] ،۲۰۰۱)

---

[1] Gumbel
[2] Kolmogorov
[3] Brownian
[4] Riemann zeta
[5] Biane

در فصل اول به توضیح تقریب نقطه‌ی زینی پرداخته‌ایم و برخی از کاربردهای آن را در آمار بیان کرده‌ایم. در فصل دوم شرح مختصری از توزیع‌های که دراین پایان‌نامه نقش دارند آمده است و یک تقریب برای توزیع نمایی وزنی. که حالت خاصی از توزیع $Z_{n,\alpha}$ است بیان شده است. فصل سوم شامل بسط مجانبی یکنواخت برای مجموع متغیرهای تصادفی مستقل و هم‌توزیع است. در فصل چهارم یک بسط تقریبا دقیق و یکنواخت برای همه‌ی $x \geq 0$ با استفاده از نسخه توسعه یافته روش نقطه‌ی زینی بدست می‌آوریم.

فصل اول : آشنایی با تقریب نقطه‌ی زینی

مقدمه

در ریاضیات روش تندترین کاهش یا روش نقطه‌ی زینی، یک تعمیم از روش لاپلاس است، برای مقدار تقریبی انتگرال، که در آن، تغییر شکل انتگرال روی منحنی ساده بسته، در صفحه مختلط در نزدیکی نقطه‌ی زینی در مسیر شدیدترین نزول صورت می‌گیرد.

۱-۱ روش لاپلاس

تابع مثبت $f(x)$ را در نظر می‌گیریم، با توجه به این که

$$h(x) \equiv ln\, f(x)$$

$$f(x) = e^{h(x)}$$

و با استفاده از بسط تیلور[1] $h(x)$ حول نقطه x_0 داریم

$$f(x) = exp\left\{h(x_0) + (x - x_0)h'(x_0) + \frac{(x - x_0)}{2} h''(x_0) + R\right\}$$

که در آن $R = O((x - x_0)^3)$ است. با انتخاب $x_0 = \hat{x}$ وقتی که $h'(\hat{x}) = 0$ داریم

$$f(x) = exp\left\{h(\hat{x}) + \frac{(x - \hat{x})^2}{2}\ h''(\hat{x})\right\}$$

پس $$\int f(x)dx \approx \int exp\left\{h(\hat{x}) + \frac{(x-\hat{x})^2}{2}\ h''(\hat{x})\right\} dx$$

[1]Taylor

اگر $\hat{x}$ نقطه ماکسیمم باشد $h''(\hat{x})$ منفی است و

$$\int f(x)dx \approx exp\{h(\hat{x})\}\left(-\frac{2\pi}{h''(\hat{x})}\right)^{1/2}$$

واین، تقریب لاپلاس نامیده می‌شود.

اگر تابع f را مانند زیر بنویسیم

$$f(x) = \int m(x,t)dt$$

که این کار برای $m(x,t)$ مثبت، قابل انجام است. برای مثال در نظر می‌گیریم،

$$m(x,t) = f(x)m_0(t)$$

وقتی $m_0(t)$ یک تابع انتگرال‌گیری باشد، بوسیله تعریف $k(x,t) = ln\, m(x,t)$ تقریب لاپلاس را بررسی می‌کنیم. برای $exp\{k(x,t)\}$ با توجه به متغیر دوم t برای هر x ثابت می‌نویسیم

$$\begin{aligned} f(x) \quad &\approx \quad \int \; exp\left\{k\big(x,\hat{t}(x)\big) + \frac{\big(t-\hat{t}(x)\big)^2}{2}\frac{\partial^2 k(x,t)}{\partial t^2}\Big|_{\hat{t}(x)}\right\}dt \\ &= \quad exp\big\{k\big(x,\hat{t}(x)\big)\big\}\left(-\frac{2\pi}{\frac{\partial^2 k(x,t)}{\partial t^2}\Big|_{\hat{t}(x)}}\right)^{1/2} \end{aligned}$$

که برای هر x و $\hat{t}(x)$ در $\frac{\partial k(x,t)}{\partial t} = 0$ صدق می‌کند. آخرین معادله بالا تقریب نقطه‌ی زینی است از $f(x)$ و $\hat{t}(x)$ یک نقطه‌ی زینی است. یک مثال کاربردی از این روش می‌تواند محاسبه چگالی حاشیه‌ای باشد.

اگر $f(x,y)$ چگالی دو متغیره باشد، داریم

$$f_X(x)=\int f(x,y)dy \approx \sqrt{2\pi}\, f(x,\hat{y})\left[\frac{\partial^2 \ln f(x,y)}{\partial y^2}\Big|_{\hat{y}}\right]^{-1/2}$$

هنگامی که $\hat{y}=\hat{y}(x)$ در عبارت $\frac{\partial \ln f(x,y)}{\partial y}=0$ صدق می‌کند.

۱-۲ لم واتسون[1]

فرض می‌کنیم انتگرال $\int_0^\infty e^{-xt} q\,(t)dt$ به xهای بزرگ همگرا باشد. با بسط دادن $q(t)$

به صورت

$$q(t)\sim\sum_{s=0}^{\infty} a_s t^{(s+\lambda-\mu)/\mu}$$

که λ و x ثابت‌های مثبت هستند، و با جانشین کردن در انتگرال $\int_0^\infty e^{-xt} q\,(t)dt$ یک بسط

مجانبی به صورت زیر بدست می‌آوریم.

$$\int_0^\infty e^{-xt} q\,(t)dt \sim \sum_{s=0}^{\infty} \Gamma\left(\frac{s+\lambda}{\mu}\right)\frac{a_s}{x^{(s+\lambda)/\mu}}$$

[1] Watson

۱-۳ تقریب نقطه‌ی زینی کلاسیک

انتگرال زیر را در نظر می‌گیریم

$$I(z)=\int_a^b e^{-zp(t)}q(t)dt$$

با فرض‌اینکه t_0 مستقل از z_0 است، داریم

$$I(z)=\int_{t_0}^b e^{-zp(t)}q(t)dt-\int_{t_0}^a e^{-zp(t)}q(t)dt$$

با به کار بردن نتایج روش لاپلاس برای هر انتگرال در قسمت راست و بسط سری تیلور $p(t)$ و $q(t)$ در $t=t_0$ می‌نویسیم

$$p(t)=p(a)+\sum_{s=0}^{\infty}p_s(t-a)^{s+\mu}$$

$$q(t)=\sum_{s=0}^{\infty}q_s(t-a)^{s+\lambda-1}$$

که t_0 یک صفر ساده از $p'(t)$ است و نقطه‌ی زینی نامیده می‌شود. با استفاده از لم واتسون مقدار $I(z)$ را بدست می‌آوریم

$$\int_a^b e^{-zp(t)}\,q(t)dt \sim 2e^{-zp(t_0)}\sum_{s=0}^{\infty}\Gamma\left(s+\frac{1}{2}\right)\frac{b_2 s}{z^{s+\left(\frac{1}{2}\right)}}$$

که در آن $b_0=\frac{q}{(2p\prime\prime)^{1/2}}$

$$b_2 = \left(2q'' - \frac{2p'''q'}{p''} + \left(\frac{5(p''')^2}{6(p'')^2} - \left(\frac{p''''}{2p''}\right) q\right)\right)\frac{1}{(2p'')^{3/2}}$$

است.

۱-٤ نقطه‌ی زینی در آمار

به یاد داریم که برای چگالی $f(x)$ تابع مولد گشتاور بصورت زیر تعریف شده است.

$$\varphi_X(t) = \int_{-\infty}^{+\infty} exp\,(tx)f(x)dx \qquad (۱-۱)$$

از $\varphi_X(t)$ می‌توانیم $f(x)$ را بوسیله فرمول وارون بدست بیاوریم

$$\begin{aligned} f(x) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \varphi_X(it)exp(-itx)dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} exp\{k_X(it) - itx\}dt \end{aligned} \qquad (۱-۲)$$

و $k_X(t)$ را برابر $ln\,\varphi_X(t)$ تعریف می‌کنیم. این فرمول در زمینه‌های آماری عمومی است. وقتی که $f(x)$ یک چگالی و $\varphi_X(it)$ یک تابع مشخصه و $k_{\mathrm{x}}(t) = ln\,\varphi_X(t)$ یک تابع مولد انباشته از متغیر تصادفی X است. با تغییر متغیر $(t' = it)$ داریم

$$f(x) = \frac{1}{2\pi}\int_{\tau-i\infty}^{\tau+i\infty} exp(k_X(t) - tx)dt \qquad (۱-۳)$$

با در نظر گرفتن $k(x,t) = k_X(t) - tx$ و همچنین پیدا کردن جواب معادله $k'_X(t) = x$ که نقطه‌ی زینی را می‌دهد، توان را در (۱-۳)، حول $\hat{t}(x)$ بسط می‌دهیم. پس رابطه‌ی زیر را داریم

$$k_X(t) - tx \approx k_X\big(\hat{t}(x)\big) - \hat{t}(x)x - \frac{\big(t-\hat{t}(x)\big)^2}{2} k''_X\big(\hat{t}(x)\big)$$

اکنون با جایگذاری در (۱–۳) انتگرال می‌گیریم و رابطه‌ی (۱–٤)بدست می‌آید.

$$f_X(x) \approx \left(\frac{1}{2\pi k''_x\big(\hat{t}(x)\big)}\right)^{1/2} exp\{k_X(\hat{t}(x) - \hat{t}(x)x)\} \qquad (٤–١)$$

یک کاربرد مستقیم ازاین فرمول در بدست آوردن توزیع $\bar{x}$ می‌باشد. اگر در نظر بگیریم که $\bar{x}$ ، میانگین نمونه‌ای از $x_1, x_2, \dots, x_n$ است، و x_iها متغیرهای تصادفی مستقل و هم‌توزیع‌اند و هر x_i دارای یک تابع مولد گشتاور، $\varphi_X(t)$ و تابع مولد انباشته، $k_X(t)$ است. نشان می‌دهیم تابع مولد گشتاور $\bar{x}$ برابر $\varphi_{\bar{X}}(t) = \varphi_X\left(t/n\right)^n$ و تابع مولد انباشته برابر $k_{\bar{X}}(t) = nk_X\left(t/n\right)$ است.

با توجه به فرمول (۱–٤) داریم

$$f_{\bar{X}}(\bar{x}) \approx \left(\frac{n}{2\pi k''_X(\hat{t}(\bar{x}))}\right)^{1/2} exp\{n[K_X\hat{t}(\bar{x}) - \hat{t}(\bar{X})\bar{x}]\}$$

تقریب نقطه‌ی زینی به شدت در تقریب‌های بدست آمده در چگالی برآورد درست نمایی ماکسیمم، به ویژه در خانواده نمایی استفاده می‌شود.

در نظر می‌گیریم $x_1, x_2, \dots, x_n$ متغیرهای تصادفی مستقل با چگالی

$$f(x|\theta) = exp\{\theta s(x) - k(\theta) - d(x)\}$$

که $S=\sum_{i=1}^{n} s(x_i)$ آماره‌های بسنده‌ی کافی است که دارای چگالی (۱-۵) می‌باشد.

$$f(s|\theta) = exp\{\theta_s - nk(\theta) - h(s)\} \tag{۵-۱}$$

با در نظر گرفتن بسندگی، می‌خواهیم به چگالی $f(s|\theta)$ برسیم، که اندکی از روش‌های مختلف بدست می‌آید.

اگر تقریب نقطه‌ی زینی را برای تمام چگالی‌های (۱-۵) بدست بیاوریم، نقطه‌ی زینی تابعی از θ خواهد شد. با حذف بخشی، تقریب نقطه‌ی زینی را تنها برای تابع $exp\{-h(s)\}$ بدست می‌آوریم.

اولین قدم برای پیدا کردن تبدیل لاپلاس و تابع مشخصه‌ی انباشته از $exp\{-h(s)\}$ است زیرا $f(s|\theta)$ یک چگالی است و این انتگرال برابر ۱ می‌شود. از این رو بوسیله انتگرال‌گیری و چینش دوباره (۱-۵) داریم

$$exp\{nk(\theta)\} = \int exp(\theta s) exp\{-h(s)\} ds$$

که طرف راست کاملاً شبیه (۱-۱) است، با θ به جای t و s به جای x. تقریب نقطه‌ی زینی را تنها برای $exp\{-h(s)\}$ استفاده می‌کنیم. چگالی $f(s|\theta)$ تقریبی است بوسیله

$$\begin{aligned} f(s|\theta) \quad &\approx \quad exp\{\theta s - nk(\theta)\} \frac{1}{\left(2\pi nk''^{(\hat{t}(s))}\right)^{1/2}} exp\left(nk(\hat{t}(s)) - \hat{t}(s)s\right) \\ &= \quad \frac{1}{\left[2\pi nk''(\hat{t}(s))\right]^{1/2}} exp\{[\theta - \hat{t}(s)]s - n[k(\theta) - k(\hat{t}(s))]\} \end{aligned}$$