



دانشگاه شاهرود

دانشکده علوم پایه

گروه آمار

پایان نامه

برای دریافت درجه کارشناسی ارشد در رشته

آمار، گرایش آمار ریاضی

عنوان

تقریب توزیع های پسین با استفاده از بسط اچوورث و اتحاد استاین

استاد راهنما

دکتر حسین باغیشنی

استاد مشاور

دکتر نگار اقبال

پژوهشگر

رحیم خیرکو

۱۳۹۳

نام خانوادگی دانشجو: خیرگو

نام: رحیم

عنوان: تقریب توزیع‌های پسین با استفاده از بسط اچوورث و اتحاد استاین

استاد راهنما: دکتر حسین باغیشنی

استاد مشاور: دکتر نگار اقبال

مقطع تحصیلی: کارشناسی ارشد

رشته: آمار

گرایش: آمار ریاضی

دانشگاه: دانشگاه شاهرود

دانشکده علوم پایه

تاریخ فارغ‌التحصیلی: ۱۳۹۳

تعداد صفحات: ۹۶

واژگان کلیدی: بسط اچوورث، اتحاد استاین، چندجمله‌ای‌های هرمیتی، توزیع پسین

چکیده

در بسیاری از کاربردها، محققین علاقه‌مند به تقریب تابع توزیع احتمال می‌باشند. یکی از روش‌هایی که برای این منظور به کار گرفته می‌شود، استفاده از بسط اچوورث است. بسط اچوورث به صورت یک سری نوشته می‌شود که توزیع احتمال را بر حسب گشتاورها یا انباشتک‌هایش، تقریب می‌زند. این بسط را می‌توان ابتدا به وسیله بسط تابع توزیع، بر حسب توابع متعامد هرمیتی، و سپس جمع کردن عبارات توانی بر حسب اندازه نمونه نتیجه گرفت. در این پایان‌نامه، با استفاده از اتحاد استاین و ویژگی چندجمله‌ای‌های هرمیتی، برای توزیع‌های پسین، با استفاده از بسط سری‌های اچوورث، تقریب مناسبی را به دست می‌آوریم. سپس مقایسه‌ای بین این روش و روش‌های قبلی معرفی‌شده انجام می‌دهیم. در پایان، عملکرد تقریب توزیع پسین حاصل از این روش را در دو مثال بررسی می‌کنیم.

تقدیم بہ

پدرم بہ استواری کوه

مادرم بہ زلالی چشمہ

ہمسرم بہ صمیمیت باران

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 بی‌آنکه دوست بداند، روزی کن.

اگر تنها ترین تنها شوم، باز خدا هست

او جانشین همه نداشتن‌هاست...

^۱مناجاتی از دکتر علی شریعتی.

سپاس گزارمی...پ

سپاس خداوندگار حکیم را که با لطف بی‌کران خود، آدمی را زیور عقل آراست. در آغاز وظیفه خود می‌دانم از زحمات بی‌دریغ استاد راهنمای خود، جناب آقای دکتر حسین باغیشنی، صمیمانه تشکر و قدردانی کنم که قطعاً بدون راهنمایی‌های ارزنده ایشان، این مجموعه به انجام نمی‌رسید. از سرکار خانم دکتر نگار اقبال که زحمت مطالعه و مشاوره این رساله را تقبل فرمودند و در آماده‌سازی این رساله، به نحو احسن اینجانب را مورد راهنمایی قرار دادند، کمال امتنان را دارم. 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 از همسرم به پاس عاطفه سرشار و گرمای امیدبخش وجودش، که در این سردترین روزگاران، بهترین پشتیبان من بود. و تمام دوستانم که همراه من بودند.

رحیم خیرکو
۱۳۹۳

پیش‌گفتار

در دیدگاه بیزی استنباط آماری، همه استنباط‌ها مبتنی بر توزیع پسین کمیت‌های مورد علاقه و توزیع‌های پیشگو به‌دست می‌آیند. در اکثر موقعیت‌های کاربردی، صورت توزیع پسین صریح یا بسته نیست و باید تقریبی از آن را به دست آورد. به‌طور کلی، می‌توان توزیع پسین را با دو روش تقریب زد: الف) روش‌های مبتنی بر نمونه‌گیری، مانند الگوریتم‌های MCMC، ب) روش‌های مجانبی، مانند تقریب لاپلاس، که مشابه تقریب حاصل از قضیه حد مرکزی برای توزیع‌های پسین است.

تقریب‌های حاصل از نظریه مجانبی توزیع‌ها، تحت شرایطی که در مواردی برقرار نیستند، دارای خطای ناچیزی هستند. اما در مواردی هم سرعت همگرایی به توزیع تقریبی چندان قابل قبول نیست. در این موارد، استفاده از عبارات با مرتبه‌های بالاتر در بسط‌های مجانبی توزیع‌های پسین می‌تواند تقریب بهتری را نتیجه دهد و سرعت همگرایی را افزایش دهد. البته در کنار دقت بالاتر، هزینه محاسبات نیز افزایش می‌یابد. یکی از این بسط‌های مجانبی، بسط اچوورث می‌باشد. در این پایان‌نامه، با استفاده از بسط اچوورث و اتحاد استاین به ارایه تقریبی برای توزیع پسین می‌پردازیم. این پایان‌نامه در چهار فصل تدوین شده است که به‌طور خلاصه در هر فصل به موارد زیر پرداخته شده است:

- در فصل اول مفاهیم و مقدمات لازم برای ورود به بحث اصلی پایان‌نامه ارایه شده‌اند.
- در فصل دوم، پس از معرفی اتحاد استاین، صورتی از توزیع مجانبی نرمال به کمک این اتحاد برای توزیع پسین، در حالت چندپارامتری، نمایش داده شده است.
- در فصل سوم، به کمک بسط اچوورث، تقریب بهتری، نسبت به تقریب فصل دوم، برای توزیع پسین ارایه شده است. سپس بسط دیگری معروف به بسط جانسون معرفی شده و مقایسه‌ای بین آن و بسط اچوورث صورت گرفته است.
- در فصل چهارم نیز، عملکرد تقریب حاصل از بسط اچوورث در دو مثال مورد ارزیابی قرار گرفته و با تقریب معرفی شده در فصل دوم مقایسه شده است. در پایان، نتیجه‌گیری و آینده تحقیق مطرح شده‌اند.
- دو پیوست نیز شامل معرفی مفاهیم نظریه اندازه و کدهای نرم‌افزار R برای بازتولید مثال‌های پایان‌نامه، در انتهای پایان‌نامه گنجانده شده‌اند.

فهرست مطالب

۱	مقدمات و پیش‌نیازها	۱
۱	۱.۱ ترتیب‌ها و ترتیب‌های تصادفی	۱
۳	۲.۱ انباشتک	۳
۴	۱.۲.۱ برخی از ویژگی‌های تابع مولد انباشتک	۴
۵	۲.۲.۱ برخی از ویژگی‌های انباشتک‌ها	۵
۶	۳.۲.۱ رابطه بین انباشتک‌ها و گشتاورها	۶
۹	۳.۱ چندجمله‌ای‌های هرمیتی	۹
۱۵	۴.۱ بسط کرنیش-فیشر	۱۵
۱۷	۵.۱ سری گرام شارلیر	۱۷
۱۸	۶.۱ سری اچوورث	۱۸
۲۴	۱.۶.۱ معایب بسط اچوورث	۲۴
۲۴	۷.۱ استنباط بیزی	۲۴
۲۷	۲ اتحاد استاین و تقریب نرمال توزیع پسین	۲۷
۲۷	۱.۲ مقدمه	۲۷
۲۹	۲.۲ اتحاد استاین	۲۹
۳۲	۳.۲ معرفی مدل	۳۲
۳۵	۴.۲ شرایط نظم	۳۵
۳۶	۵.۲ توزیع مجانبی پسین	۳۶
۴۱	۳ تقریب توزیع پسین با بسط اچوورث	۴۱
۴۱	۱.۳ مقدمه	۴۱
۴۳	۲.۳ اتحاد استاین برای بسط توزیع پسین	۴۳

۴۴ بسط اچوورث بیزی	۳.۳
۴۹ شرط انتگرال‌پذیری	۱.۳.۳
۵۱ بسط اچوورث توزیع پسین بر اساس چندجمله‌های هریتی و گشتاورها	۲.۳.۳
۵۲ خطای تقریب مبتنی بر بسط اچوورث	۴.۳
۵۶ بسط توزیع پسین جانسون	۵.۳
۵۹ مطالعه دقت بسط اچوورث بیزی	۴
۵۹ مدل دوجمله‌ای	۱.۴
۶۱ مدل نرمال دومتغیره	۲.۴
۶۴ نتیجه‌گیری و پیشنهادات برای آینده تحقیق	۳.۴
۶۷ نمادها و تعاریف	آ
۷۲ کدهای زبان برنامه‌نویسی R برای بازتولید نتایج شبیه‌سازی پایان‌نامه	ب
۹۰ مراجع	

فصل ۱

مقدمات و پیش‌نیازها

در این فصل، مفاهیم و تعاریف مورد نیاز برای ورود به بحث‌های فصل‌های بعدی را بیان می‌کنیم.

۱.۱ ترتیب‌ها و ترتیب‌های تصادفی

نمادهای ترتیب‌های تصادفی O_p و o_p و غیرتصادفی O و o ، رایج‌ترین نمادها برای توصیف مرتبه مجانبی کمیت‌های تصادفی و غیرتصادفی هستند. برای ورود به این بحث، ابتدا لازم است تا تعریفی از همگرایی در احتمال را ارائه دهیم.

تعریف ۱.۱.۱ (گات، ۲۰۰۵) دنباله متغیرهای تصادفی Y_n را در احتمال به Y ، همگرا گوئیم و می‌نویسیم

$$Y_n \xrightarrow{P} Y, \text{ اگر به ازای هر } \epsilon > 0,$$

$$\lim_{n \rightarrow \infty} P[\|Y_n - Y\|_d < \epsilon] = 1,$$

که در آن $\|\cdot\|_d$ معرف نرم اقلیدسی است.

تعریف ۲.۱.۱ (ون در وارت، ۱۹۹۸) دو دنباله از متغیرهای تصادفی $\{X_n\}$ و $\{Y_n\}$ را در نظر بگیرید.

گوئیم $X_n = O_p(Y_n)$ ، اگر برای هر $\epsilon > 0$ ، $K(\epsilon) > 0$ و $n.(\epsilon)$ وجود داشته باشند، به طوری که به ازای

همه مقادیر $n.(\epsilon)$ ، $n > n.(\epsilon)$

$$P\left(\left|\frac{X_n}{Y_n}\right| \leq K(\epsilon)\right) > 1 - \epsilon.$$

جدول ۱.۱: a و b اعداد حقیقی هستند و $k = \max(a, b)$

سطر	ضرب	جمع	ترکیب
۱	$o(n^a)o(n^b) = o(n^{a+b})$	$o(n^a) + o(n^b) = o(n^k)$	$O_p(O(n^a)) = O_p(n^a)$
۲	$O(n^a)O(n^b) = O(n^{a+b})$	$O(n^a) + O(n^b) = O(n^k)$	$O_p(O_p(n^a)) = O_p(n^a)$
۳	$O(n^a)o(n^b) = o(n^{a+b})$	$O(n^a) + o(n^b) = O(n^k)$	$O(O(n^a)) = O(n^a)$
۴	$o_p(n^a)o_p(n^b) = o_p(n^{a+b})$	$o_p(n^a) + o_p(n^b) = o_p(n^k)$	$O(O_p(n^a)) = O_p(n^a)$
۵	$O_p(n^a)O_p(n^b) = O_p(n^{a+b})$	$O_p(n^a) + O_p(n^b) = O_p(n^k)$	$o_p(O_p(n^a)) = o_p(n^a)$
۶	$O_p(n^a)o_p(n^b) = o_p(n^{a+b})$	$O_p(n^a) + o_p(n^b) = O_p(n^k)$	$o(O_p(n^a)) = o_p(n^a)$
۷	$o_p(n^a)O(n^b) = o_p(n^{a+b})$	$O_p(n^a) + O(n^b) = O_p(n^k)$	$O_p(o_p(n^a)) = O_p(n^a)$
۸	$O_p(n^a)o(n^b) = o_p(n^{a+b})$	$O(n^a) + o_p(n^b) = O_p(n^k)$.
۹	$O(n^a)o_p(n^b) = o_p(n^{a+b})$.	.
۱۰	$O(n^a)O_p(n^b) = O_p(n^{a+b})$.	.

تعریف ۳.۱.۱. (ون دروات، ۱۹۹۸) یک دنباله از متغیرهای تصادفی $\{X_n\}$ و دنباله‌ای از متغیرهای

تصادفی $\{R_n\}$ را در نظر بگیرید. گوئیم $X_n = o_p(R_n)$ ، اگر وقتی $n \rightarrow \infty$ ، آن‌گاه

$$\frac{X_n}{R_n} \xrightarrow{P} 0.$$

تعریف ۴.۱.۱. (ون دروات، ۱۹۹۸) دو دنباله از اعداد ثابت $\{a_n\}$ و $\{b_n\}$ را در نظر بگیرید. گوئیم

$$b_n = o(a_n) \text{ هرگاه}$$

$$\frac{b_n}{a_n} \xrightarrow{P} 0.$$

تعریف ۵.۱.۱. (ون دروات، ۱۹۹۸) دو دنباله از اعداد ثابت $\{a_n\}$ و $\{b_n\}$ را در نظر بگیرید. گوئیم

$b_n = O(a_n)$ هرگاه $K > 0$ و عدد صحیح مثبت N وجود داشته باشند، به طوری که برای $n \geq N$ ،

$$\left| \frac{b_n}{a_n} \right| < K.$$

با استفاده از تعاریف بالا، برای هر ثابت حقیقی c ، ترتیب‌های $o_p(a_n)$ ، $O_p(a_n)$ ، $o(a_n)$ و $O(a_n)$ به

ترتیب معادل با $ca_n O_p(1)$ ، $ca_n o_p(1)$ ، $ca_n O(1)$ و $ca_n o(1)$ هستند. همچنین $X_n = O_p(n^c)$ ، برقراری

$X_n = o_p(n^{c+\epsilon})$ را به ازای هر $\epsilon > 0$ ، نتیجه می‌دهد، اما $X_n = O_p(n^{c+\epsilon})$ لزوماً به معنی برقراری

$X_n = O_p(n^c)$ نیست. برخی از ویژگی‌های ترتیب‌های تصادفی و غیرتصادفی در جدول ۱.۱ خلاصه

شده‌اند (اسحاقی، ۱۳۹۲).

۲.۱ انباشتک

در نظریه احتمال و آمار، انباشتک‌های^۱ یک توزیع احتمال، κ_n ، مجموعه‌ای از کمیت‌هایی هستند که جایگزینی برای گشتاورهای توزیع فراهم می‌کنند (بریلینگر، ۱۹۶۹). در برخی موارد، رفتارهای نظری مسایل احتمالی با عبارات انباشتک‌ها ساده‌تر از گشتاورهای توزیع می‌باشد. همان‌طور که در مورد گشتاورها، برای گردایه‌ای از متغیرهای تصادفی، گشتاورهای توأم تعریف می‌شوند، انباشتک‌های توأم را نیز می‌توان تعریف کرد. برای یک متغیر تصادفی X ، تابع مولد انباشتک در نقطه t با $g(t)$ نشان داده و به صورت زیر تعریف می‌شود:

$$g(t) = \log E[e^{tX}].$$

که با استفاده از بسط مک لورن^۲ داریم

$$g(t) = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}.$$

در این صورت n امین انباشتک X به صورت زیر محاسبه می‌شود:

$$\kappa_n = \left(\frac{\partial^n}{\partial t^n} g(t) \Big|_{t=0} \right).$$

جدول ۲.۱ انباشتک‌های اول و دوم برخی از توزیع‌های آماری را خلاصه کرده است.

جدول ۲.۱: انباشتک‌های اول و دوم چند توزیع آماری

انباشتک دوم κ_2	انباشتک اول κ_1	توزیع
$p(1-p)$	p	برنولی: $B(1, p)$
μ	μ	پواسن: $P(\mu)$
$\kappa_1 p^{-1}$	$p^{-1} - 1$	هندسی: $Ge(p)$
$\kappa_1(1-p)$	np	دوجمله‌ای: $B(n, p)$
σ^2	μ	نرمال: $N(\mu, \sigma^2)$
λ^{-2}	λ^{-1}	نمایی: $E(\lambda)$

^۱Cumulants

^۲Maclaurin series

۱.۲.۱ برخی از ویژگی‌های تابع مولد انباشتک

در این زیر بخش برخی از ویژگی‌های مهم تابع مولد انباشتک را فهرست می‌کنیم. جزییات بیشتر این ویژگی‌ها را می‌توانید در هالد (۲۰۰۰) ببینید.

- تابع مولد انباشتک $g(t)$ محدب است، از مبدأ عبور می‌کند و در صورت وجود بی‌نهایت بار مشتق‌پذیر است.
- مشتق اول به‌طور یکنوا در بازه‌ای باز بین اینفیمم و سوپریمم تکیه‌گاه توزیع احتمال قرار می‌گیرد. مشتق دوم آن به جز برای توزیع‌های تباهیده^۳ به‌طور اکید مثبت می‌باشد. این ویژگی شاید با این واقعیت که مقدار آن در $t = 0$ ، واریانس توزیع را نتیجه می‌دهد، بهتر درک شود.
- تابع مولد انباشتک وجود دارد اگر و تنها اگر دم‌های توزیع، به‌وسیله یک نرخ کاهشی نمایی^۴ (فروپاشی نمایی)، ماژور شده^۵ (بیشانده) باشند. به بیان دیگر دم‌های توزیع بالاتر از دم‌های توزیع نمایی قرار گرفته و آن را محدود کند. یعنی

$$\exists C > 0, \quad F(x) = O(e^{Cx}), \quad x \rightarrow -\infty;$$

$$\exists d > 0, \quad 1 - F(x) = O(e^{-dx}), \quad x \rightarrow \infty,$$

که در آن $F(\cdot)$ تابع توزیع می‌باشد.

- در صورت وجود اینفیمم c و سوپریمم d ، تابع مولد انباشتک دارای مجانب قائم در آن نقاط می‌باشد. در غیر این صورت برای همه اعداد حقیقی تعریف شده است.
- برای یک انتقال k در توزیع، داریم

$$g_{X+k}(t) = g_X(t) + kt.$$

^۳Degenerate Distribution

^۴Exponential decay

^۵Majorized

- یک ویژگی جالب که تابع مولد انباشتک دارد، آن است که تابع مولد انباشتک مجموع دو متغیر تصادفی مستقل، برابر مجموع دو تابع مولد انباشتک متناظر آن دو متغیر است. یعنی

$$\begin{aligned} g_{X+Y}(t) &= \log E[e^{t(X+Y)}] \\ &= \log(E[e^{tX}]E[e^{tY}]) \\ &= \log E[e^{tX}] + \log E[e^{tY}] \\ &= g_X(t) + g_Y(t). \end{aligned}$$

- برای یک توزیع تباهیده در k ، تابع مولد انباشتک، خط $g_k(t) = kt$ می‌باشد.
- اگر $f(x|\theta)$ تابع (چگالی) احتمال یک خانواده نمایی کانونی^۶ و توابع $h(t)$ و $g(t) = \log M(t)$ به ترتیب توابع مولد احتمال و انباشتک این خانواده باشند، آن‌گاه (کسلا و برگر، ۲۰۰۱)

$$f(x|\theta) = \frac{1}{M(\theta)} e^{\theta x} h(x),$$

و

$$g(t|\theta) = g(t + \theta) - g(\theta).$$

۲.۲.۱ برخی از ویژگی‌های انباشتک‌ها

به منظور درک بهتر عملکرد انباشتک‌ها، برخی از ویژگی‌های مهم آن‌ها را در این زیر بخش ذکر می‌کنیم. جزئیات این ویژگی‌ها را می‌توانید در بریلینگر (۱۹۶۹) ببینید. این ویژگی‌ها به راحتی از تعریف مستقیم تابع مولد انباشتک و مشتق‌گیری از آن قابل نتیجه‌گیری هستند.

هم‌وردایی و ناوردایی

با انتقال یک متغیر تصادفی (با اضافه کردن C)، اولین انباشتک به اندازه C انتقال می‌یابد. یعنی تحت انتقال هم‌وردایی^۷ می‌باشد. اما این انتقال روی بقیه انباشتک‌ها تاثیری ندارد. یعنی بقیه انباشتک‌ها ناوردایی^۸ هستند. به

^۶ Canonical exponential family

^۷ Equivariance

^۸ Invariance

عبارت دیگر، اگر $\kappa_n(X)$ ، n امین انباشتک توزیع احتمال متغیر تصادفی X باشد، آن‌گاه برای هر ثابت C

$$\kappa_1(X + C) = \kappa_1(X) + C,$$

$$\kappa_n(X + C) = \kappa_n(X), \quad n \geq 2.$$

همگنی

n امین انباشتک، همگن از درجه n می‌باشد. به این معنی که اگر C مقداری ثابت باشد، آن‌گاه

$$\kappa_n(CX) = C^n \kappa_n(X).$$

جمع‌پذیری

اگر X و Y متغیرهای تصادفی مستقل باشند، آن‌گاه

$$\kappa_n(X + Y) = \kappa_n(X) + \kappa_n(Y).$$

۳.۲.۱ رابطه بین انباشتک‌ها و گشتاورها

با نوشتن تابع مولد گشتاور، بهتر می‌توان رابطه بین گشتاورها و انباشتک‌های یک توزیع را درک کرد. تابع

مولد گشتاور یک توزیع عبارت است از

$$M(t) = 1 + \sum_{n=1}^{\infty} \frac{\mu'_n t^n}{n!} = \exp\left(\sum_{n=1}^{\infty} \frac{\kappa_n t^n}{n!}\right) = \exp(g(t)),$$

که در آن μ'_n گشتاور غیر مرکزی مرتبه n ام است. به عبارتی تابع مولد انباشتک، لگاریتم تابع مولد گشتاور

می‌باشد. انباشتک اول، مقدار امید ریاضی توزیع است. انباشتک‌های دوم و سوم به ترتیب گشتاورهای مرکزی

دوم و سوم هستند. اما انباشتک‌های بالاتر، نه گشتاور و نه گشتاور مرکزی هستند، بلکه توابع چندجمله‌ای

پیچیده‌تری از گشتاورها می‌باشند. وابستگی انباشتک‌ها به گشتاورها را می‌توان به وسیله رابطه بازگشتی زیر

مشاهده کرد:

$$\kappa_n = \mu'_n - \sum_{m=1}^{n-1} C_{m-1}^{n-1} \kappa_m \mu'_{n-m},$$

که در آن $C_x^y = \frac{y!}{x!(y-x)!}$. گشتاور غیرمرکزی n ام، یک چندجمله‌ای درجه n بر حسب n انباشتک اول می‌باشد. این واقعیت را برای شش گشتاور اول در زیر مشاهده می‌کنید:

$$\mu'_1 = \kappa_1,$$

$$\mu'_2 = \kappa_2 + \kappa_1^2,$$

$$\mu'_3 = \kappa_3 + 3\kappa_2\kappa_1 + \kappa_1^3,$$

$$\mu'_4 = \kappa_4 + 4\kappa_3\kappa_1 + 3\kappa_2^2 + 6\kappa_2\kappa_1^2 + \kappa_1^4,$$

$$\mu'_5 = \kappa_5 + 5\kappa_4\kappa_1 + 10\kappa_3\kappa_2 + 10\kappa_3\kappa_1^2 + 15\kappa_2^2\kappa_1 + 10\kappa_2\kappa_1^3 + \kappa_1^5,$$

$$\begin{aligned} \mu'_6 = & \kappa_6 + 6\kappa_5\kappa_1 + 15\kappa_4\kappa_2 + 15\kappa_4\kappa_1^2 + 10\kappa_3^2 + 60\kappa_3\kappa_2\kappa_1 + 20\kappa_3\kappa_1^3 + 15\kappa_2^3 \\ & + 45\kappa_2^2\kappa_1^2 + 15\kappa_2\kappa_1^4 + \kappa_1^6. \end{aligned}$$

نماد μ_n برای گشتاور مرکزی مرتبه n استفاده می‌شود. گشتاورهای مرکزی را نیز می‌توان به صورت توابعی از انباشتک‌ها بیان کرد. به عنوان مثال

$$\mu_1 = 0,$$

$$\mu_2 = \kappa_2,$$

$$\mu_3 = \kappa_3,$$

$$\mu_4 = \kappa_4 + 3\kappa_2^2,$$

$$\mu_5 = \kappa_5 + 10\kappa_3\kappa_2,$$

$$\mu_6 = \kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^3.$$

برعکس انباشتک n نیز به صورت یک چندجمله‌ای درجه n بر حسب n گشتاور غیرمرکزی اول قابل تعریف است. به عنوان مثال

$$\kappa_1 = \mu'_1,$$

$$\kappa_2 = \mu'_2 - \mu_1'^2,$$

$$\kappa_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2\mu_1'^3,$$

$$\kappa_4 = \mu'_4 - 4\mu'_3\mu'_1 - 3\mu_2'^2 + 12\mu'_2\mu_1'^2 - 6\mu_1'^4,$$

$$\kappa_5 = \mu'_5 - 5\mu'_4\mu'_1 - 10\mu'_3\mu'_2 + 20\mu'_3\mu_1'^2 + 30\mu_2'^2\mu'_1 - 60\mu'_2\mu_1'^3 + 24\mu_1'^5,$$

$$\begin{aligned} \kappa_6 = & \mu'_6 - 6\mu'_5\mu'_1 - 15\mu'_4\mu'_2 + 30\mu'_4\mu_1'^2 - 10\mu_3'^2 + 120\mu'_3\mu_2'\mu'_1 \\ & - 120\mu'_3\mu_1'^3 + 30\mu_2'^3 - 270\mu_2'^2\mu_1'^2 + 360\mu'_2\mu_1'^4 - 120\mu_1'^6. \end{aligned}$$

برای نمایش انباشتک‌های κ_n ، برای $n > 1$ ، به صورت توابعی از گشتاورهای مرکزی، از چندجمله‌ای‌های بر حسب گشتاورهای غیرمرکزی همه عباراتی را که μ'_1 در آن‌ها ظاهر می‌شود را حذف می‌کنیم و بقیه را به گشتاور مرکزی تبدیل می‌کنیم. به عنوان نمونه، با این دست‌ورالعمل، شش انباشتک اول در بالا بر حسب گشتاورهای مرکزی به صورت زیر بازنویسی می‌شوند:

$$\kappa_2 = \mu_2,$$

$$\kappa_3 = \mu_3,$$

$$\kappa_4 = \mu_4 - 3\mu_2^2,$$

$$\kappa_5 = \mu_5 - 10\mu_3\mu_2,$$

$$\kappa_6 = \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3.$$

برای بیان انباشتک‌های κ_n برای $n > 2$ به صورت توابعی از گشتاورهای مرکزی استاندارد شده، نیز مشابه قبل عمل کرده و در چندجمله‌ای‌ها قرار می‌دهیم $\mu_2 = 1$. در نتیجه مثلاً داریم

$$\kappa_2 = 1,$$

$$\kappa_3 = \mu_3,$$

$$\kappa_4 = \mu_4 - 3,$$

$$\kappa_5 = \mu_5 - 10\mu_3,$$

$$\kappa_6 = \mu_6 - 15\mu_4 - 10\mu_3^2 + 30.$$

۳.۱ چندجمله‌ای‌های هرمیتی

در این قسمت به مرور چندجمله‌ای‌های هرمیتی^۹ می‌پردازیم. این چندجمله‌ای‌ها نقش اساسی در نمایش بسط اچورث^{۱۰} ایفا می‌کنند. چندجمله‌ای هرمیتی مرتبه k ، $q_k(\cdot)$ ، به صورت زیر تعریف می‌شود:

$$q_k(z)\phi(z) = (-1)^k \frac{d^k}{dz^k} \phi(z), \quad (1.1)$$

که در آن $\phi(\cdot)$ تابع چگالی نرمال استاندارد است. به عنوان مثال، به ازای $k = 0, 1, \dots, 5$ داریم

$$q_0(z) = 1,$$

$$q_1(z) = z,$$

$$q_2(z) = z^2 - 1,$$

$$q_3(z) = z^3 - 3z,$$

$$q_4(z) = z^4 - 6z^2 + 3,$$

$$q_5(z) = z^5 - 10z^3 + 15z.$$

^۹Hermite Polynomials

^{۱۰}Edgeworth series

این چندجمله‌ای‌ها یک دنباله از توابع متعامد را تشکیل می‌دهند. نمایش این چندجمله‌ای‌ها در (۱.۱) یک نمایش احتمالی می‌باشد، در حالی که نمایش فیزیکی آن‌ها به صورت زیر است:

$$q_k^{phy}(z)e^{-z^2} = (-1)^k \frac{d^k}{dz^k} e^{-z^2}.$$

به سادگی قابل درک است که این دو نمایش فقط در مقیاس تفاوت دارند، یعنی $q_k^{phy}(z) = 2^{\frac{k}{2}} q_k(\sqrt{2}z)$. کورانت و هیلبرت (۱۹۵۳) چندین ویژگی از $q_k^{phy}(\cdot)$ را ارائه دادند. در حقیقت چندجمله‌ای‌های هرمیتی جواب‌های محرک هارمونیک ساده ماشین‌های کوانتوم هستند (بواس، ۲۰۰۶ و ویرو ارفکن، ۲۰۰۴) و بخشی لازم در ریاضی فیزیک می‌باشند. در زیر سه ویژگی این چندجمله‌ای‌ها را که در فصل‌های بعدی مورد استفاده قرار می‌گیرند، بیان می‌کنیم.

لم ۱.۳.۱. فرض کنید $q'_k(z)$ مشتق چندجمله‌ای نسبت به z باشد. بنابراین

$$q'_k(z) = kq_{k-1}(z), \quad (2.1)$$

$$q_{k+1}(z) = zq_k(z) - kq_{k-1}(z), \quad (3.1)$$

$$\int_{\mathbb{R}} q_k(z)q_j(z)d\Phi(z) = \begin{cases} 0 & k \neq j \\ k! & k = j \end{cases} \quad (4.1)$$

که در آن $\Phi(\cdot)$ معرف تابع توزیع نرمال استاندارد است.

برهان. برای اثبات این سه ویژگی تابع مولد زیر را تعریف می‌کنیم:

$$\begin{aligned} \psi(z, t) &= e^{-\frac{t}{2} + tz} \\ &= e^{-\frac{t}{2} + tz + \frac{z^2}{2} - \frac{z^2}{2}} \\ &= e^{\frac{z^2}{2} - \frac{(t-z)^2}{2}}. \end{aligned}$$

با استفاده از بسط سری مک‌لورن می‌توان نوشت

$$\begin{aligned}\psi(z, t) &= \sum_{n=0}^{\infty} \left(\frac{\partial^n \psi(z, t)}{\partial t^n} \right) \Big|_{t=0} \times \frac{t^n}{n!} \\ &= \sum_{n=0}^{\infty} (-1)^n e^{\frac{z}{\nu}} \frac{d^n e^{-\frac{z}{\nu}}}{dz^n} \times \frac{t^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{q_n(z)}{n!} t^n,\end{aligned}\tag{۵.۱}$$

که خط آخر از (۱.۱) نتیجه می‌شود. از طرفی

$$\frac{\partial \psi(z, t)}{\partial z} = [z + (t - z)] e^{\frac{z}{\nu} - \frac{(t-z)}{\nu}} = t\psi(z, t).\tag{۶.۱}$$

بنابراین با استفاده از (۵.۱) و (۶.۱)

$$\begin{aligned}\sum_{n=0}^{\infty} \frac{q'_n(z)}{n!} t^n &= \sum_{n=0}^{\infty} \frac{q_n(z)}{n!} t^{n+1} \\ &= \sum_{n=1}^{\infty} \frac{q_{n-1}(z)}{(n-1)!} t^n \\ &= \sum_{n=0}^{\infty} \frac{n q_{n-1}(z)}{n!} t^n.\end{aligned}$$

در نتیجه

$$q'_k(z) = k q_{k-1}(z).$$

برای اثبات ویژگی دوم می‌توان نوشت

$$\frac{\partial \psi(z, t)}{\partial t} = -(t - z) e^{\frac{z}{\nu} - \frac{(t-z)}{\nu}} = -(t - z)\psi(z, t).\tag{۷.۱}$$

با استفاده از (۵.۱) و (۷.۱) داریم

$$\begin{aligned} \frac{\partial \psi(z, t)}{\partial t} + (t - z)\psi(z, t) &= \sum_{n=1}^{\infty} \frac{nq_n(z)}{n!} t^{n-1} + t \sum_{n=0}^{\infty} \frac{q_n(z)}{n!} t^n - z \sum_{n=0}^{\infty} \frac{q_n(z)}{n!} t^n \\ &= \sum_{n=1}^{\infty} \frac{q_n(z)}{(n-1)!} t^{n-1} + \sum_{n=0}^{\infty} \frac{q_n(z)}{n!} t^{n+1} - \sum_{n=0}^{\infty} \frac{zq_n(z)}{n!} t^n \\ &= \sum_{n=0}^{\infty} \frac{q_{n+1}(z)}{n!} t^n + \sum_{n=0}^{\infty} \frac{(n+1)q_n(z)}{(n+1)!} t^{n+1} - \sum_{n=0}^{\infty} \frac{zq_n(z)}{n!} t^n \\ &= \sum_{n=0}^{\infty} [q_{n+1}(z) + nq_{n-1}(z) - zq_n(z)] \frac{t^n}{n!} \\ &= 0. \end{aligned}$$

در نتیجه داریم

$$q_{n+1}(z) + nq_{n-1}(z) - zq_n(z) = 0.$$

در لم ۱.۳.۱ ویژگی متعامد (۴.۱) نیز با استدلال زیر قابل اثبات است:

$$\begin{aligned} \int_{-\infty}^{\infty} q_m(z)q_n(z)e^{-\frac{z^2}{\nu}} dz &= (-1)^n \int_{-\infty}^{\infty} q_m(z) \frac{d^n e^{-\frac{z^2}{\nu}}}{dz^n} dz \\ &= (-1)^n \left[q_m(z) \frac{d^{n-1} e^{-\frac{z^2}{\nu}}}{dz^{n-1}} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{\infty} q'_m(z) \frac{d^{n-1} e^{-\frac{z^2}{\nu}}}{dz^{n-1}} dz \right] \\ &= (-1)^{n-1} \left[-q_m(z) \frac{d^{n-1} e^{-\frac{z^2}{\nu}}}{dz^{n-1}} \Big|_{-\infty}^{+\infty} + m \int_{-\infty}^{\infty} q_{m-1}(z) \frac{d^{n-1} e^{-\frac{z^2}{\nu}}}{dz^{n-1}} dz \right] \\ &= \dots \\ &= (-1)^{n-m} m! \int_{-\infty}^{\infty} q_0(z) \frac{d^{n-m} e^{-\frac{z^2}{\nu}}}{dz^{n-m}} dz = 0. \end{aligned}$$

آخرین تساوی، برای $n > m$ با تکرار انتگرال‌گیری جزء به جزء و این امر که $e^{-\frac{z^2}{\nu}}$ و تمام مشتقاتش در

□

$z = \pm\infty$ به صفر میل می‌کنند، نتیجه می‌شود.

لم ۲.۳.۱. فرض کنید Z یک متغیر تصادفی نرمال استاندارد باشد، آن‌گاه

۱.

$$\sum_{i=1}^n C_i^n q_{i-1}(x) E(Z^{n-i}) = \sum_{i=0}^n C_i^n q_i(x) E(Z^{n+1-i}). \quad (۸.۱)$$

۲.

$$\sum_{i=0}^n C_i^{n+1} q_i(x) E(Z^{n+1-i}) = \sum_{i=0}^{n-1} C_i^n q_{i+1}(x) E(Z^{n-i}) + \sum_{i=0}^n C_i^n q_i(x) E(Z^{n+1-i}).$$

برهان. ۱: در توزیع نرمال استاندارد اگر r عددی زوج باشد، داریم $E(Z^r) = (r-1)(r-3)\dots(3)(1)$ و برای r فرد $E(Z^r) = 0$. حال اگر n زوج باشد، $n/2$ عبارت غیرصفر در هر طرف رابطه (۸.۱) وجود دارد. فرض کنید $j = 2$ ، $m = 2j$ ، آن‌گاه زامین عبارت غیرصفر در دو طرف چپ و راست به ترتیب به صورت زیر می‌باشند:

$$C_m^n m q_{m-1} E(Z^{n-m}) = \frac{n(n-1)\dots(n-m+1)}{m!} m q_{m-1} \times (n-m-1)(n-m-3)\dots(3)(1)$$

و

$$C_{m-1}^n q_{m-1} E(Z^{n-m+2}) = \frac{n(n-1)\dots(n-m+2)}{(m-1)!} q_{m-1} (n-m+1)(n-m-1)(n-m-3)\dots(3)(1),$$

که هر دو عبارت مساوی می‌باشند. برهان برای n فرد مشابه است.

۲: می‌دانیم $C_i^{m+1} = C_i^m + C_{i-1}^m$. در نتیجه داریم

$$\begin{aligned} & \sum_{i=0}^n C_i^{m+1} q_i(x) E(Z^{n+1-i}) - \sum_{i=0}^n C_i^m q_i(x) E(Z^{n+1-i}) \\ &= \sum_{i=1}^n C_{i-1}^m q_i(x) E(Z^{n+1-i}) \\ &= \sum_{i=0}^{n-1} C_i^m q_{i+1}(x) E(Z^{n-i}). \end{aligned}$$

□

گزاره ۳.۳.۱. فرض کنید Z متغیر تصادفی نرمال استاندارد باشد. برای $k = 1, 2, \dots$ داریم

$$q_k(x) = x^k - \sum_{i=0}^{k-1} C_i^k q_i(x) E(Z^{k-i}).$$

برهان. اثبات به روش استقرا انجام می‌شود. بدیهی است که برای $k = 1, 2$ رابطه برقرار است. فرض

می‌کنیم که برای $k = n - 1, n$ رابطه برقرار باشد. حال با استفاده از (۳.۱) در لم ۱.۳.۱ داریم

$$q_n(x) = x^n - \sum_{i=0}^{n-1} C_i^n q_i(x) E(Z^{n-i}) = xq_{n-1}(x) - (n-1)q_{n-2}(x).$$

با مشتق‌گیری از دو طرف رابطه فوق نسبت به x داریم

$$nx^{n-1} - \sum_{i=0}^{n-1} C_i^n q_i'(x) E(Z^{n-i}) = q_{n-1}(x) + xq_{n-1}'(x) - (n-1)q_{n-2}'(x).$$

حال با استفاده از (۲.۱) و (۳.۱) در لم ۱.۳.۱ می‌توان نوشت

$$\begin{aligned} nx^{n-1} - \sum_{i=0}^{n-1} C_i^n i q_{i-1}(x) E(Z^{n-i}) &= q_{n-1}(x) + xq_{n-1}'(x) - (n-1)q_{n-2}'(x) \\ &= q_{n-1}(x) + x(n-1)q_{n-2}(x) - (n-1)(n-2)q_{n-3}(x) \\ &= q_{n-1}(x) + (n-1)[xq_{n-2}(x) - (n-2)q_{n-3}(x)] \\ &= q_{n-1}(x) + (n-1)q_{n-1}(x) \\ &= nq_{n-1}. \end{aligned}$$

در نتیجه

$$\begin{aligned} nx^{n-1} &= \sum_{i=1}^{n-1} C_i^n i q_{i-1}(x) E(Z^{n-i}) + nq_{n-1}(x) \\ &= \sum_{i=1}^{n-1} C_i^n i q_{i-1}(x) E(Z^{n-i}) + C_n^n q_n'(x) E(Z^{n-n}) \\ &= \sum_{i=1}^n C_i^n i q_{i-1}(x) E(Z^{n-i}). \end{aligned} \tag{۹.۱}$$

حال با استفاده از (۳.۱) در لم ۱.۳.۱ و فرض استقرا داریم

$$\begin{aligned} q_{n+1}(x) &= xq_n(x) - nq_{n-1}(x) \\ &= x^{n+1} - \sum_{i=0}^{n-1} C_i^n xq_i(x)E(Z^{n-i}) - [nx^{n-1} - n \sum_{i=0}^{n-2} C_i^{n-1} q_i(x)E(Z^{n-1-i})]. \quad (۱۰.۱) \end{aligned}$$

از طرفی با استفاده از (۳.۱) در لم ۱.۳.۱ می‌توان نوشت

$$\begin{aligned} &\sum_{i=0}^{n-1} C_i^n xq_i(x)E(Z^{n-i}) - n \sum_{i=0}^{n-2} C_i^{n-1} q_i(x)E(Z^{n-1-i}) \\ &= xq_n(x) + \sum_{i=1}^{n-1} C_i^n xq_i(x)E(Z^{n-i}) - n \sum_{i=1}^{n-1} C_{i-1}^{n-1} q_{i-1}(x)E(Z^{n-i}) \\ &= xq_n(x) + \sum_{i=1}^{n-1} C_i^n (xq_i(x) - iq_{i-1}(x))E(Z^{n-i}) \\ &= \sum_{i=0}^{n-1} C_i^n q_{i+1}(x)E(Z^{n-i}) \\ &= \sum_{i=0}^n C_i^n q_i(x)E(Z^{n+1-i}). \quad (۱۱.۱) \end{aligned}$$

آخرین تساوی از لم ۲.۳.۱ قسمت ۱ نتیجه می‌شود. با استفاده از (۹.۱) و (۱۱.۱) می‌توان (۱۰.۱) را به صورت زیر بازنویسی کرد:

$$\begin{aligned} q_{n+1}(x) &= x^{n+1} - \sum_{i=0}^n C_i^n q_i(x)E(Z^{n+1-i}) - \sum_{i=0}^n C_i^n q_{i+1}(x)E(Z^{n-i}) \\ &= x^{n+1} - \sum_{i=0}^{n-1} C_i^{n+1} q_i(x)E(Z^{n+1-i}). \end{aligned}$$

که تساوی آخر از لم ۲.۳.۱ قسمت ۲ نتیجه شده است. بنابراین حکم استقرا برقرار و برهان کامل می‌شود.

□

۴.۱ بسط کرنیش-فیش

مشخص‌سازی توزیع‌های آماری با استفاده از گشتاورها و انباشتک‌های آن‌ها، و به کمک معرفی بسط‌های مجانبی حاصل از آن‌ها، از حدود ۱۹۳۷ توسط محققین مختلفی شروع شد (فیش و کرنیش، ۱۹۶۰). کرنیش

و فیشر (۱۹۳۸) یکی از این بسط‌ها را که به بسط کرنیش-فیشر^{۱۱} معروف شد، ارائه دادند. بسط کرنیش-فیشر، بسطی مجانبی است که برای تقریب چندک‌های توزیع یک متغیر تصادفی، بر اساس انباشتک‌هایش و چندک‌های توزیع نرمال استاندارد، به کار گرفته می‌شود.

لم ۱.۴.۱. (داس‌گوپتا، ۲۰۰۸) بسط کرنیش-فیشر برای چندک توزیع: برای یک متغیر تصادفی Z با میانگین μ و واریانس متناهی، چندک مرتبه p از Z ، Q_p ، به صورت $Q_p = \mu + \sigma w$ تعریف می‌شود که در آن

$$\begin{aligned} w = & x + [\gamma_1 h_1(x)] \\ & + [\sigma^2 h_2(x) + \gamma_1^2 h_{11}(x)] \\ & + [\gamma_3 h_3(x) + \gamma_1 \gamma_2 h_{12}(x) + \gamma_1^3 h_{111}(x)] \\ & + \dots, \end{aligned}$$

که در آن

$$x = \Phi^{-1}(p).$$

و

$$\gamma_{r-2} = \frac{\kappa_r}{\kappa_2^{r/2}}, \quad r \in \{3, 4, \dots\},$$

$$h_1(x) = \frac{q_2(x)}{6},$$

$$h_2(x) = \frac{q_3(x)}{24},$$

$$h_3(x) = \frac{q_4(x)}{120},$$

$$h_{11}(x) = -\frac{[2q_3(x) + q_1(x)]}{36},$$

$$h_{12}(x) = -\frac{[q_4(x) + q_2(x)]}{24},$$

$$h_{111}(x) = \frac{[12q_4(x) + 19q_2(x)]}{324},$$

^{۱۱}Cornish-Fisher expansion

به طوری که $q_n(x)$ چند جمله‌ای هریمیتی احتمالی n ام می‌باشد. مقادیر γ_1 و γ_2 نیز چولگی و کشیدگی متغیر تصادفی می‌باشند.

مثال ۲.۴.۱. فرض کنید X یک متغیر تصادفی با میانگین ۱۰، واریانس ۲۵، چولگی ۵ و کشیدگی ۲ باشد. می‌توان از چند جمله اول بسط کرنیش-فیشر که تنها به چولگی و کشیدگی وابسته‌اند، برای برآورد چندک‌های این متغیر تصادفی، به کمک چندک‌های توزیع نرمال استاندارد، استفاده کرد. صدک ۹۵ ام توزیع نرمال استاندارد ۱/۶۵ می‌باشد. بنابراین وزن w را می‌توان به صورت زیر محاسبه کرد:

$$w \approx 1/65 + 5 \frac{1/65^2 - 1}{6} + 5 \frac{1/65^3 - 3 \times 1/65}{24} - 5^2 \frac{2 \times 1/65^3 - 5 \times 1/65}{36} = 2/55621$$

بنابراین صدک ۹۵ برآورد شده X ، $10 + 5(2/55621)$ یا تقریباً ۲۲/۷۸۱ می‌باشد.

۵.۱ سری گرام شارلیر

سری‌های گرام شارلیر^{۱۲} که به افتخار جورجین پدرسون گرام و کارل شارلیر نام‌گذاری شده‌اند، سری‌هایی هستند که توزیع احتمال را به صورت عباراتی برحسب انباشتک‌ها تقریب می‌کنند. در این زمینه می‌توان به استوارت و اورد (۱۹۹۴) و کرامر (۱۹۹۹) مراجعه کرد.

هدف کلیدی این بسط‌ها، تقریب تابع مشخصه توزیعی با تابع چگالی احتمال $f(\cdot)$ است که به صورت عباراتی برحسب تابع مشخصه توزیعی دیگر (که معمولاً توزیع نرمال می‌باشد)، با ویژگی‌های مناسب و معلوم، نوشته شده است و $f(\cdot)$ به وسیله تبدیل فوریه معکوس به دست می‌آید.

متغیر تصادفی پیوسته X را در نظر بگیرید و فرض کنید $\Psi(\cdot)$ تابع مشخصه آن با تابع (چگالی) احتمال $f(\cdot)$ و انباشتک‌های K_{γ_r} باشد. بسط را بر حسب عبارات توزیعی معلوم با تابع چگالی احتمال $g(\cdot)$ ، تابع مشخصه $\psi(\cdot)$ و انباشتک‌های γ_r انجام می‌دهیم. معمولاً $g(\cdot)$ تابع چگالی نرمال در نظر گرفته می‌شود. اما

^{۱۲}Gram-Charlier series

انتخاب‌های دیگر نیز ممکن است. با توجه به تعریف انباشتک‌ها، اتحاد زیر را داریم

$$\Psi(t) = \exp\left[\sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(it)^r}{r!}\right] \psi(t).$$

به‌وسیله ویژگی‌های تبدیل فوریه، $it^r \psi(t)$ تبدیل فوریه $(-1)^r D^r g(x)$ می‌باشد، که در آن D عملگر دیفرانسیل نسبت به x می‌باشد. بنابراین

$$f(x) = \exp\left[\sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(-D)^r}{r!}\right] g(x).$$

اگر $g(x)$ را تابع چگالی نرمال با میانگین و واریانس مفروض برای f ، یعنی میانگین $\mu = \kappa_1$ و واریانس $\sigma^2 = \kappa_2$ در نظر بگیریم. آن‌گاه

$$f(x) = \exp\left[\sum_{r=3}^{\infty} \kappa_r \frac{(-D)^r}{r!}\right] \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right].$$

حال به کمک بسط تابع نمایی و متحد کردن عبارات متناظر با مشتقات هم‌مرتب، به سری گرام شارلیر دست می‌یابیم. اگر تنها دو عبارت اول را در نظر بگیریم، خواهیم داشت

$$f(x) \approx \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \left[1 + \frac{\kappa_3}{3!\sigma^3} q_3\left(\frac{x-\mu}{\sigma}\right) + \frac{\kappa_4}{4!\sigma^4} q_4\left(\frac{x-\mu}{\sigma}\right)\right], \quad (12.1)$$

که در آن $q_3(x) = x^3 - 3x$ و $q_4(x) = x^4 - 6x^2 + 3$ چندجمله‌ای‌های هرمیتی می‌باشند.

تضمینی وجود ندارد که عبارت (۱۲.۱) مثبت و در نتیجه یک توزیع احتمال معتبر باشد. همچنین، سری‌های گرام شارلیر در بسیاری از موارد واگرا هستند. این سری تنها در صورتی همگراست که برای مقادیر خیلی بزرگ x ، $f(x)$ سریع‌تر از $\exp(-\frac{x^4}{4})$ کاهش یابد و هرگاه همگرا نباشد، سری نتیجه‌شده یک بسط مجانبی معتبر نیست و امکان برآورد خطای آن میسر نخواهد بود. به همین دلیل سری‌های اچورث به سری گرام شارلیر ترجیح داده می‌شوند.

۶.۱ سری اچورث

اچورث (۱۸۴۵-۱۹۲۶) بسطی مشابه با سری گرام شارلیر ارائه داد. این بسط تعمیمی از قضیه حد مرکزی می‌باشد. والاک (۱۹۵۸) پیشرفت‌ها و کاربردهای مختلف سری اچورث را تا زمان خود برشمرد. حال

(۱۹۹۲) به معرفی و تشریح کاربردهای بسط اچوورث در روش‌های خودگردان‌سازی^{۱۳} پرداخت. بسط اچوورث در حوزه‌های دیگر نیز به کار گرفته شده است. به عنوان نمونه، موسنر (۱۹۹۸) بسط‌های گرام شارلیر، گوس-هرمیت^{۱۴} و اچوورث را در مسایل فیزیک نجوم مورد استفاده قرار داد و فیلهو و روسنفلد (۲۰۰۴) مساله آزمون قیمت‌گذاری با بسط اچوورث را مورد بررسی قرار دادند. مزیت سری‌های اچوورث در کنترل‌پذیر بودن خطای تقریب حاصل از آن می‌باشد. بنابراین سری، یک بسط مجانبی معتبر می‌باشد. فرض کنید $\{X_i\}$ دنباله‌ای از متغیرهای تصادفی مستقل و هم‌توزیع با میانگین μ و واریانس متناهی σ^2 باشد و Y_n به صورت زیر تعریف شود:

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - \mu}{\sigma}.$$

همچنین فرض کنید $G_n(\cdot)$ تابع توزیع تجمعی متغیر تصادفی Y_n باشد. آن‌گاه با توجه به قضیه حد مرکزی، برای هر x داریم

$$\lim_{n \rightarrow \infty} G_n(x) = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.$$

حال فرض کنید متغیرهای تصادفی X_i دارای انباشتک‌های مرتبه بالاتر $\kappa_r = \sigma^r \lambda_r$ باشند. اگر قرار دهیم

$$\Psi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right),$$

آن‌گاه اختلاف انباشتک در عبارات تابع مشخصه $g_n(t)$ از G_n به صورت زیر می‌باشند:

$$\kappa_1^{G_n} - \gamma_1 = 0,$$

$$\kappa_2^{G_n} - \gamma_2 = 0,$$

$$\kappa_r^{G_n} - \gamma_r = \frac{\kappa_r}{\sigma^r n^{r/2-1}} = \frac{\lambda_r}{n^{r/2-1}}, \quad r \geq 3.$$

سری‌های اچوورث مشابه سری‌های گرام شارلیر توسعه داده شده‌اند، با این تفاوت که در این سری عبارات

^{۱۳} Bootstrap

^{۱۴} Gauss-Hermit

بر حسب توان n متحد شده‌اند. بنابراین

$$g_n(t) = \left[1 + \sum_{j=1}^{\infty} \frac{P_j(it)}{n^{j/2}} \right] \exp\left(-\frac{t^2}{2}\right),$$

به طوری که $P_j(x)$ یک چندجمله‌ای از درجه $1 - 3j$ است. دوباره بعد از اعمال تبدیل فوریه معکوس، تابع

توزیع $G_n(\cdot)$ به صورت

$$G_n(x) = \Phi(x) + \sum_{j=1}^{\infty} \frac{P_j(-D)}{n^{j/2}} \Phi(x), \quad (13.1)$$

به دست می‌آید. پنج عبارت اول بسط به صورت زیر می‌باشند:

$$\begin{aligned} G_n(x) = & \Phi(x) \\ & - \frac{1}{n^{1/2}} \left(\frac{1}{6} \lambda_3 \Phi^{(3)}(x) \right) \\ & + \frac{1}{n} \left(\frac{1}{24} \lambda_4 \Phi^{(4)}(x) + \frac{1}{72} \lambda_3^2 \Phi^{(3)}(x) \right) \\ & - \frac{1}{n^{3/2}} \left(\frac{1}{120} \lambda_5 \Phi^{(5)}(x) + \frac{1}{144} \lambda_3 \lambda_4 \Phi^{(4)}(x) + \frac{1}{1296} \lambda_3^3 \Phi^{(3)}(x) \right) \\ & + \frac{1}{n^2} \left(\frac{1}{720} \lambda_6 \Phi^{(6)}(x) + \left(\frac{1}{1152} \lambda_4^2 + \frac{1}{720} \lambda_3 \lambda_5 \right) \Phi^{(5)}(x) \right) \\ & + \frac{1}{1728} \lambda_3 \lambda_4 \Phi^{(4)}(x) + \frac{1}{3210.4} \lambda_3^4 \Phi^{(3)}(x) \\ & + O(n^{-5/2}), \end{aligned}$$

به طوری که $\Phi^{(j)}(x)$ مشتق j ام $\Phi(\cdot)$ در نقطه x می‌باشد. بلینیکو و موسنر (۱۹۹۸)، الگوریتم ساده‌ای برای

محاسبه عبارات با مرتبه‌های بالاتر ارائه دادند.

قضیه حد مرکزی برای میانگین‌ها، به دلیل متقارن بودن توزیع نرمال، نمی‌تواند چولگی موجود در توزیع واقعی را وارد تقریب کند. با استفاده از بسط اچوورث می‌توان چولگی را دخالت داد. جمله دوم بسط (غیر از تقریب توزیع نرمال) نیز چولگی را وارد می‌کند. اگرچه بسط برای هر تعداد جمله، در صورت وجود گشتاورهای کافی، قابل دست‌یابی است، اما معمولاً بسط با دو جمله اول، بعد از تابع توزیع نرمال، پرکاربردتر است. بسط‌های با سه جمله یا بیشتر، به دلیل حضور چندجمله‌ای‌های هرمیتی با مرتبه‌های بالا،

می‌توانند ناپایدار باشند (هال، ۱۹۹۲). به دلیل کاربرد وسیع بسط اچوورث با دو جمله اول، آن را خاص کرده و در قضیه زیر ارایه می‌دهیم. برای اثبات آن می‌توانید به هال (۱۹۹۲) مراجعه کنید. قبل از بیان قضیه، شرط کرامر را که برای سنجش خطای تقریب لازم است، بیان می‌کنیم.

شرط کرامر: تابع توزیع $F(\cdot)$ در شرط کرامر صدق می‌کند، هرگاه

$$\limsup_{t \rightarrow \infty} \chi |E_F(e^{itX})| < 1.$$

که در آن χ تکیه‌گاه X است.

تمام توزیع‌های به‌طور مطلق پیوسته در شرط کرامر صدق می‌کنند (لاهی، ۲۰۰۳). برای توزیع‌هایی که در این شرط صدق نمی‌کنند، نمی‌توان از بسط بیان‌شده در قضیه زیر برای آن‌ها استفاده کرد. البته برای این‌گونه توزیع‌ها بررسی‌هایی برای بهره‌برداری از بسط اچوورث وجود دارند که محاسبات بیشتری را به همراه دارند (اسین، ۱۹۴۵).

قضیه ۱.۶.۱. اگر F در شرط کرامر صدق کند و $E_F(X^4) < \infty$ ، آن‌گاه

$$F_n(x) = \Phi(x) + \frac{C_1(F)P_1(x)\phi(x)}{\sqrt{n}} + \frac{C_2(F)P_2(x) + C_3(F)P_3(x)}{n}\phi(x) + O(n^{-3/2}),$$

به‌طوری‌که

$$C_1(F) = \frac{E(X - \mu)^3}{6\sigma^3},$$

$$C_2(F) = \frac{\frac{E(X - \mu)^4}{\sigma^4} - 3}{24},$$

$$C_3(F) = \frac{C_1^2(F)}{72},$$

$$P_1(x) = (1 - x^2),$$

$$P_2(x) = 3x - x^3,$$

$$P_3(x) = 10x^3 - 15x - x^5.$$

علی‌رغم وجود این قضیه برای توزیع‌هایی که در شرط کرامر صدق می‌کنند، همواره سوال مهمی که برای

هر تقریبی مطرح می‌شود این است که تقریب حاصل تا چه حد دقیق است. برای بسط اچورث، به‌طور کلی و به صراحت نمی‌توان به این سوال پاسخ داد. اما به نظر می‌رسد اگر توزیع مورد نظر چوله یا n کوچک باشد، آن‌گاه بسط اچورث با یک یا حتی دو جمله اول نمی‌تواند دقیق باشد.

برای نمایش عملکرد تقریب حاصل از بسط اچورث، در ادامه دو مثال ساده را تشریح می‌کنیم.

مثال ۲.۶.۱. فرض کنید X_1, X_2, X_3 یک نمونه تصادفی ساده از توزیع کی‌دو با ۲ درجه آزادی باشد. می‌خواهیم توزیع میانگین این نمونه ۳ تایی، $\bar{X} = 1/3 \sum_{i=1}^3 X_i$ ، را به‌دست آوریم. توزیع دقیق \bar{X} گاما است. یعنی

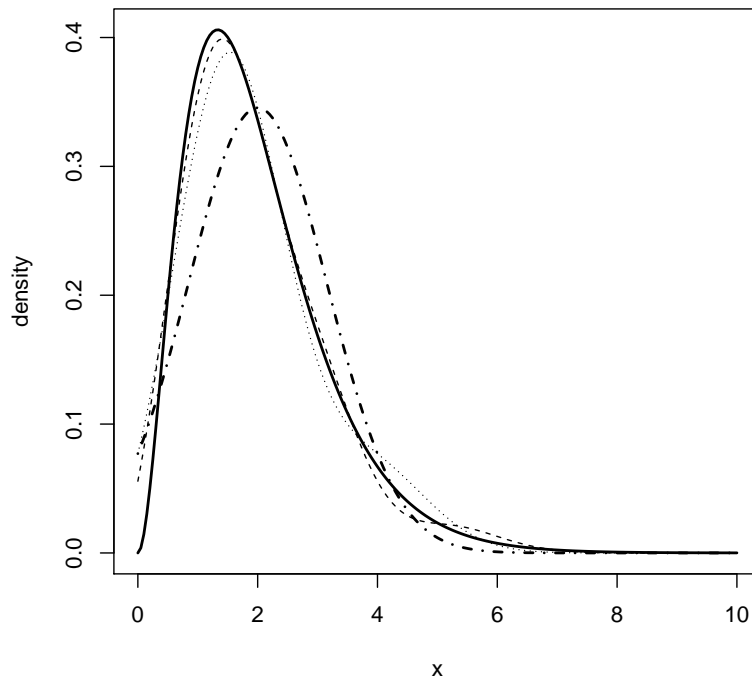
$$\bar{X} \sim \text{Gamma}(\alpha = nk/2, \theta = 2/n) = \text{Gamma}(\alpha = 3, \theta = 2/3).$$

برای ارزیابی عملکرد تقریب حاصل از بسط اچورث، تقریب‌های حاصل از قضیه حد مرکزی (توزیع مجانبی نرمال)، $\bar{X} \sim N(k, 2k/n) = N(2, 4/3)$ ، و دو بسط اچورث درجه دوم و سوم را با توزیع دقیق مقایسه می‌کنیم. شکل ۲.۶.۱ منحنی‌های تابع چگالی واقعی و سه تقریب حاصل از قضیه حد مرکزی و دو بسط اچورث را نشان می‌دهد. مشاهده می‌کنید که دقت تقریب با بسط اچورث مرتبه دوم و سوم خیلی بالاست و با منحنی واقعی تفاوت چندانی ندارند؛ در حالی که، تقریب توزیع نرمال با خطای کوچکی همراه نیست و با منحنی واقعی فاصله دارد. کد برنامه R برای باز تولید این مثال در پیوست ب آمده است.

مثال ۳.۶.۱. (داس‌گوپتا، ۲۰۰۸) فرض کنید X_1, X_2, \dots, X_5 یک نمونه تصادفی ۵ تایی از توزیع نمایی با میانگین ۱ باشد. جدول ۳.۱ دقت بسط اچورث با یک جمله، دو جمله و قضیه حد مرکزی را در تقریب توزیع میانگین این نمونه نشان می‌دهد. همان‌طور که ملاحظه می‌کنید تقریب‌های حاصل از بسط اچورث یک جمله‌ای و دو جمله‌ای بهتر از تقریب قضیه حد مرکزی عمل می‌کنند. توجه کنید که نادقیق بودن تقریب حد مرکزی برای $x = 0$ ، به دلیل ناتوانی آن در وارد کردن چولگی می‌باشد.

کاربردهای مختلفی از بسط سری اچورث را می‌توانید در بیکل (۱۹۷۴)، باتاچاریا و گوش (۱۹۷۸)،

بارندورف-نیلسن و کاکس (۱۹۹۱)، بنتکاس و همکاران (۱۹۹۷) و آلبرز (۲۰۰۱) بیابید.



شکل ۱.۱: تقریب تابع چگالی احتمال میانگین ۳ متغیر کی دو با دو درجه آزادی؛ خط ممتد: چگالی واقعی، نقطه خط: تقریب نرمال، نقطه چین: بسط اچوورث درجه دو، خط تیره: بسط اچوورث درجه سه

جدول ۳.۱: مقایسه دقت بسط اچوورث و تقریب حد مرکزی در تقریب توزیع میانگین

x	مقدار دقیق توزیع	تقریب حد مرکزی	بسط اچوورث یک جمله‌ای	بسط اچوورث دو جمله‌ای
۰	۰/۵۶	۰/۵۰	۰/۵۵۹	۰/۵۵۹
۰/۴	۰/۷۰۱	۰/۶۵۵	۰/۷۰۲	۰/۷۰۰
۱	۰/۸۴۷	۰/۸۴۱	۰/۸۴۱	۰/۸۴۹
۲	۰/۹۵۹	۰/۹۷۷	۰/۹۵۳	۰/۹۵۹

۱.۶.۱ معایب بسط اچوورث

در کنار ویژگی‌های خوب و کاربردهای متعدد بسط اچوورث، این بسط دارای معایبی نیز می‌باشد که در زیر به آن‌ها اشاره می‌کنیم:

- ضمانتی وجود ندارد که تقریب حاصل از بسط اچوورث یک تابع توزیع احتمال باشد به این معنی که

۱. انتگرال تابع چگالی حاصل از آن لزوماً برابر یک نمی‌شود.

۲. احتمال‌های محاسبه‌شده بر اساس آن ممکن است منفی باشند.

- احتمال‌های حاصل از آن، به‌ویژه احتمال‌های دمی، می‌توانند با خطای قابل ملاحظه‌ای همراه باشند.

دو دلیل اصلی برای این امر عبارتند از:

۱. این بسط تحت یک سری تیلور حول میانگین به‌دست می‌آید.

۲. این بسط به‌طور مجانبی یک خطای مطلق را متحمل می‌شود نه یک خطای نسبی، و این امر

زمانی که بخواهیم کمیت‌های خیلی کوچک را تقریب کنیم، مساله‌ساز می‌شود. چون ممکن

است خطای مطلق کوچک و خطای نسبی قابل ملاحظه باشد.

۷.۱ استنباط بیزی

در دیدگاه کلاسیک آمار، پارامتر θ را یک مقدار ثابت نامعلوم در نظر می‌گیرند و یک نمونه تصادفی

X_1, X_2, \dots, X_n از جمعیتی که دارای تابع چگالی احتمال $f(x|\theta)$ است، جمع‌آوری کرده و بر اساس

آن در مورد θ تصمیم‌گیری می‌کنند. در دیدگاه بیزی، پارامتر θ را کمیتی در نظر می‌گیرند که خود یک متغیر تصادفی است و تغییرات آن توسط یک توزیع احتمالی (که به آن توزیع پیشین^{۱۵} گفته می‌شود) بیان می‌گردد (رابرتز، ۲۰۰۸). این توزیع پیشین بر اساس اعتقادات و تجربیات قبلی آزمایشگر و قبل از مشاهده داده‌ها تعیین می‌گردد. سپس از جمعیت یک نمونه گردآوری می‌شود و بر اساس آن توزیع پیشین به‌روز می‌گردد. توزیع پیشین به‌روز شده را توزیع پسین^{۱۶} می‌نامند و بر اساس این توزیع استنباط‌های آماری استخراج می‌شوند. برای مثال فرض کنید طول عمر قطعات تولیدی یک کارخانه دارای توزیع نمایی با میانگین θ ساعت باشند. میانگین طول عمر دستگاه‌ها می‌تواند به علت فرسودگی آن‌ها در طی سال تغییر کند و این تغییرات از یک توزیع احتمالی، توزیع پیشین، پیروی کند. در این حالت استفاده از روش‌های استنباط بیزی می‌تواند مناسب باشد.

برای درک شیوه به‌روز رسانی اعتقاد پیشین، یعنی محاسبه توزیع پسین، فرض کنید θ دارای توزیع پیشین $\pi(\theta)$ باشد و $X = (X_1, \dots, X_n)$ یک نمونه تصادفی گردآوری شده از جمعیتی باشد که توزیع آن به θ بستگی دارد و تابع (چگالی) احتمال آن $f(x|\theta)$ است. تابع (چگالی) احتمال توزیع پسین به صورت زیر محاسبه می‌شود:

$$\pi(\theta|x) = \frac{f(x, \theta)}{m(x)},$$

که در آن $f(x, \theta)$ تابع (چگالی) احتمال توأم X و θ و $m(x)$ تابع (چگالی) احتمال حاشیه‌ای X است. از طرفی می‌توان نوشت

$$f(x, \theta) = \pi(\theta)f(x|\theta), \quad m(x) = \int f(x, \theta)d\Pi(\theta),$$

که در آن $\Pi(\theta)$ تابع توزیع پیشین است.

بنابراین

$$\pi(\theta|x) = \frac{\pi(\theta)f(x, \theta)}{m(x)} = c(x)\pi(\theta)f(x|\theta) \propto \pi(\theta)f(x|\theta).$$

^{۱۵}Prior distribution

^{۱۶}Posterior distribution

یعنی برای محاسبه توزیع احتمال پسین θ کافی است حاصل ضرب $\pi(\theta)$ در $f(x|\theta)$ (تابع درست‌نمایی) را به‌دست آورده و آن را به یک توزیع احتمال تبدیل کنیم. برای اطلاعات بیشتر در این زمینه به رابرتز (۲۰۰۸) مراجعه کنید.

فصل ۲

اتحاد استاین و تقریب نرمال توزیع پسین

در این فصل، به کمک اتحادی معروف به اتحاد استاین، یک تقریب نرمال برای تابع توزیع پسین، در حالت چندپارامتری، ارائه می‌دهیم. مطالب مربوط به این فصل، تشریحی از ونگ و تی‌سی (۲۰۰۸) می‌باشند.

۱.۲ مقدمه

تحقیق و مطالعه بر روی نرمال بودن توزیع مجانبی پسین به زمان لاپلاس برمی‌گردد (ونگ و تی‌سی، ۲۰۰۸) و توجه بسیاری از نویسندگان را جلب کرده است. روش معمول برای نمایش نرمال بودن توزیع مجانبی پسین، بر مبنای بسط سری تیلور تابع لگاریتم درستنمایی حول برآوردگر درستنمایی ماکسیمم (MLE) عمل می‌کند و بسطی را می‌سازد که اولین مولفه آن توزیع نرمال استاندارد است، به طوری که، به شرط معلوم بودن داده‌ها، در احتمال یا قریب به یقین^۱ (a.s.) برقرار است.

نرمال مجانبی بودن توزیع‌های پسین توسط بسیاری از آماردان‌ها، تحت شرایط متفاوت، نشان داده شده است. به عنوان نمونه، در چارچوب متغیرهای تصادفی مستقل و هم‌توزیع می‌توان به له کم (۱۹۵۳)، واکر (۱۹۶۹)، دیوید (۱۹۷۰)، لیندلی (۱۹۸۰) و جانسون (۱۹۷۰) اشاره کرد و در چارچوب فرآیندهای تصادفی از هاید و جانستون (۱۹۷۹)، باساوا و رآوو (۱۹۸۰)، چن (۱۹۸۵) و سویتینگ و آدکولا (۱۹۸۷) یاد کرد.

^۱Almost surely

در تحقیقاتی که اسم بردیم و تحقیقات مشابه، معمولاً سه شرط اساسی وجود دارند. این سه شرط به رشد اطلاع، پیوستگی اطلاع و رفتار دنباله‌ای تابع درستنمایی مربوط می‌شوند. به‌طور خاص، رشد اطلاع فرض می‌کند که نرم ماتریس اطلاع مشاهده‌شده به بی‌نهایت میل می‌کند؛ پیوستگی اطلاع شامل این فرض است که تابع اطلاع در یک همسایگی از برآورد درستنمایی ماکسیمم (یا مقدار واقعی پارامتر)، هموار است؛ و رفتارهای دنباله‌ای اساساً پاسخی به این سوال هستند که مقدار لگاریتم درستنمایی خارج از همسایگی مشخص‌شده، با چه سرعتی به صفر میل می‌کند؟ محققانی (چن، ۱۹۸۵؛ سویتینگ، ۱۹۹۲) نشان داده‌اند بررسی دو شرط اول نسبتاً آسان است ولی بررسی سوم که به رفتار غیرموضعی تابع درستنمایی مربوط می‌شود، پیچیده‌تر و مشکل است.

ونگ (۲۰۰۳)، در حالت تک‌پارامتری، یک روش جایگزین برای اثبات نرمال مجانبی بودن توزیع پسین فرآیندهای تصادفی، بر اساس اتحاد استاین^۲ (وودروف، ۱۹۸۹) پیشنهاد کرد. سپس ونگ و تی‌سی (۲۰۰۸) آن را به حالت چندپارامتری تعمیم دادند. در این روش از یک تبدیل مناسب مانند Z استفاده می‌شود. سپس برای یک تابع اندازه‌پذیر کراندار مانند $h(\cdot)$ ، صورتی از اتحاد استاین برای جداسازی مولفه‌های باقی‌مانده امید ریاضی پسین $h(Z)$ به کار می‌رود، به‌طوری که نرمال مجانبی بودن این توزیع پسین به‌سادگی قابل نتیجه باشد.

در این فصل، با استفاده از اتحاد استاین و مطابق با ونگ و تی‌سی (۲۰۰۸) همگرایی توزیع پسین به نرمال را نشان می‌دهیم. در ادامه، ابتدا تعاریف و نمادهای لازم را معرفی می‌کنیم. سپس شرایط نظم مورد نیاز را بیان می‌کنیم. و در پایان همگرایی توزیع مجانبی پسین به نرمال را نشان می‌دهیم.

^۲Stein's identity

۲.۲ اتحاد استاین

روش استاین، برای تقریب نرمال اولین بار توسط استاین (۱۹۷۲) بیان شد، که با انگیزه معرفی معیاری برای مشخص سازی توزیع نرمال انجام گرفت. به عبارتی $Z \sim N(0, \sigma^2)$ اگر و تنها اگر

$$E[Zf(Z)] = \sigma^2 E\left[\frac{d}{dz}f(Z)\right],$$

به طوری که $f(\cdot)$ هر تابع پیوسته مطلق است و $f'(\cdot)$ بیانگر مشتق تابع $f(\cdot)$ است. چن (۱۹۷۵) از اتحاد استاین برای مشخص سازی توزیع پواسن استفاده کرد. در تحقیق دیگری، استاین (۱۹۸۱) از تقریب نرمال حاصل از این اتحاد برای یافتن برآوردگر می نیکس میانگین یک توزیع نرمال چندمتغیره با بعد ۳ یا بیشتر بهره برد. برای مشاهده تعمیمها و کاربردهای مختلف این اتحاد به چن و همکاران (۲۰۱۱) مراجعه کنید. برای بیان اتحاد استاین، ابتدا باید برخی نمادها را معرفی کنیم.

فرض کنید $\Phi_p(\cdot)$ تابع توزیع نرمال استاندارد p متغیره و $\phi_p(\cdot)$ تابع چگالی آن را نشان دهد. همچنین به اختصار Φ_1 و ϕ_1 را با Φ و ϕ نشان می دهیم. برای توابع $h(\cdot)$ که انتگرال آنها متناهی است، قرار می دهیم

$$\Phi_p h = \int h d\Phi_p.$$

همچنین فرض کنید Γ یک اندازه علامت دار متناهی^۳ (گات، ۲۰۰۵) به صورت

$$d\Gamma = f d\Phi_p,$$

باشد که $f(\cdot)$ یک تابع حقیقی مقدار روی \mathbb{R}^p است، به طوری که $\int |f| d\Phi_p < \infty$. به طور مشابه تعریف می کنیم $\Gamma h = \int h d\Gamma$. برای $s > 0$ مجموعه همه توابع اندازه پذیر $h : \mathbb{R}^p \rightarrow \mathbb{R}$ را که برای $b > 0$ در شرط $h(z) < b(1 + \|z\|^s)$ صدق می کنند، با H_s نشان می دهیم. برای $h \in H_s$ تعریف می کنیم

$$\langle h_p, h \rangle = \Phi_p h$$

$$h_k(y_1, \dots, y_k) = \int_{\mathbb{R}^{p-k}} h(y_1, \dots, y_k, w) \Phi_{p-k}(dw), \quad (1.2)$$

^۳Finite signed measure

و

$$g_k(y_1, \dots, y_p) = e^{\frac{1}{\nu} y_k^\nu} \int_{y_k}^{\infty} [h_k(y_1, \dots, y_{k-1}, w) - h_{k-1}(y_1, \dots, y_{k-1})] e^{-\frac{1}{\nu} w^\nu} dw, \quad (2.2)$$

به طوری که $-\infty < y_1, \dots, y_p < \infty$ و $k = 1, \dots, p$. سپس تعریف می‌کنیم $Uh = (g_1, \dots, g_p)^T$.

دقت داشته باشید که در (2.2)، $g_k(y_1, \dots, y_p) = g_k(y_1, \dots, y_k)$ ، که شامل y_{k+1}, \dots, y_p نیست.

مثال 1.2.2. برای $z \in \mathbb{R}^3$ اگر قرار دهیم $h(z) = z_1 z_2$ ، نشان می‌دهیم $Uh(z) = (0, z_1, 0)^T$. با توجه

به (1.2) داریم

$$\begin{aligned} h. &= \int z_1 z_2 d\Phi_3(z) = 0, \\ h_1(z_1) &= \int h(z_1, w_1, w_2) d\Phi_2(w_1, w_2) = \int z_1 w_1 d\Phi_2(w_1, w_2) = 0, \\ h_2(z_1, z_2) &= \int h(z_1, z_2, w) d\Phi_1(w) = \int z_1 z_2 d\Phi(w) = z_1 z_2, \\ h_3(z_1, z_2, z_3) &= h(z_1, z_2, z_3) = z_1 z_2. \end{aligned}$$

در نتیجه بنا بر (2.2) می‌توان نوشت

$$\begin{aligned} g_1(z) &= e^{\frac{z_1^\nu}{\nu}} \int_{z_1}^{\infty} [h_1(w) - h.] e^{-\frac{w^\nu}{\nu}} dw = 0, \\ g_2(z) &= e^{\frac{z_2^\nu}{\nu}} \int_{z_2}^{\infty} [h_2(z_1, w) - h_1(z_1)] e^{-\frac{w^\nu}{\nu}} dw = e^{\frac{z_2^\nu}{\nu}} \int_{z_2}^{\infty} z_1 w e^{-\frac{w^\nu}{\nu}} dw = z_1, \\ g_3(z) &= e^{\frac{z_3^\nu}{\nu}} \int_{z_3}^{\infty} [h_3(z_1, z_2, w) - h_2(z_1, z_2)] e^{-\frac{w^\nu}{\nu}} dw = 0. \end{aligned}$$

بنابراین

$$Uh(z) = (g_1, g_2, g_3)^T = (0, z_1, 0)^T.$$

به طور مشابه مثال بالا، می‌توان نشان داد برای $z \in \mathbb{R}^p$ اگر $h(z) = z_1$ آن‌گاه $Uh(z) = (1, 0, \dots, 0)^T$

و اگر $h(z) = \|z\|^2$ آن‌گاه $Uh(z) = z$

لم ۲.۲.۲. (اتحاد استاین) فرض کنید r عدد صحیح نامنفی باشد. با فرض آن که $d\Gamma = f d\Phi_p$ ، که $f(\cdot)$ تابعی مشتق‌پذیر روی \mathbb{R}^p است و

$$\int_{\mathbb{R}^p} |f(z)| \Phi_p(dz) + \int_{\mathbb{R}^p} (1 + \|z\|^r) \|\nabla f(z)\|_d \Phi_p(dz) < \infty, \quad (۳.۲)$$

آن‌گاه برای همه $h \in H_r$

$$\Gamma h = \Gamma \mathbf{1} \cdot \Phi_p h + \int_{\mathbb{R}^p} (U h(z))^T \nabla f(z) \Phi_p(dz). \quad (۴.۲)$$

برهان. اثبات این لم، همان اثبات گزاره ۱ وودروف (۱۹۸۹) می‌باشد. با این حال، اثبات را برای حالتی که $p = ۱$ است، بیان می‌کنیم؛ زیرا در فصل سوم از این اثبات استفاده خواهیم کرد.

فرض کنید عمل مشتق‌گیری را نشان دهد. با توجه به پذیره‌های لم ۲.۲.۲ داریم $f(x) = \int_{-\infty}^x f'(y) dy$. همچنین داریم

$$\begin{aligned} \Gamma h - \Gamma \mathbf{1} \cdot \Phi h &= \Phi(fh) - \Phi f \cdot \Phi h \\ &= \int_{\mathbb{R}} \left\{ \int_{-\infty}^x f'(y) dy \right\} \phi(x) [h(x) - \Phi h] dx \\ &= \int_{\mathbb{R}} \left\{ \int_y^{\infty} [h(x) - \Phi h] \phi(x) dx \right\} f'(y) dy \\ &= \int_{\mathbb{R}} U h(y) f'(y) \phi(y) dy \end{aligned}$$

□ تعویض ترتیب انتگرال‌ها به ضمانت شرایط انتگرال‌پذیری تعریف‌شده در لم قابل انجام است.

لم زیر در بخش‌های بعد مورد استفاده قرار می‌گیرد که اثبات آن را می‌توانید در وودروف و کواد (۱۹۹۷) یا ونگ و وودروف (۲۰۰۰) ببینید.

لم ۳.۲.۲. (وودروف و کواد، ۱۹۹۷) اگر $h(z) \in H$ ، آن‌گاه $U h \in H$. همچنین اگر $h(z) = \|z\|^p$ ، $p \geq ۱$ آن‌گاه

$$\|U h(z)\| \leq C (1 + \|z\|^{p-1}).$$

که در آن $0 < C < \infty$.

یادآوری می‌کنیم که $h \in H$ ، به معنی آن است که تابع $h(\cdot)$ کران‌دار است.

۳.۲ معرفی مدل

فرض کنید X_t یک بردار تصادفی باشد که دارای توزیعی با تابع (چگالی) احتمال $p_t(x_t|\theta)$ است، به طوری که t یک پارامتر گسسته یا پیوسته (که معمولاً بیانگر حجم نمونه است) و $\theta \in \Theta$ زیرمجموعه‌ای باز در \mathbb{R}^p است. همچنین فرض کنید $\ell_t(\theta) = \ln(L_t(\theta; x_t))$ تابع لگاریتم درست‌نمایی مدل باشد که نسبت به θ دو بار مشتق‌پذیر است. علاوه بر این، فرض کنید برآوردگر درست‌نمایی ماکسیمم، $\hat{\theta}$ ، وجود دارد و در معادله $\nabla \ell(\hat{\theta}_t) = 0$ صدق می‌کند و $-\nabla^2 \ell_t(\hat{\theta}_t)$ معین مثبت است، که در آن $\nabla \ell_t(\theta)$ و $\nabla^2 \ell_t(\theta)$ نشان‌دهنده به ترتیب بردار مشتقات جزئی مرتبه اول و ماتریس مشتقات جزئی مرتبه دوم نسبت به θ هستند.

ساخت مدل بیزی نیازمند تعیین توزیع پیشین برای پارامترهای مدل است. فرض کنید پارامترهای θ دارای تابع چگالی پیشین $\xi(\theta)$ باشد. در این صورت، تابع چگالی پسین θ ، به شرط داده‌های x_t ، به صورت زیر تعریف می‌شود:

$$\xi_t(\theta) \propto L(\theta; x_t)\xi(\theta) = \exp(\ell_t(\theta))\xi(\theta).$$

کمیت‌های Σ_t و Z_t را به صورت زیر تعریف می‌کنیم:

$$\Sigma_t^T \Sigma_t = -\nabla^2 \ell_t(\hat{\theta}_t), \quad (5.2)$$

$$Z_t = \Sigma_t(\theta - \hat{\theta}_t). \quad (6.2)$$

با توجه به این که

$$\frac{\xi_t(\theta)}{\xi(\theta)} \propto \exp(\ell_t(\theta)),$$

تابع چگالی پسین Z_t عبارتست از

$$\zeta_t(z) = \xi_t(\theta(z)) \propto \exp(\ell_t(\theta) - \ell_t(\hat{\theta}))\xi(\theta), \quad (7.2)$$

که در آن رابطه θ و z در (۶.۲) بیان شده است. فرض کنید θ مقدار واقعی (نامعلوم) پارامترهای مدل باشد. علاوه بر این، فرض کنید P_t^c و E_t^c به ترتیب تابع احتمال و امید ریاضی شرطی، به شرط بردار x_t باشند. در ادامه همه عبارات احتمالی نسبت به توزیع احتمال واقعی (تحت θ) بیان می‌شوند. برای اثبات همگرایی توزیع پسین پارامترها به توزیع نرمال، باید نشان دهیم وقتی $t \rightarrow \infty$ ، آن‌گاه

$$P_t^c(Z_t \in B) \rightarrow \Phi_p(B), \quad (۸.۲)$$

که در آن B مجموعه بورل دلخواه در \mathbb{R}^p است.

به منظور تبدیل $\ell_t(\theta)$ به شکلی مناسب برای اتحاد استاین، به معرفی روابط زیر نیاز داریم. برای تبدیل

$\ell_t(\theta)$ به شکلی نزدیک به هسته چگالی نرمال، بسط تیلور آن حول $\hat{\theta}$ را به صورت

$$\ell_t(\theta) \approx \ell_t(\hat{\theta}) + \frac{1}{\gamma}(\theta - \hat{\theta})^T \nabla^2 \ell_t(\theta^*)(\theta - \hat{\theta}),$$

در نظر بگیرید، که در آن θ^* بین θ و $\hat{\theta}$ قرار دارد. قرار می‌دهیم

$$u_t(\theta) = -\frac{1}{\gamma}(\theta - \hat{\theta})^T [\nabla^2 \ell_t(\hat{\theta}) - \nabla^2 \ell_t(\theta^*)](\theta - \hat{\theta}).$$

بنابراین

$$\begin{aligned} \ell_t(\theta) &\approx \ell_t(\hat{\theta}) + \frac{1}{\gamma}(\theta - \hat{\theta})^T \nabla^2 \ell_t(\hat{\theta})(\theta - \hat{\theta}) \\ &\quad - \frac{1}{\gamma}(\theta - \hat{\theta})^T \nabla^2 \ell_t(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{\gamma}(\theta - \hat{\theta})^T \nabla^2 \ell_t(\theta^*)(\theta - \hat{\theta}) \\ &\approx \ell_t(\hat{\theta}) \mp \frac{1}{\gamma} \|z_t\|_d^2 + \frac{1}{\gamma}(\theta - \hat{\theta})^T \nabla^2 \ell_t(\theta^*)(\theta - \hat{\theta}) \\ &\approx \ell_t(\hat{\theta}) - \frac{1}{\gamma} \|z_t\|_d^2 + u_t(\theta). \end{aligned} \quad (۹.۲)$$

بنابراین (۷.۲) را می‌توان به صورت زیر بازنویسی کرد:

$$\begin{aligned} \zeta_t(Z_t) &\propto e^{u_t(\theta) - \frac{1}{\gamma} \|z_t\|_d^2} \xi(\theta(z)) \\ &= \phi_p(z) f_t(z), \end{aligned} \quad (۱۰.۲)$$

که در آن $f_t(z) = \xi(\theta(z)) \exp(u_t(\theta))$. همان طور که مشاهده می کنید اکنون توزیع پسین Z_t دارای صورتی است که برای اتحاد استاین مناسب می باشد. اگر توزیع پیشین ξ در \mathbb{R}^p دو بار مشتق پذیر باشد، آن گاه $f_t(z)$ نیز همین طور خواهد بود. علاوه بر این، اگر (۳.۲) برقرار باشد، آن گاه با توجه به لم ۲.۲.۲، داریم

$$E_t^c\{h(Z_t)\} = \Phi_p h + E_t^c\left\{[Uh(z_t)]^T \frac{\nabla_{Z_t} f_t(Z_t)}{f_t(Z_t)}\right\}. \quad (11.2)$$

به عنوان مثال اگر $h(z) = z_i$ باشد، آن گاه $Uh(z) = e_i$ ، که در آن e_i برداری p بعدی از صفرهاست که فقط عضو i ام آن ۱ است. در این صورت

$$E_t^c(Z_t) = E_t^c\left(\frac{\nabla_{Z_t} f_t(Z_t)}{f_t(Z_t)}\right), \quad (12.2)$$

که در آن $\nabla_{z_t} f_t(z_t)$ بردار مشتقات جزئی مرتبه اول تابع f_t نسبت به Z_t است.

گزاره ۱.۳.۲. با توجه به رابطه (۱۱.۲)، شرط لازم برای برقراری (۸.۲) آن است که

$$E_t^c\left\{[Uh(z_t)]^T \frac{\nabla_{Z_t} f_t(Z_t)}{f_t(Z_t)}\right\} \xrightarrow{P_{\theta_t}} 0.$$

از رابطه (۹.۲) داریم

$$\nabla u_t(\theta) = \nabla \ell_t(\theta) - \nabla^2 \ell_t(\hat{\theta})(\theta - \hat{\theta}).$$

با نوشتن بسط تیلور $\nabla \ell_t(\theta)$ خواهیم داشت

$$\nabla u_t(\theta) = \left[\frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta^{*ij}} - \nabla^2 \ell_t(\hat{\theta}) \right] (\theta - \hat{\theta}),$$

که در آن $\theta^{*il} = \theta^{*ir}$ ، $i, l, r = 1, \dots, p$ قرار دارند و $(\partial^2 \ell_t(\theta) / \partial \theta_i \partial \theta_j) |_{\theta^{*ij}}$ ماتریس هسیان^۴

^۴Hessian

$\ell_t(\theta)$ است که مولفه (i, j) ام آن در θ^{*ij} مقداردهی شده است. بنابراین

$$\begin{aligned}
 (\Sigma_t^T)^{-1} \nabla u_t(\theta) &= (\Sigma_t^T)^{-1} \left[\frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta^{*ij}} - \nabla^2 \ell_t(\hat{\theta}) \right] (\theta - \hat{\theta}) \\
 &= (\Sigma_t^T)^{-1} \left[-\nabla^2 \ell_t(\hat{\theta}) + \frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta^{*ij}} \right] \Sigma_t^{-1} \Sigma_t (\theta - \hat{\theta}) \\
 &= (\Sigma_t^T)^{-1} \left[-\nabla^2 \ell_t(\hat{\theta}) + \frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta^{*ij}} \right] \Sigma_t^{-1} Z_t \\
 &= (\Sigma_t^T)^{-1} \left[\Sigma_t^T \Sigma_t + \frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta^{*ij}} \right] \Sigma_t^{-1} Z_t \\
 &= \left\{ I_p - (\Sigma_t^T)^{-1} \left[-\frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta^{*ij}} \right] \Sigma_t^{-1} \right\} Z_t, \quad (13.2)
 \end{aligned}$$

که در آن I_p ماتریس همانی با بعد p است. بنابراین

$$\frac{\nabla_{Z_t} f_t(Z)}{f_t(Z)} = (\Sigma_t^T)^{-1} \left[\frac{\nabla \xi(\theta)}{\xi(\theta)} + \nabla u_t(\theta) \right]. \quad (14.2)$$

قرار می‌دهیم $\{ \nabla \ell_t(\hat{\theta}_t) = 0, -\nabla^2 \ell_t(\hat{\theta}_t) > 0 \}$. در این جا $A > 0$ به معنی معین مثبت بودن ماتریس A است.

۴.۲ شرایط نظم

در این بخش، شرایط نظم لازم برای برقراری نتایج مجانبی را ارائه می‌دهیم. برای این منظور، فرض کنید $B(a; d)$ نشان‌دهنده یک همسایگی p بعدی حول a به شعاع $r > 0$ باشد. علاوه بر این، تعریف می‌کنیم

$$S_t = \{ z : z = \Sigma_t(\theta - \hat{\theta}), \theta \in \Theta_\xi \}.$$

شرایط نظم لازم برای توزیع پیشین:

(P1) $\xi(\theta)$ روی \mathbb{R}^p به‌طور پیوسته مشتق‌پذیر است.

(P2) $\xi(\theta)$ دارای تکیه‌گاه فشرده $\mathbb{R}^k \subset \Theta_\xi$ است.

(P3) مقادیر $\varepsilon > 0$ و $\delta > 0$ وجود دارند، به طوری که در همسایگی $(B(\theta_0, \delta), \xi(\theta) > \varepsilon)$.

شرایط نظم لازم برای تابع لگاریتم درستنمایی:

(L۱) هرگاه $t \rightarrow \infty$ ، آن گاه $\hat{\theta} \xrightarrow{P_{\theta_0}} \theta$ ، $\|\Sigma_t^{-1}\| \xrightarrow{P_{\theta_0}} 0$ ، $P_{\theta_0}(D'_t) \rightarrow 0$ ، که منظور از D'_t متمم مجموعه D_t است.

(L۲) دنباله‌ای صعودی و همگرا به ∞ از مقادیر ثابت و مثبت $\{b_t\}$ وجود دارد به طوری که

$$\sup_{\eta^{ij} \in \{\theta: \|z_t\| \leq b_t\}} \|I_p + (\Sigma_t^T)^{-1} \left(\frac{\partial^2 \ell_t(\eta^{ij})}{\partial \theta_i \partial \theta_j} \right) \Sigma_t^{-1}\| \xrightarrow{P_{\theta_0}} 0.$$

(L۳) (i) برای دنباله b_t در (L۲)، ثابت‌های $r \geq 1$ و $c \geq 0$ وجود دارند به طوری که برای همه

$$\theta \in \{\|z_t\| > b_t\} \cap \Theta_\varepsilon$$

$$(\Sigma_t^T)^{-1} \nabla u_t(\theta) \leq c \|z_t\|^r.$$

(ii) ثابت $r \geq 1$ و تابع نامنفی $g: \mathbb{R}^+ \times \mathbb{R}^p \rightarrow \mathbb{R}$ وجود دارند به طوری که با احتمال متمایل به

۱ و به ازای هر $\theta \in \Theta_\varepsilon$ ، $[\ell_t(\hat{\theta}) - \ell_t(\theta)] \geq g(t, \theta)$ ، $\|z_t\|^r e^{-g(t, \theta)}$ نسبت

به t به طور یکنواخت انتگرال پذیرند و $\int_{\Theta_\varepsilon} m_t(\theta) d\theta$ نسبت به t به طور یکنواخت کراندارند.

در شرایط نظم لازم برای تابع لگاریتم درستنمایی، شرط (L۱) مربوط به رشد اطلاع، شرط (L۲) مربوط به پیوستگی اطلاع و شرط (L۳) که برخی از ویژگی‌های انتگرال پذیری را بیان می‌کند، مربوط به رفتار دمی تابع درستنمایی می‌باشند.

وَنگ و تی‌سی (۲۰۰۸) شرط جانشین دیگری را برای (ii) (L۳) بیان کردند که همه نتایج مجانبی که

در بخش بعدی ارایه می‌شوند، با جانشین کردن آن معتبر باقی می‌مانند.

۵.۲ توزیع مجانبی پسین

در این بخش نشان می‌دهیم توزیع پسین پارامترها به طور مجانبی نرمال است. برای این منظور، ابتدا چند

لم را بیان می‌کنیم که اثبات آن‌ها همگی در وَنگ و تی‌سی (۲۰۰۸) قابل دسترس هستند. با در نظر گرفتن

رابطه (۹.۲) و شرایط نظم، سه لم زیر به سادگی اثبات می‌شوند.

لم ۱.۵.۲. الف) تحت شرط (L۲)، ثابت‌های s و q وجود دارند به طوری که

$$\sup_{\{\theta: \|z_t\| \leq s\}} \left\{ \ell_t(\hat{\theta}) - \ell_t(\theta) \right\} \stackrel{P}{\leq} q,$$

که در آن منظور از نماد $\stackrel{P}{\leq}$ ، برقراری نامساوی با احتمال متمایل به ۱ است. به این معنی که

$$P\left(\sup_{\{\theta: \|z_t\| \leq s\}} \left\{ \ell_t(\hat{\theta}) - \ell_t(\theta) \right\} \leq q \right) \rightarrow 1.$$

ب) تحت شرط (L۳)، برای $0 < M < \infty$ ،

$$\begin{aligned} \int_{S_t} e^{\ell_t(\theta) - \ell_t(\hat{\theta})} dz_t &\stackrel{P}{\leq} M, \\ \int_{S_t} \|z_t\| e^{\ell_t(\theta) - \ell_t(\hat{\theta})} dz_t &\stackrel{P}{\leq} M, \\ \int_{S'_t} \|z_t\|^r e^{\ell_t(\theta) - \ell_t(\hat{\theta})} dz_t &\xrightarrow{P} 0, \end{aligned}$$

که در آن $S'_t = S_t \cap \{\|z_t\| > b_t\}$.

لم ۲.۵.۲. فرض کنید $\xi(\theta)$ در شرایط (P۱) تا (P۳) صدق کند. در این صورت

(i) اگر شرایط (L۱) و (L۲) نیز برقرار باشند، آن‌گاه $K_1 > 0$ وجود دارد به طوری که

$$\int_{S_t} \phi_p(z_t) f_t(z_t) dz_t \stackrel{P}{\geq} K_1.$$

(ii) اگر شرط (ii) (L۳) برقرار باشد، آن‌گاه $K_2 > 0$ وجود دارد به طوری که

$$\int_{S_t} \phi_p(z_t) f_t(z_t) dz_t \stackrel{P}{\leq} K_2.$$

لم ۳.۵.۲. تحت شرایط مشابه لم ۲.۵.۲، ثابت $0 < C < \infty$ وجود دارد به طوری که با احتمال مایل به ۱،

$$E_t^c \left(\frac{\|\nabla \xi(\theta)\|}{\xi(\theta)} \right) \geq C.$$

اکنون می‌توان قضیه اصلی را بیان کرد. قضیه زیر نشان می‌دهد توزیع پسین پارامترها به طور مجانبی

نرمال است.

قضیه ۴.۵.۲. فرض کنید $h(\cdot)$ تابع اندازه‌پذیر کراندار باشد، یا به عبارتی $h \in H_s$ و شرایط (P۱) تا (P۳)

برای پیشین $\xi(\theta)$ و شرایط (L۱) تا (L۳) برای $\ell_t(\theta)$ برقرار باشند. اگر $t \rightarrow \infty$ ، آن‌گاه

$$E_t^c[h(Z_t)] \xrightarrow{P_{\theta}} \Phi h.$$

برهان. با توجه به لم ۳.۲.۲ و شرایط (P۱) تا (P۳)، توابع Uh و $\xi(\theta)$ کراندار هستند. از (۱۱.۲) و

(۱۴.۲)، تقریباً همه جا روی D_t داریم

$$E_t^c[h(Z_t)] - \Phi h = E_{t,Z_t}^c,$$

به طوری که

$$\begin{aligned} E_{t,Z_t}^c &= E_t^c \left\{ (Uh(Z_t))^T (\Sigma_t^T)^{-1} \frac{\nabla \xi(\theta)}{\xi(\theta)} \right\} \\ &+ E_t^c \left\{ (Uh(Z_t))^T (\Sigma_t^T)^{-1} \nabla u_t(\theta) \right\} = I_{z_t} + II_{z_t}. \end{aligned} \quad (۱۵.۲)$$

از آن‌جا که بنا بر شرط (L۱)، $P_t^c(D_t') \rightarrow 0$ ، برای اثبات قضیه کفایت نشان دهیم

$$I_{z_t} + II_{z_t} \xrightarrow{P} 0.$$

بنا بر لم ۳.۵.۲ و رابطه $\|\Sigma_t^{-1}\| \xrightarrow{P} 0$ در شرط (L۱)، داریم $I_{z_t} \xrightarrow{P} 0$. از (۱۵.۲) داریم

$$II_{z_t} = \frac{\int_{S_t} (Uh(z_t))^T (\Sigma_t^T)^{-1} \nabla u_t(\theta) \phi_p(z_t) f_t(z_t) dz_t}{\int_{S_t} \phi_p(z_t) f_t(z_t) dz_t}. \quad (۱۶.۲)$$

بنا بر قسمت الف) لم ۲.۵.۲، مخرج کسر (۱۶.۲) از پایین توسط $K_1 > 0$ کراندار است. بنابراین تنها

باید نشان دهیم صورت کسر در احتمال به صفر میل می‌کند. ابتدا صورت کسر (۱۶.۲) را به دو انتگرال با

دامنه‌های $\|z_t\| \leq b_t$ و $\|z_t\| > b_t$ تجزیه می‌کنیم و به ترتیب با $II_{z_t,1}$ و $II_{z_t,2}$ نشان می‌دهیم. بنا بر شرایط

(P۱) و (P۲)، لم ۳.۲.۲ و رابطه (۱۳.۲)، ثابتی مانند $C_1 > 0$ وجود دارد که

$$\begin{aligned} |II_{z_t,1}| &\leq \int_{\|z_t\| \leq b_t} |(Uh(z_t))^T (\Sigma_t^T)^{-1} \nabla u_t(\theta) \xi(\theta(z_t)) e^{\ell_t(\theta) - \ell_t(\hat{\theta})} dz_t \\ &\leq C_1 \left\{ \sup_{\theta: \|z_t\| \leq b_t} \|I_p - (\Sigma_t^T)^{-1} \left[-\left(\frac{\partial^2 \ell_t(\theta)}{\partial \theta_i \partial \theta_j} \Big|_{\theta^{*ij}} \right) \right] \Sigma_t^{-1} \| \right\} \\ &\times \int_{\|z_t\| \leq b_t} \|z_t\| e^{\ell_t(\theta) - \ell_t(\hat{\theta})} dz_t. \end{aligned}$$

بنابراین با استفاده از شرط (L۲) و قسمت دوم لم ۱.۵.۲، می‌توان نتیجه گرفت $II_{z_t, 1} \xrightarrow{P} 0$. بنا بر شرایط (P۱) تا (P۳)، شرط (L۳) و لم ۳.۲.۲، ثابتی مانند $C_2 > 0$ وجود دارد که

$$|II_{z_t, 2}| \leq C_2 \int_{S_t \cap \{\|z_t\| > b_t\}} \|z_t\|^r e^{\ell_t(\theta) - \ell_t(\hat{\theta})} dz_t. \quad (17.2)$$

بنا به قسمت دوم لم ۱.۵.۲، طرف راست نامساوی (۱۷.۲) در احتمال به صفر میل می‌کند. بنابراین $II_{z_t, 2} \xrightarrow{P} 0$. با این نتیجه برهان قضیه کامل می‌شود. \square

نتیجه ۵.۵.۲. اگر $h(Z_t) = I_A(Z_t)$ که در آن

$$I(Z_t) = \begin{cases} 1 & Z_t \leq z_t \\ 0 & o.w \end{cases}$$

آن‌گاه

$$E_t^c(h(Z_t)) = P_t^c(Z_t \leq z_t)$$

که همان توزیع پسین Z_t است.

دو شرط (P۱) و (P۲) برای $\xi(\theta)$ ، توزیع‌های پیشین را به توزیع‌هایی با دامنه فشرده و به‌طور پیوسته مشتق‌پذیر محدود می‌کنند. به عنوان مثال، بنا بر شرایط این قضیه، انتخاب توزیع پیشین نرمال یا گاما (که انتخاب‌هایی معمول در استنباط بیزی برای پارامترهای مکان و مقیاس هستند) مجاز نیست. نتیجه زیر نشان می‌دهد همگرایی توزیع‌های پسین به توزیع نرمال، با انتخاب توزیع‌های پیشین کلی‌تر نیز برقرار خواهد بود.

نتیجه ۶.۵.۲. با دنبال کردن دو قضیه ۴.۲ و ۴.۳ در ونگ و تی‌سی (۲۰۰۸) و استدلال قضیه ۴.۵.۲، همگرایی توزیع پسین به توزیع نرمال با انتخاب توزیع‌های پیشینی که دامنه فشرده ندارند یا کران‌دار و مشتق‌پذیر نیستند، برقرار خواهد بود.

فصل ۳

تقریب توزیع پسین با بسط اچوورث

در این فصل نشان می‌دهیم که چگونه با استفاده از بسط اچوورث و اتحاد استاین می‌توان تقریب بهتری (بر مبنای مرتبه بالاتر از ۲) برای توزیع پسین ارائه کرد. سپس بسط دیگری برای تقریب توزیع پسین با نام بسط جانسون معرفی می‌کنیم و به مقایسه آن با بسط اچوورث می‌پردازیم.

۱.۳ مقدمه

پایه‌ای‌ترین نتیجه بسط اچوورث برای متغیرهای تصادفی مستقل و هم‌توزیع X_1, \dots, X_n با میانگین θ و واریانس متناهی σ^2 (رابطه (۱۳.۱) در فصل اول) است؛ اگر $\hat{\theta}_n$ میانگین نمونه X_i ها باشد، تحت شرایط نظم، تابع توزیع $Y = n^{1/2}(\hat{\theta}_n - \theta)$ به صورت زیر قابل بسط است:

$$P\left(\frac{n^{1/2}(\hat{\theta}_n - \theta)}{\sigma} \leq x\right) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + \dots + n^{-j/2}p_j(x)\phi(x) + \dots$$

با توجه به مطالب فصل اول، توابع p_j چندجمله‌ای‌هایی هستند که ضرایب آن‌ها به انباشتک‌های $\hat{\theta}_n - \theta$ وابسته هستند. همچنین p_j یک چندجمله‌ای از درجه $3j - 1$ است و بر حسب آن که j زوج یا فرد باشد، این توابع فرد یا زوج هستند.

محققان مختلفی بسط اچوورث را برای مواردی که داده‌ها مستقل و هم‌توزیع نیستند، به دست آورده‌اند. به عنوان نمونه، جینگ و وانگ (۲۰۰۳)، بسط‌ها را برای U -آماره‌ها^۱ معرفی کردند. مطالعاتی در مورد این

^۱U-statistics

بسط از دیدگاه بیزی نیز انجام شده‌اند. فرض کنید $g(\cdot)$ تابعی هموار از پارامتر θ باشد. رهیافت معمول برای بسط مجانبی توزیع پسین پارامتر، از نوشتن میانگین پسین $g(\theta)$ به عنوان نسبت دو انتگرال به صورت

$$E_{\xi} [g(\theta)|x_t] = \frac{\int g(\theta) \exp(\ell_t(\theta)) \xi(\theta) d\theta}{\int \exp(\ell_t(\theta)) \xi(\theta) d\theta},$$

شروع می‌شود. سپس برای صورت و مخرج، بسط سری‌های تیلور حول MLE نوشته می‌شود و با تقسیم دو سری نتیجه‌شده، تقریبی از میانگین پسین $g(\theta)$ حاصل می‌شود. جانسون (۱۹۶۷ و ۱۹۷۰) محتوایی کامل برای این نوع رهیافت فراهم کرده است. مطالعات دیگری انجام شده‌اند که تقریب لاپلاس را برای صورت و مخرج کسر محاسبه کرده و نسبت آن‌ها را به عنوان تقریب میانگین پسین ارائه داده‌اند که به عنوان مثال می‌توان به لیندلی (۱۹۶۱ و ۱۹۸۰)، ماستر و والاس (۱۹۶۴)، تیرنی و کدین (۱۹۸۶) و منابع داخل آن‌ها اشاره کرد. البته دقت داشته باشید که این بسط‌های مجانبی بر حسب تابع درستنمایی و پیشین نوشته می‌شوند نه انباشتک‌ها یا گشتاورهای توزیع پسین.

پس از ارائه تقریبی از توزیع پسین با توزیع نرمال به کمک اتحاد استاین توسط ونگ (۲۰۰۳)، در حالت تک‌پارامتری، و سپس ونگ و تی‌سی (۲۰۰۸)، در حالت چندپارامتری، که در فصل دوم به تفصیل به آن پرداخته شد، ونگ و لین (۲۰۱۰) با تعمیم کار آن‌ها یک بسط مجانبی را بر اساس اتحاد استاین معرفی کردند و برای کاربردهای رتبه‌دهی برخط بیزی^۲ مورد استفاده قرار دادند. ونگ (۲۰۱۰) و ونگ و سو (۲۰۱۳) جزئیات کاملی از این نوع بسط ارائه داده‌اند.

تقریبی که در این فصل ارائه می‌شود مطابق با کار ونگ (۲۰۱۰) و ونگ و سو (۲۰۱۳) است. این تقریب از تقریب حاصل از نرمال مجانبی، که در فصل دوم ارائه شد، بهتر است؛ زیرا از جملات مرتبه‌های بالاتر از ۲ استفاده می‌کند. البته باید متذکر شویم که بسط اچوورث ارائه‌شده در این فصل برای توزیع‌های پسین کناری^۳ می‌باشد نه توزیع پسین توام. همچنین توجه کنید که این بسط از دو منظر شبیه به بسط اچوورث کلاسیک می‌باشد؛ یکی این که هر دو به‌طور مستقیم با انباشتک‌ها یا گشتاورها در ارتباط هستند و دیگر این که می‌توان آن‌ها را یک بسط از توزیع احتمال بر حسب توابع متعامد هرمیتی و عباراتی که بر پایه

^۲Bayesian online ranking

^۳Marginal posterior distribution

توان‌های مختلف حجم نمونه متحد شده‌اند، در نظر گرفت. این دو ویژگی در بسط‌های پسین موجود، وجود ندارند. مزیت بیان یک توزیع بر حسب گشتاورها یا انباشتک‌هایش، این است که اطلاعات توزیع به‌طوری کارا ذخیره و نمایش داده می‌شوند.

۲.۳ اتحاد استاین برای بسط توزیع پسین

علاوه بر تعاریف بخش ۲.۲، در فصل دوم، فرض کنید $Vh = (U^\top h + (U^\top h)^T)/2$ به‌طوری که $U^\top h$ ماتریسی است $p \times p$ که ستون k ام آن Ug_k می‌باشد و g_k در (۲.۲) تعریف شده است. برای روشن‌تر شدن محاسبه Vh مثال ۱.۲.۲ در فصل دوم را در نظر بگیرید. در این مثال نتیجه شد

$$Uh(z) = (g_1, g_2, g_3)^T = (0, z_1, 0)^T.$$

همچنین $Ug_1 = Ug_3 = [0, 0, 0]^T$ و $Ug_2 = [1, 0, 0]^T$. بنابراین

$$Vh = \frac{1}{4}(U^\top h + (U^\top h)^T) = \frac{1}{4} \left(\begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} + \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix} \right) = \begin{bmatrix} \cdot & \cdot & \cdot \\ \cdot & 1/2 & \cdot \\ \cdot & \cdot & \cdot \end{bmatrix}.$$

اکنون صورت مناسبی از لم استاین را برای استفاده در بسط اچوورث توزیع پسین، معرفی می‌کنیم. در ادامه، منظور از $tr(A)$ اثر ماتریس A است.

لم ۱.۲.۳. (اتحاد استاین تعمیم‌یافته) مشابه لم ۲.۲.۲، فرض کنید r عدد صحیح نامنفی باشد. همچنین فرض کنید $d\Gamma = f d\Phi_p$ که f تابع مشتق‌پذیر روی \mathbb{R}^p است (دقت داشته باشید که تحت این شرایط،

لم ۲.۲.۲ برقرار است). اگر $\frac{\partial}{\partial z_j} f(z)$ ، $j = 1, \dots, p$ ، مشتق‌پذیر باشند و

$$\int_{\mathbb{R}^p} (1 + \|z\|^r) \|\nabla^\top f(z)\| \Phi_p(dz) < \infty, \quad (1.3)$$

آن‌گاه برای همه $h \in H_r$

$$\Gamma h = \Gamma f \cdot \Phi_p h + (\Phi_p U h)^T \int_{\mathbb{R}^p} \nabla f(z) \Phi_p(dz) + \int_{\mathbb{R}^p} tr[(Vh(z)) \nabla^\top f(z)] \Phi_p(dz). \quad (2.3)$$

برهان. می‌توان نوشت

$$(Uh(z))^T \nabla f(z) = \sum_{i=1}^p g_i(z) \frac{\partial f(z)}{\partial z_i}. \quad (3.3)$$

اکنون با به کار بردن (۴.۲) در لم ۲.۲.۲ برای g_i و $\frac{\partial f}{\partial z_i}$ به جای h و f ، با شرط وجود همه انتگرال‌ها، نتیجه می‌شود

$$\int g_i \frac{\partial f}{\partial z_i} d\Phi_p(z) = \Phi_p(g_i) \int \frac{\partial f}{\partial z_i} d\Phi_p(z) + \int (U(g_i))^T \nabla \left(\frac{\partial f}{\partial z_i} \right) d\Phi_p(z). \quad (۴.۳)$$

توجه کنید $\Phi_p(g_i)$ در معادله بالا یک مقدار ثابت است که به صورت زیر تعریف می‌شود:

$$\Phi_p(g_i) = \int g_i(z) \phi_p(z) dz.$$

با جمع‌بندی هر دو طرف تساوی (۴.۳) روی مقادیر $i = 1, \dots, p$ ، رابطه (۴.۲) در لم ۲.۲.۲ را می‌توان به صورت زیر بازنویسی کرد:

$$\begin{aligned} \int h(z) f(z) d\Phi_p(z) &= \int f(z) d\Phi_p(z) \cdot \int h(z) d\Phi_p(z) + (\Phi_p U h)^T \int \nabla f(z) d\Phi_k(z) \\ &\quad + \int \text{tr}[(V h(z)) \nabla^\top f(z)] d\Phi_p(z), \end{aligned}$$

□

که در نتیجه (۲.۳) اثبات می‌شود.

در لم ۱.۲.۳، $\Phi_p U h = (\Phi_p(g_1), \dots, \Phi_p(g_p))^T$ ، در این لم، با در نظر گرفتن $f(z)$ به صورت $z_i z_j$ و $z_i z_j$ ، می‌توان نتیجه گرفت

$$\Phi_p(Uh) = \int_{\mathbb{R}^p} z h(z) \Phi_p(dz), \quad (۵.۳)$$

$$\Phi_p(U^\top h) = \int_{\mathbb{R}^p} \frac{1}{2} (z z^T - 1) h(z) \Phi_p(dz). \quad (۶.۳)$$

۳.۳ بسط اچوورث بیزی

در این بخش، ابتدا به بیان روابطی که بعداً مورد استفاده قرار می‌گیرند، می‌پردازیم. سپس بسط اچوورث بیزی را توسعه می‌دهیم.

مشابه رابطه (۱۱.۲) در فصل دوم، اگر (۱.۳) برقرار باشد از لم ۱.۲.۳ داریم

$$E_t^c \{h(Z_t)\} = \Phi_p h + (\Phi_p U h)^T E_t^c \left[\frac{\nabla_{Z_t} f_t(Z_t)}{f_t(Z_t)} \right] + E_t^c \left\{ \text{tr} \left[V h(Z_t) \frac{\nabla_{Z_t}^\top f_t(Z_t)}{f_t(Z_t)} \right] \right\}.$$

به عنوان مثال اگر $h(z) = z_i z_j$ ، $i < j$ ، آن گاه $Uh(z) = z_i e_j$. در این صورت

$$E_t^c(Z_{ti} Z_{tj}) = \delta_{ij} + E_t^c \left[\frac{\nabla_{Z_t}^2 f_t(Z_t)}{f_t(Z_t)} \right]_{ij}, \quad (۷.۳)$$

که در آن $\delta_{ij} = 1$ هرگاه $i = j$ و به جز این صفر خواهد بود و $[\cdot]_{ij}$ نشان دهنده مولفه (i, j) ام یک ماتریس است.

ملاحظه ۱.۳.۳. از آن جا که در استنباط بیزی، معمولاً، برآوردگر پارامتر و واریانس آن، میانگین و واریانس توزیع پسین هستند، در اکثر کاربردها محاسبه روابط (۱۲.۲) و (۷.۳) برای اهداف استنباط کافیت.

سایر روابطی را که در بخش های بعدی مورد استفاده قرار می گیرند، در زیر فهرست کرده ایم. با توجه به

تعریف تابع $f_t(Z_t)$ در فصل دوم، می توان نوشت

$$\begin{aligned} \nabla_{Z_t} f_t(Z_t) &= \frac{\partial \xi(\theta(Z))}{\partial Z} e^{u_t(\theta)} + \frac{\partial e^{u_t(Z)}}{\partial Z} \xi(\theta(Z)) \\ &= \nabla \xi(\theta) \times \frac{\partial \theta}{\partial Z} e^{u_t(\theta)} + \nabla u_t(\theta) \times \frac{\partial \theta}{\partial Z} e^{u_t(\theta)} \xi(\theta(Z)) \\ &= \nabla \xi(\theta) (\Sigma_t^T)^{-1} e^{u_t(\theta)} + \nabla u_t(\theta) (\Sigma_t^T)^{-1} \xi(\theta(Z)) \\ &= (\Sigma_t^T)^{-1} e^{u_t(\theta)} \xi(\theta(Z)) \left(\frac{\nabla \xi(\theta)}{\xi(\theta)} + \nabla u_t(\theta) \right) \\ &= f_t(Z_t) (\Sigma_t^T)^{-1} \left[\frac{\nabla \xi(\theta)}{\xi(\theta)} + \nabla u_t(\theta) \right]. \end{aligned}$$

بنابراین $\nabla_{Z_t}^2 f_t(Z_t)$ به صورت زیر نتیجه می شود:

$$\begin{aligned} \nabla_{Z_t}^2 f_t(Z_t) &= \nabla_{Z_t} f_t(Z_t) (\Sigma_t^T)^{-1} \left(\frac{\nabla \xi(\theta)}{\xi(\theta)} + \nabla u_t(\theta) \right) \\ &\quad + (\Sigma_t^T)^{-1} \left[\frac{\nabla^2 \xi}{\xi} - \nabla \xi^T \frac{\nabla \xi}{\xi} + \nabla^2 u_t(\theta) \right] f_t(Z_t) (\Sigma_t^T)^{-1} \\ &= (\Sigma_t^T)^{-1} \left[\frac{\nabla^2 \xi}{\xi} + \nabla^2 u_t(\theta) + \frac{\nabla \xi}{\xi} \nabla u_t^T + \nabla u_t \frac{\nabla \xi^T}{\xi} + \nabla u_t \nabla u_t^T \right] (\Sigma_t)^{-1} f_t(Z_t). \end{aligned}$$

در نتیجه

$$\frac{\nabla_{Z_t}^2 f_t(Z_t)}{f_t(Z_t)} = (\Sigma_t^T)^{-1} \left[\frac{\nabla^2 \xi}{\xi} + \nabla^2 u_t(\theta) + \frac{\nabla \xi}{\xi} \nabla u_t^T + \nabla u_t \frac{\nabla \xi^T}{\xi} + \nabla u_t \nabla u_t^T \right] (\Sigma_t)^{-1}. \quad (۸.۳)$$

با استفاده از رابطه (۹.۲) می‌توان نوشت

$$\begin{aligned}\nabla u_t(\theta) &= \nabla \ell_t(\theta) + \Sigma_t^T \cdot Z_t \\ &= \nabla \ell_t(\theta) + \Sigma_t^T \Sigma_t(\theta - \hat{\theta}) \\ &= \nabla \ell_t(\theta) - \nabla^2 \ell_t(\hat{\theta})(\theta - \hat{\theta}).\end{aligned}\quad (9.3)$$

و همچنین

$$\nabla^2 u_t(\theta) = \nabla^2 \ell_t(\theta) - \nabla^2 \ell_t(\hat{\theta}). \quad (10.3)$$

یادآوری می‌کنیم که $Uh = (g_1, \dots, g_p)^T$ که در (۲.۲) تعریف شده است. در لم زیر $g_l = (Uh)_l$ می‌باشد.

لم ۲.۳.۳. فرض کنید $h \in H_r$ و $h(z) = h^*(z_i)$ که $i \in \{1, \dots, p\}$ و $h^* : \mathbb{R} \rightarrow \mathbb{R}$ آن‌گاه $(Uh)_l = 0$ هرگاه $l \neq i$ و $(Uh)_i(z) = (Uh)_i(z_i) = Uh^*(z_i)$ که تنها به z_i وابسته است.

برهان. از آنجایی که $h(z) = h^*(z_i)$ ، با استفاده از (۱.۲) برای $l = 0, \dots, i-1$ ، می‌توان نتیجه گرفت $h_l = \Phi_p h$ و برای $l = i, \dots, p$ ، $h_l(z) = h(z_i)$ ، بنابراین با توجه به تعریف (۲.۲)، نتایج مورد نظر حاصل می‌شوند. \square

گزاره زیر از دو لم ۱.۲.۳ و ۲.۳.۳ حاصل می‌شود. این گزاره برای محاسبه بسط اچورث برای توزیع‌های پسین کناری مفید است.

گزاره ۳.۳.۳. فرض کنید r و s اعداد صحیح نامنفی باشند و $d\Gamma = f d\Phi_p$ که f تابعی مشتق‌پذیر روی \mathbb{R}^p است. همچنین فرض کنید $h \in H_r$ به طوری که برای $i \in \{1, \dots, p\}$ ، $h(z) = h^*(z_i)$ و $h^* : \mathbb{R} \rightarrow \mathbb{R}$

علاوه بر این برای $s \leq k$ فرض کنید

$$\int_{\mathbb{R}^p} |f(z)| \Phi_p(dz) + \int_{\mathbb{R}^p} (1 + |z_i|^r) \left\| \frac{\partial^k f(z)}{\partial z_i^k} \right\| \Phi_p(dz) < \infty. \quad (11.3)$$

در این صورت

$$\Gamma h = \Gamma \mathbf{1} \cdot \Phi h^* + \sum_{j=1}^{s-1} (\Phi U^j h^*) \int_{\mathbb{R}^p} \frac{\partial^j f(z)}{\partial z_i^j} \Phi_p(dz) + \int_{\mathbb{R}^p} U^s h^*(z_i) \frac{\partial^s f(z)}{\partial z_i^s} \Phi_p(dz). \quad (12.3)$$

برهان. اگر $h(z) = h^*(z_i)$ بنا بر لم ۲.۳.۳ و (۳.۳)، می توان (۴.۲) را به صورت زیر نوشت:

$$\Gamma h = \Gamma h^* = \Gamma \mathbf{1} \cdot \Phi h^* + \int_{\mathbb{R}^p} U h^*(z_i) \frac{\partial f(z)}{\partial z_i} \Phi_p(dz). \quad (۱۳.۳)$$

سپس، به کار گرفتن (۱۳.۳) با $U h^*$ و $\partial f / \partial z_i$ به جای h^* و f ، نتیجه می شود

$$\Gamma h = \Gamma h^* = \Gamma \mathbf{1} \cdot \Phi h^* + \Phi U h^*(z_i) \int_{\mathbb{R}^p} \frac{\partial f(z)}{\partial z_i} \Phi_p(dz) + \int_{\mathbb{R}^p} U^2 h^*(z_i) \frac{\partial^2 f(z)}{\partial z_i^2} \Phi_p(dz).$$

□ با به کارگیری مکرر (۱۳.۳) با $U^j h^*$ و $\partial^j f / \partial z_i^j$ به جای h^* و f ، (۱۲.۳) نتیجه می شود.

به کار بردن گزاره ۳.۳.۳ برای توزیع پسین Z_t ، نیازمند برقراری شرط انتگرال پذیری (۱۱.۳) است که برای بررسی آن باید مولفه های $\frac{\partial^k f_t(z)}{\partial z_i^k}$ را محاسبه کنیم. برای $k = 1, 2$ ، $\frac{\partial^k f_t(z)}{\partial z_i^k}$ را می توان به کمک (۱۴.۲) و (۸.۳) محاسبه کرد. برای $k \geq 3$ عبارات پیچیده ترند. اما به منظور ارزیابی برقراری (۱۱.۳)، استفاده از یک نماد یک بعدی کافیت. برای هر تابع $g(\theta)$ فرض کنید $g^{(k)}$ مشتق k ام تابع $g(\theta)$ نسبت به θ را نشان دهد. به استقرا ثابت می کنیم

$$\frac{d^k f_t(z)}{dz^k} = \left(\frac{d\theta}{dz}\right)^k f_t(z) G_k(\theta), \quad (۱۴.۳)$$

که در آن

$$G_1 = \frac{\xi^{(1)}}{\xi} + u_t^{(1)}, \quad G_k = G_1 G_{k-1} + G_{k-1}^{(1)}. \quad (۱۵.۳)$$

به عنوان مثال

$$\begin{aligned} G_1(\theta) &= \frac{\xi^{(1)}}{\xi} + u_t^{(1)}, \\ G_2(\theta) &= [u_t^{(1)}]^2 + u_t^{(2)} + 2u_t^{(1)} \frac{\xi^{(1)}}{\xi} - \left(\frac{\xi^{(1)}}{\xi}\right)^2, \\ G_3(\theta) &= [u_t^{(1)}]^3 + 3u_t^{(1)} u_t^{(2)} + u_t^{(3)} + 3[u_t^{(1)}]^2 \frac{\xi^{(1)}}{\xi} + 3u_t^{(2)} \frac{\xi^{(1)}}{\xi} - u_t^{(1)} \left(\frac{\xi^{(1)}}{\xi}\right)^2 \\ &\quad + 2u_t^{(1)} \frac{\xi^{(2)}}{\xi} + \left(\frac{\xi^{(1)}}{\xi}\right)^3 - 2\left(\frac{\xi^{(1)}}{\xi}\right) \left(\frac{\xi^{(2)}}{\xi}\right), \end{aligned}$$

به طوری که $G_1(\cdot)$ و $G_2(\cdot)$ نسخه های یک بعدی (۱۴.۲) و (۸.۳) هستند. به طور کلی، می توان نشان داد

$G_k(\cdot)$ به صورت زیر است:

$$G_k(\theta) = \sum_l c_{kl} \left\{ \left(\prod_{i=1}^k [u_t^{(i)}]^{r_{ki}} \right) \left[\prod_{j=1}^k \left(\frac{\xi^{(j)}}{\xi} \right)^{s_{kj}} \right] \right\}, \quad (16.3)$$

که در آن r_{ki} و s_{kj} در رابطه زیر صدق می‌کنند:

$$\sum_{i=1}^k (i r_{ki}) + \sum_{j=1}^k (j s_{kj}) = k. \quad (17.3)$$

البته توجه داشته باشید که r_{ki} و s_{kj} به l وابسته‌اند، اما در نمادگذاری نمایش داده نشده است.

اکنون به اثبات روابط (15.3) تا (17.3) می‌پردازیم. با توجه به تعریف تابع $f_t(z)$ ، بدیهی است (15.3)

برای $k = 1$ برقرار است. فرض کنید

$$\frac{d^{k-1} f_t(z)}{dz^{k-1}} = \left(\frac{d\theta}{dz} \right)^{k-1} f_t(z) G_{k-1}(\theta).$$

بنابراین

$$\begin{aligned} \frac{d^k f_t(z)}{dz^k} &= \left(\frac{d\theta}{dz} \right)^{k-1} \left[\frac{df_t(z)}{dz} G_{k-1}(\theta) + f_t(z) \frac{dG_{k-1}(\theta)}{d\theta} \frac{d\theta}{dz} \right] \\ &= \left(\frac{d\theta}{dz} \right)^{k-1} \left[\frac{d\theta}{dz} f_t(z) G_1(\theta) G_{k-1}(\theta) + f_t(z) \frac{dG_{k-1}(\theta)}{d\theta} \frac{d\theta}{dz} \right] \\ &= \left(\frac{d\theta}{dz} \right)^k f_t(z) \left(G_1 G_{k-1} + G_{k-1}^{(1)} \right). \end{aligned}$$

در نتیجه (15.3) برقرار است. روابط (16.3) و (17.3) را به روش استقرا اثبات می‌کنیم. برای اثبات آن‌ها

فرض کنید برای G_k ، (16.3) و (17.3) برقرارند. کافی است نشان دهیم G_{k+1} نیز این دو ویژگی را دارد.

می‌توان نوشت $G_{k+1} = G_1 G_k + G_k^{(1)}$. عبارت اول سمت راست تساوی به شکل

$$G_1 G_k = \left(\frac{\xi^{(1)}}{\xi} + u_t^{(1)} \right) \sum_l c_{kl} \left\{ \left(\prod_{i=1}^k [u_t^{(i)}]^{r_{ki}} \right) \left[\prod_{j=1}^k \left(\frac{\xi^{(j)}}{\xi} \right)^{s_{kj}} \right] \right\},$$

و عبارت دوم به صورت $G_k^{(1)} = \frac{dG_k}{d\theta}$ قابل بازنویسی هستند. بنابراین G_{k+1} به شکل (16.3) می‌باشد.

برای (17.3) باید نشان دهیم

$$\sum_{i=1}^{k+1} (i r_{k+1,i}) + \sum_{j=1}^{k+1} (j s_{k+1,j}) = k + 1. \quad (18.3)$$

چون G_k در عامل $\frac{\xi^{(1)}}{\xi}$ ضرب شده است، به توان متناظر برای این عامل عدد ۱ افزوده می‌شود؛ یعنی $S_{k+1,1} = s_{k1} + 1$ و سایر توان‌ها تغییر نمی‌کنند؛ یعنی برای هر i ، $r_{k+1,i} + r_{ki}$ و برای $j \neq 1$ ، $s_{k+1,j} = s_{kj}$ از این رو (۱۸.۳) برای عبارات $G_k(\frac{\xi^{(1)}}{\xi})$ برقرار است. استدلال‌های مشابه برای عبارات $G_k u_t^{(1)}$ به کار می‌رود. اکنون $G_k^{(1)}$ را در نظر بگیرید. این تابع شامل مشتقات $[u_t^{(i)}]^{r_{ki}}$ یا $(\frac{\xi^{(j)}}{\xi})^{s_{kj}}$ نسبت به θ می‌باشد. توجه کنید

$$\frac{d[u_t^{(i)}]^{r_{ki}}}{d\theta} = r_{ki}[u_t^{(i)}]^{r_{ki}-1} u_t^{(i+1)}.$$

بنابراین $r_{k+1,i+1} = r_{k,i+1} + 1$ و $r_{k+1,i} = r_{ki} - 1$ در نتیجه

$$i r_{k+1,i} + (i+1) r_{k+1,i+1} = i r_{ki} + (i+1) r_{k,i+1} + 1$$

که (۱۸.۳) را تایید می‌کند. استدلال برای $(\frac{d}{d\theta})(\frac{\xi^{(j)}}{\xi})^{s_{kj}}$ به‌طور مشابه است و برهان کامل است.

۱.۳.۳ شرط انتگرال‌پذیری

برای برقراری (۱۱.۳) شرایط زیر لازمند:

$$(A1) \text{ برای هر } r > 0, E_t^c(\|Z_t\|^r) = O(1),$$

$$(A2) \text{ برای هر } k \geq 3, \frac{\xi^{(k)}(\theta)}{t} \text{ به‌طور یکنواخت در } t \text{ و در } \theta \in \Theta \text{ کراندار است.}$$

$$(A3) \text{ برای برخی } b > 0 \text{ و } s > 0, \|\frac{\xi^{(k)}}{\xi}\| \leq b(1 + \|\theta\|^s),$$

به یاد بیاورید که $O(1)$ به معنی همگرایی یک دنباله از اعداد حقیقی است، هرگاه $t \rightarrow \infty$. توجه کنید که شرط (A3) برای یک رده گسترده از توزیع‌ها برقرار است و نشان می‌دهد برای برخی $0 < b_t < \infty$ ، که دنباله b_t ممکن است به داده‌های x_t وابسته باشند، می‌توان نوشت

$$\|\frac{\xi^{(k)}}{\xi}\| \leq b(1 + \|\hat{\theta}_t + \Sigma_t^{-1} z_t\|^s) \leq b_t(1 + \|z_t\|^s).$$

اکنون با استفاده از نمادگذاری یک بعدی بالا، می‌توانیم برقراری (۱۱.۳) را تایید کنیم. ابتدا، از آنجایی که ζ_t در (۱۰.۲) یک چگالی پسین است، انتگرال $\int |f_t| \Phi_p(dz)$ متناهی است و ما آن را با C_t نشان می‌دهیم.

از طرفی، از (۱۴.۳) داریم

$$\begin{aligned} & \int (\mathbb{1} + |z|^r) \left| \frac{d^k f_t(z)}{dz^k} \right| \Phi(dz) \\ &= \int (\mathbb{1} + |z|^r) \left| \left(\frac{d\theta}{dz} \right)^k f_t(z) G_k(\theta) \right| \Phi(dz) \\ &= C_t \left(\frac{d\theta}{dz} \right)^k E_t^c [(\mathbb{1} + |Z_t|^r) |G_k(\theta)|] \\ &\leq b_t^* C_t \left(\frac{d\theta}{dz} \right)^k E_t^c \left((\mathbb{1} + |Z_t|^r) (\mathbb{1} + |Z_t|^s) \left(\prod_{i=1}^k [u_t^{(i)}]^{r_i} \right) \right) \end{aligned} \quad (19.3)$$

که $0 < b_t^* < \infty$ و خط آخر از (۱۶.۳) و شرط (A۳) نتیجه می‌شود. علاوه براین از (۹.۳) و بسط تیلور

$\ell_t^{(1)}(\theta)$ حول $\hat{\theta}$ داریم

$$\begin{aligned} u_t^{(1)}(\theta) &= \ell_t^{(1)}(\theta) - \ell_t^{(2)}(\hat{\theta}_t)(\theta - \hat{\theta}_t) \\ &= \ell_t^{(1)}(\hat{\theta}) + \ell_t^{(2)}(\hat{\theta})(\theta - \hat{\theta}) + \frac{1}{\Psi} \ell_t^{(3)}(\eta_t)(\theta - \hat{\theta})^2 - \ell_t^{(2)}(\hat{\theta})(\theta - \hat{\theta}) \\ &= 0 + \frac{1}{\Psi} \ell_t^{(3)}(\eta_t)(\theta - \hat{\theta})^2 \\ &= \frac{1}{\Psi} \ell_t^{(3)}(\eta_t)(\theta - \hat{\theta})^2. \end{aligned} \quad (20.3)$$

و با استفاده از (۱۰.۳) و قضیه مقدار میانگین، می‌توان نوشت

$$u_t^{(2)}(\theta) = \ell_t^{(3)}(\omega_t)(\theta - \hat{\theta}) \quad (21.3)$$

و همچنین با توجه به (۹.۳) داریم

$$u_t^{(3)}(\theta) = \ell_t^{(3)}(\eta_t), \quad (22.3)$$

که ω_t و η_t بین θ و $\hat{\theta}$ قرار دارند. بنابراین، بنا به شرایط (A۱) و (A۲)، سمت راست (۱۹.۳) متناهی می‌باشد. با این تعاریف و اثبات‌ها، می‌توان قضیه زیر را ارائه کرد.

قضیه ۴.۳.۳. (ونگ، ۲۰۱۰). فرض کنید $\xi(\theta)$ و $\ell_t(\theta)$ ، s بار مشتق‌پذیر باشند و شرط‌های (A۱) تا

(A۳) برقرار باشند. بنابراین، برای $k \leq s$ ،

$$\int_{\mathbb{R}^p} |f_t(z)| \Phi_p(dz) + \int_{\mathbb{R}^p} (\mathbb{1} + |z_i|^r) \left\| \frac{\partial^k f_t(z)}{\partial z_i^k} \right\| \Phi_p(dz) < \infty,$$

و از این رو برای $h^*(\cdot)$ در گزاره ۳.۳.۳، داریم

$$E_t^c(h^*(Z_{ti})) = \Phi h^* + \sum_{j=1}^{s-1} (\Phi U^j h^*) E_t^c \left[\frac{\partial^j f_t / \partial z_{ti}^j}{f_t}(Z_t) \right] + E_t^c \left\{ [U^s h^*(Z_{ti})] \frac{\partial^s f_t / \partial z_{ti}^s}{f_t}(Z_t) \right\}. \quad (23.3)$$

۲.۳.۳ بسط اچورث توزیع پسین بر اساس چندجمله‌های هرمیتی و گشتاورها

در ادامه بسط اچورث توزیع پسین کناری پارامترها را بر اساس چندجمله‌های هرمیتی و گشتاورهای توزیع‌های کناری، نمایش می‌دهیم. دو گزاره بعد، بسط پسین (۲۳.۳) را با چندجمله‌ای‌های هرمیتی $q_k(\cdot)$ و گشتاورهای Z_{ti} مرتبط می‌کند.

گزاره ۵.۳.۳. فرض کنید $h^* : \mathbb{R} \rightarrow \mathbb{R}$ تابعی اندازه‌پذیر باشد. در این صورت، برای $k = 1, 2, \dots$

$$\Phi(U^k h^*) = \frac{1}{k!} \int_{\mathbb{R}} q_k(z) h^*(z) \Phi(dz). \quad (24.3)$$

برهان. اثبات به روش استقرا انجام می‌شود. برای $k = 1, 2$ ، (۲۴.۳) دقیقاً همان به ترتیب (۵.۳) و (۶.۳)

هستند. اکنون فرض کنید (۲۴.۳) برای $k = 1, \dots, n-1$ برقرار باشد. در گزاره ۳.۳.۳، قرار دهید

$f(z) = z_i^n$ و $s = n+1$. توجه داشته باشید با انتخاب این f ، (۱۱.۳) برقرار خواهد بود و $\frac{\partial^{n+1} f}{\partial z_i^{n+1}} = 0$

با این f و با توجه به فرض استقرا، (۱۲.۳) را می‌توان به صورت زیر بازنویسی کرد:

$$\Phi U^n h^* = \frac{1}{n!} \int_{\mathbb{R}} \left[z^n - \sum_{i=1}^{n-1} C_i^n q_i(z) E(Z^{n-i}) \right] h^*(z) d\Phi(z), \quad (25.3)$$

که در آن Z متغیر تصادفی نرمال استاندارد می‌باشد. بنا بر گزاره ۳.۳.۱، سمت راست تساوی (۲۵.۳)

□

می‌باشد. بنابراین حکم استقرا برقرار است.

گزاره ۶.۳.۳. فرض کنید $E_t^c(Z_{ti}^k) < \infty$. بنابراین

$$E_t^c \left(\frac{\partial^k f_t / \partial z_{ti}^k}{f_t}(Z_t) \right) = E_t^c(q_k(Z_{ti})). \quad (26.3)$$

برهان. ابتدا در (۲۳.۳)، $h^*(z) = q_k(z)$ و $s = k$ را اختیار می‌کنیم. بنابراین $\Phi h^* = 0$ ، $U^k h^*(z) = 1$

و

$$E_t^c(q_k(Z_{ti})) = \sum_{j=1}^{k-1} (\Phi U^j h^*) E_t^c \left(\frac{\partial^j f_t / \partial Z_{ti}^j}{f_t}(Z_t) \right) + E_t^c \left(\frac{\partial^k f_t / \partial Z_{ti}^k}{f_t}(Z_t) \right)$$

که بنا به گزاره ۵.۳.۳ و ویژگی متعامد بودن (۴.۱)، برای $j \neq k$ داریم

$$\Phi U^j h^* = \frac{1}{j!} \int q_j(z) h^*(z) d\Phi = \frac{1}{j!} \int q_j(z) q_k(z) d\Phi = 0.$$

بنابراین تساوی (۲۶.۳) نتیجه می‌شود. \square

توجه کنید که برای $k = 1, 2$ ، گزاره ۶.۳.۳ نسخه‌های تک بعدی (۱۲.۲) و (۷.۳) را نتیجه می‌دهد.

نتیجه ۷.۳.۳. اگر در (۲۳.۳)، h^* را به صورت تابع نشانگر $1(z_{ti} < \omega)$ ، $\omega \in \mathbb{R}$ ، تعریف کنیم، آنگاه از

گزاره‌های ۵.۳.۳ و ۶.۳.۳ و رابطه

$$\int_{-\infty}^{\omega} q_k(z) \phi(z) dz = -q_{k-1}(\omega) \phi(\omega), \quad (27.3)$$

تابع چگالی پسین کناری Z_{ti} به صورت زیر نتیجه می‌شود:

$$\zeta_t(z_i) = \sum_{k=0}^{\infty} c_k q_k(z_i) \phi(z_i), \quad (28.3)$$

که در آن

$$c_k = \frac{1}{k!} \int_{-\infty}^{\infty} \zeta_t(z_i) q_k(z_i) dz_i = \frac{1}{k!} E_t^c(q_k(Z_{ti})).$$

۴.۳ خطای تقریب مبتنی بر بسط اچوورث

برای ارزیابی خطای حاصل از تقریب توزیع پسین توسط بسط اچوورث ارائه‌شده، در این بخش، قضیه‌ای را مطرح می‌کنیم که سرعت همگرایی را بر حسب t بیان می‌کند. قبل از بیان قضیه، مرتبه‌های عبارت‌های (۲۳.۳) را به تفکیک به دست می‌آوریم.

با توجه به روابط (۲۰.۳) تا (۲۲.۳)، و شرط‌های (A۱) تا (A۳) در (۱۶.۳) عبارات مرتبط با $|u_t^{(i)}|^{r_i}$ ،

در $i \geq 3$ ، $E_t^c[G_k(\theta)]$ دارای مرتبه $O(t^{ri})$ هستند؛ در حالی که $|u_t^{(1)}|^{r_1}$ دارای مرتبه $O(1)$ و $|u_t^{(2)}|^{r_2}$ دارای

مرتبه $O(t^{\frac{r}{3}})$ هستند. به عنوان مثال، با استفاده از روابط (۲۰.۳) تا (۲۲.۳)، داریم

$$\begin{aligned} E_t^c \left\{ [u_t^{(1)}]^2 [u_t^{(3)}]^3 \left(\frac{\xi^{(1)}}{\xi} \right)^2 \right\} &= E_t^c \left\{ [\frac{1}{\sqrt{t}} \hat{\ell}_t^{(3)}(\eta_t) \delta_t]^2 [\hat{\ell}_t^{(3)}]^3 \left(\frac{\xi^{(1)}}{\xi} \right)^2 \right\} \\ &\leq Ct^3 E_t^c \left(\frac{\xi^{(1)}}{\xi} \right)^2 \\ &= O(t^3), \end{aligned}$$

به طوری که نامساوی خط دوم از شرطهای (A۱) و (A۲) نتیجه می شود و خط آخر از شرطهای (A۱) و

(A۳). همچنین با توجه به محدودیت (۱۷.۳)، بالاترین مرتبه G_k ، $\lfloor \frac{k}{3} \rfloor$ می باشد، که بزرگترین عدد صحیحی

است که از $\frac{k}{3}$ بزرگتر نیست (ونگ، ۲۰۱۰). علاوه بر این، اگر $-\nabla^2 \hat{\ell}_t = O(t)$ ، آن گاه چنانچه $k \in J_i$

داریم

$$E_t^c \left(\frac{d^k f_t(z)/dz^k}{f_t} \right) = E_t^c \left[\left(\frac{d\theta}{dz} \right)^k G_k(\theta) \right] = O(t^{-\frac{k}{3} + \lfloor \frac{k}{3} \rfloor}) = O(t^{-\frac{i}{3}}), \quad (29.3)$$

به طوری که $J_1 = \{1, 3\}$ و برای $i > 1$ ، $J_i = \{3i - 4, 3i - 2, 3i\}$. به عنوان نمونه

$$J_2 = \{2, 4, 6\}, \quad J_3 = \{5, 7, 9\}, \quad J_4 = \{8, 10, 12\}.$$

بنابراین اگر $h \in H_r$ و $h(z) = h^*(z_p)$ ، آن گاه بنا بر لم ۳.۲.۲ نتیجه می شود اگر $r > s$ ، $U^s h^* \in H_{r-s}$ و

اگر $r \leq s$ ، $U^s h^* \in H$. بنابراین

$$\sup_{h \in H_r} \left| E_t^c \left\{ [U^s h^*(Z_{ti})] \frac{\partial^s f_t / \partial z_{ti}^s}{f_t}(Z_t) \right\} \right| = O(t^{-\frac{s}{3} + \lfloor \frac{s}{3} \rfloor}).$$

اکنون با توجه به این استدلالها، می توان قضیه زیر را مطرح کرد.

قضیه ۱.۴.۳. (ونگ، ۲۰۱۰). فرض کنید $\xi(\theta)$ و $\ell_t(\theta)$ (۳s + ۱) بار مشتق پذیر باشند، شرطهای (A۱)

تا (A۳) برقرار باشند و $-\nabla^2 \hat{\ell}_t = O(t)$. در این صورت

$$\sup_{h \in H_r} \left| E_t^c(h^*(Z_{ti})) - \Phi h^* - \sum_{\substack{k \in \{1, \dots, 3s\} \\ k \neq 3s-1}} (\Phi U^k h^*) E_t^c \left[\frac{\partial^k f_t / \partial z_{ti}^k}{f_t}(Z_t) \right] \right| = O(t^{-\frac{r_{3s+1}}{3} + s}). \quad (30.3)$$

نتیجه ۲.۴.۳. بنا بر گزاره ۶.۳.۳ و (۲۹.۳)، برای $k \in J_j$

$$E_t^c(q_k(Z_{ti})) = O(t^{-\frac{j}{3}}). \quad (31.3)$$

حال فرض کنید Σ_t در (۵.۲) به روش تجزیه چولسکی (جنتل، ۱۹۹۸) به دست آمده باشد. در این

صورت Σ_t یک ماتریس بالا مثلثی است و در نتیجه Z_{tp} صورت ساده‌تر زیر را دارد:

$$Z_{tp} = [\Sigma_t]_{pp}(\theta_p - \hat{\theta}_{tp}). \quad (۳۲.۳)$$

نتیجه ۳.۴.۳. فرض کنید Σ_t در (۵.۲) بالا مثلثی باشد، به طوری که Z_{tp} به صورت (۳۲.۳) نتیجه شود.

اگر $h^*(.)$ را در (۲۳.۳) به صورت تابع نشانگر $\mathbb{1}(z_{tp} \leq \omega)$ ، $\omega \in \mathfrak{R}$ ، تعریف کنیم، آن‌گاه توزیع پسین

کناری پارامترتکی θ_p به صورت $P_t^c(\theta_p \leq a) = P_t^c(Z_{tp} \leq \omega)$ است و

$$\sup_{\omega \in \mathfrak{R}} \left| P_t^c(Z_{tp} \leq \omega) - \Phi(\omega) - \sum_{\substack{i \in \{1, \dots, \mathfrak{r}s\} \\ i \neq \mathfrak{r}s-1}} \frac{1}{i!} q_{i-1}(\omega) \phi(\omega) E_t^c(q_i(Z_{tp})) \right| = O(t^{-\frac{\mathfrak{r}s+1}{\mathfrak{r}}+s}), \quad (۳۳.۳)$$

که در آن $\omega = [\Sigma_t]_{pp}(a - \hat{\theta}_{tp})$. علاوه بر این، تابع چگالی پسین کناری برای θ_p به صورت زیر می‌باشد:

$$\xi_t^p(a) = [\Sigma_t]_{pp} \left\{ \phi(\omega) + \sum_{\substack{i \in \{1, \dots, \mathfrak{r}s\} \\ i \neq \mathfrak{r}s-1}} \frac{1}{i!} q_i(\omega) \phi(\omega) E_t^c(q_i(Z_{tp})) + O(t^{-\frac{\mathfrak{r}s+1}{\mathfrak{r}}+s}) \right\}. \quad (۳۴.۳)$$

برهان. از معادله (۳۰.۳)، دو گزاره ۵.۳.۳ و ۶.۳.۳ و رابطه (۲۷.۳)، داریم

$$\begin{aligned} & \sup_{h \in H_r} \left| E_t^c(h^*(Z_{ti})) - \Phi h^* - \sum_{\substack{k \in \{1, \dots, \mathfrak{r}s\} \\ k \neq \mathfrak{r}s-1}} (\Phi U^k h^*) E_t^c \left[\frac{\partial^k f_t / \partial z_{ti}^k}{f_t}(Z_t) \right] \right| \\ &= \sup_{\omega \in \mathfrak{R}} \left| P_t^c(Z_{tp} \leq \omega) - \Phi(\omega) - \sum_{\substack{k \in \{1, \dots, \mathfrak{r}s\} \\ k \neq \mathfrak{r}s-1}} \frac{1}{k!} \int_{\mathfrak{R}} q_k(z) h^*(z) \Phi(dz) E_t^c(q_k(Z_{tp})) \right| \\ &= \sup_{\omega \in \mathfrak{R}} \left| P_t^c(Z_{tp} \leq \omega) - \Phi(\omega) - \sum_{\substack{k \in \{1, \dots, \mathfrak{r}s\} \\ k \neq \mathfrak{r}s-1}} \frac{1}{k!} \int_{-\infty}^{\omega} q_k(z) \Phi(dz) E_t^c(q_k(Z_{tp})) \right| \\ &= \sup_{\omega \in \mathfrak{R}} \left| P_t^c(Z_{tp} \leq \omega) - \Phi(\omega) - \sum_{\substack{k \in \{1, \dots, \mathfrak{r}s\} \\ k \neq \mathfrak{r}s-1}} \frac{1}{k!} q_{k-1}(\omega) \phi(\omega) E_t^c(q_k(Z_{tp})) \right| \\ &= O(t^{-\frac{\mathfrak{r}s+1}{\mathfrak{r}}+s}). \end{aligned}$$

معادله (۳۴.۳) با مشتق‌گیری از (۳۳.۳) نسبت به a یکی از ویژگی‌های چندجمله‌ای‌های هرمیتی، یعنی

$$\left(\frac{d}{d\omega}\right)[q_{i-1}(\omega)\phi(\omega)] = -q_i(\omega)\phi(\omega),$$

به دست می‌آید. برای درک تساوی آخر، با استفاده از (۱.۱) و (۲.۱) و (۳.۱)، می‌توان نوشت

$$\begin{aligned} \left(\frac{d}{d\omega}\right)[q_{i-1}(\omega)\phi(\omega)] &= q'_{i-1}(\omega)\phi(\omega) + q_{i-1}(\omega)\left(\frac{d\phi(\omega)}{d\omega}\right) \\ &= (i-1)q_{i-2}(\omega)\phi(\omega) - q_{i-1}(\omega)q_1(\omega)\phi(\omega) \\ &= [(i-1)q_{i-2} - \omega q_{i-1}]\phi(\omega) \\ &= -q_i(\omega)\phi(\omega). \end{aligned}$$

□

رابطه (۳۳.۳) را می‌توان به صورت زیر بازنویسی کرد:

$$P_t^c(Z_{tp} \leq \omega) = \Phi(\omega) + \sum_{i=1}^m R_i(\omega)\phi(\omega) + O(t^{-\frac{m+1}{r}}), \quad (35.3)$$

که در آن، بنا به (۳۱.۳)،

$$R_i(\omega) = \sum_{j \in J_i} \frac{1}{j!} q_{j-1}(\omega)\phi(\omega) E_t^c(q_j(Z_{tp})) = O(t^{-\frac{i}{r}}).$$

علاوه بر این، تابع $R_i(\cdot)$ یک چندجمله‌ای از درجه حداکثر $i-1$ می‌باشد و برحسب زوج یا فرد بودن i ، تابعی فرد یا زوج می‌باشد. همچنین، ضرایب این چندجمله‌ای به گشتاورهای Z_{tp} وابسته است. بنابراین (۳۵.۳) نیز ویژگی‌های بسط اچورث را دارد. به همین دلیل ونگ (۲۰۱۰) آن را یک بسط اچورث بیزی نامید.

به‌طور مشابه، (۳۴.۳) را می‌توان به صورت زیر بازنویسی کرد:

$$\xi_t^p(a) = [\Sigma_t]_{pp} \left\{ \phi(\omega) + \sum_{i=1}^m Q_i(\omega)\phi(\omega) + O(t^{-\frac{m+1}{r}}) \right\}, \quad (36.3)$$

که در آن

$$Q_i(\omega) = \sum_{j \in J_i} \frac{1}{j!} q_j(\omega)\phi(\omega) E_t^c(q_j(Z_{tp})) = O(t^{-\frac{i}{r}}).$$

به ویژه اگر $z = 2$ ، تقریب‌های (۳۵.۳) و (۳۶.۳) دارای خطایی هم مرتبه با $O(t^{-3/2})$ هستند که معمولاً به چنین تقریب‌هایی در متون آمار مجانبی (ون در وارت، ۱۹۹۸) تقریب‌های مرتبه دوم می‌گویند.

۵.۳ بسط توزیع پسین جانسون

در راستای بسط توزیع پسین، جانسون در سال ۱۹۶۷ با بررسی‌های گسترده‌ای، یک بسط مجانبی برای خانواده بزرگی از توزیع‌های پسین ارایه داد. پس از آن، جانسون (۱۹۷۰) نشان داد که توزیع پسینی که مرکزی و مقیاس‌بندی شده است، دارای بسطی مجانبی بر حسب توان‌های $t^{-1/4}$ (که در آن t حجم نمونه است) می‌باشد، که اولین جمله آن توزیع نرمال استاندارد است. فرض کنید ψ متغیر مرکزی و مقیاس‌بندی شده باشد که به صورت زیر تعریف می‌شود:

$$\psi = (\theta - \hat{\theta}_t)b(\hat{\theta}_t),$$

به طوری که

$$b(\hat{\theta}_t) = \left[-\frac{1}{t} \sum_{i=1}^t \frac{\partial^2}{\partial \theta^2} \log f(x_i, \theta) \Big|_{\theta=\hat{\theta}_t} \right]^{1/2}. \quad (37.3)$$

فرض کنید F_t تابع توزیع تجمعی پسین $t^{1/4}\psi$ باشد. جانسون (۱۹۷۰) (در قضیه ۲۰.۱ خود) برای توزیع پسین F_t ، بسطی را به صورت زیر ارایه کرد:

$$\left| F_t(\omega) - \Phi(\omega) - \sum_{j=1}^k \gamma_j(\omega, x) t^{-j/4} \right| \leq D_1 t^{-1/4(k+1)}. \quad (38.3)$$

همچنین نتیجه ۲۰.۱ که وی ارایه داد، نشان داد توابع $\gamma_j(\omega, x)$ چندجمله‌ای‌هایی بر حسب ω هستند که ضرایب آن‌ها بر حسب x کراندارند و در تابع چگالی نرمال استاندارد ضرب شده‌اند. به ویژه، توابع γ_1 و γ_2 (که در بخش ۲۰.۴ مقاله وی، صفحه ۸۵۸ ارایه شده‌اند) به صورت زیر هستند:

$$\gamma_1(\omega, x) = -\phi(\omega) c_{\cdot 1}^{-1} [c_{1 \cdot} (\omega^2 + 2) + c_{\cdot 1}], \quad (39.3)$$

$$\gamma_2(\omega, x) = -\phi(\omega) c_{\cdot 1}^{-1} [c_{2 \cdot} \omega^5 + (5c_{2 \cdot} + c_{11}) \omega^3 + (15c_{2 \cdot} + 3c_{11} + c_{\cdot 2}) \omega], \quad (40.3)$$

به طوری که ضرایب c_{lm} , $l, m = 0, 1, 2$ ، را می توان بر حسب پیشین ξ و درستنمایی به همراه مشتقاتشان $(\xi^{(1)}, a_{3t}, a_{4t})$ بیان کرد. یعنی

$$\begin{aligned} c_{00} &= \xi(\hat{\theta}), & c_{01} &= b^{-1} \xi^{(1)}(\hat{\theta}_t), & c_{02} &= b^{-2} \xi^{(2)}(\hat{\theta}_t) \\ c_{10} &= b^{-3} a_{3t}(\hat{\theta}_t) \xi(\hat{\theta}_t), & c_{11} &= b^{-4} a_{4t}(\hat{\theta}_t) \xi(\hat{\theta}_t) + b^{-4} a_{3t}(\hat{\theta}_t) \xi^{(1)}(\hat{\theta}_t) \\ c_{20} &= 2^{-1} b^{-6} a_{3t}^2(\hat{\theta}_t) \xi(\hat{\theta}_t), \end{aligned}$$

که در آن b از رابطه (۳۷.۳) به دست می آید و a_{kt} به صورت زیر تعریف می شوند:

$$a_{kt}(\theta) = t^{-1} \sum_{i=1}^t \frac{\partial^k}{\partial \theta^k} \log f(x_i, \theta) / k!, \quad k = 3, 4, \dots$$

از آنجایی که کمیت نرمال شده Z_t در (۶.۲)، نسخه چندمتغیره ψ در (۳۷.۳) است، مقایسه بسط جانسون با بسط معرفی شده در این فصل می تواند جالب توجه باشد. ابتدا شباهت ها را فهرست می کنیم:

۱. عبارات با $i = 1, 3$ در (۳۳.۳) از مرتبه $t^{-\frac{1}{2}}$ بوده و با چندجمله ای های هرمیتی q_0 و q_2 متناظر هستند؛ که با درجه چندجمله ای ها در γ_1 (۳۹.۳) معادل می باشند.

۲. عبارات با $i = 2, 4, 6$ در (۳۳.۳) از مرتبه t^{-1} بوده و با چندجمله ای های هرمیتی q_1, q_3 و q_5 متناظر می باشند؛ که با درجه چندجمله ای ها در γ_2 (۴۰.۳) مطابقت دارند.

۳. در واقع در (۳۳.۳)، اگر گشتاورهای پسین را با تقریب های گشتاوری مجانبی با مرتبه های مناسب جایگزین کنیم، بسط جانسون نتیجه خواهد شد.

تفاوت اصلی بین این دو بسط این است که بسط معرفی شده در این فصل بر حسب گشتاورها است، در حالی که بسط جانسون بر حسب توزیع پیشین، درستنمایی و مشتقات آن ها نوشته می شود. چنین تفاوتی به دلیل استفاده از رهیافت های متفاوت می باشد: بسط جانسون مبتنی بر بسط تیلور به دست می آید، در حالی که بسط معرفی شده بر مبنای اتحاد استاین نتیجه می شود. و البته باید توجه داشت که عبارات با مرتبه های بالاتر برای γ_j ها ممکن است پیچیده باشند.

فصل ۴

مطالعه دقت بسط اچورث بیزی

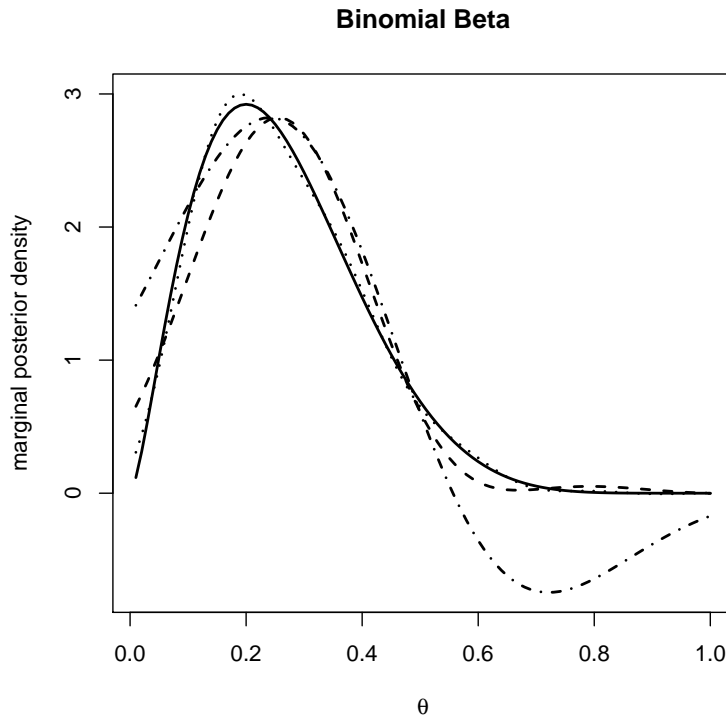
در این فصل، بر اساس دو مثال شبیه‌سازی، عملکرد بسط اچورث تابع چگالی پسین کناری (۳۴.۳) را ارزیابی می‌کنیم. با توجه به نتایج این مطالعه شبیه‌سازی، نشان می‌دهیم که بسط پیشنهادی قادر است تقریب‌های خوبی را حتی برای توزیع‌های پسینی که چوله هستند یا تک‌مدی نیستند، ارائه دهد. همچنین نشان می‌دهیم، مثال ۱.۴، تابع چگالی پسین را می‌توان با تعداد کمی از گشتاورها بازسازی کرد و بسط ارائه‌شده در موقعیت‌های کاربردی، پیچیدگی‌های زیادی وارد مساله نمی‌کند.

هر دو مثال در محیط نرم‌افزار R (تیم مرکزی توسعه R، ۲۰۱۳) اجرا شده‌اند و کدهای مربوط برای بازتولید نتایج مثال‌ها در پیوست ب موجود هستند.

۱.۴ مدل دوجمله‌ای

برای مثال اول، یک مدل دوجمله‌ای $X \sim Bin(t, \theta)$ را در نظر گرفتیم که توزیع پیشین پارامتر θ ، $Beta(a, b)$ است. در این مدل بیزی، توزیع پسین $Beta(a+x, b+t-x)$ است. در شبیه‌سازی، فرض کردیم $a = 0.5$ ، $b = 4$ ، $t = 5$ و $x = 2$ است. با این انتخاب، به‌سادگی می‌توان فهمید که حجم نمونه کوچک است و توزیع پسین θ یعنی $Beta(2/5, 7)$ چوله می‌باشد (شکل ۱.۴ را ببینید).

شکل ۱.۴ چگالی پسین واقعی θ و تقریب‌های حاصل از (۳۴.۳) را با $s = 2$ ، 13 نشان می‌دهد. دقت داشته باشید که مرتبه تقریب متناظر با $s = 2$ ، $O(t^{-2})$ است و مرتبه تقریب متناظر با $s = 13$ ، $O(t^{-13})$



شکل ۱.۴: تابع چگالی پسین کناری θ در مدل دوجمله‌ای؛ خط ممتد: توزیع دقیق، خط چین: تقریب با مرتبه $O(t^{-\frac{3}{4}})$ ، نقطه چین: تقریب با مرتبه $O(t^{-\frac{1}{2}})$ ، نقطه خط: تقریب جانسون با مرتبه $O(t^{-1})$ می‌باشد. در این مثال، گشتاورهای لازم برای محاسبه (۳۴.۳) با انتگرال‌گیری عددی به دست آمده‌اند. همانطور که در شکل ۱.۴ مشهود است، تقریب متناظر با مرتبه $O(t^{-\frac{3}{4}})$ رضایت‌بخش نیست. در حالی که، تقریب متناظر با مرتبه $O(t^{-\frac{1}{2}})$ خیلی دقیق است و اختلاف بسیار جزئی با توزیع واقعی دارد.

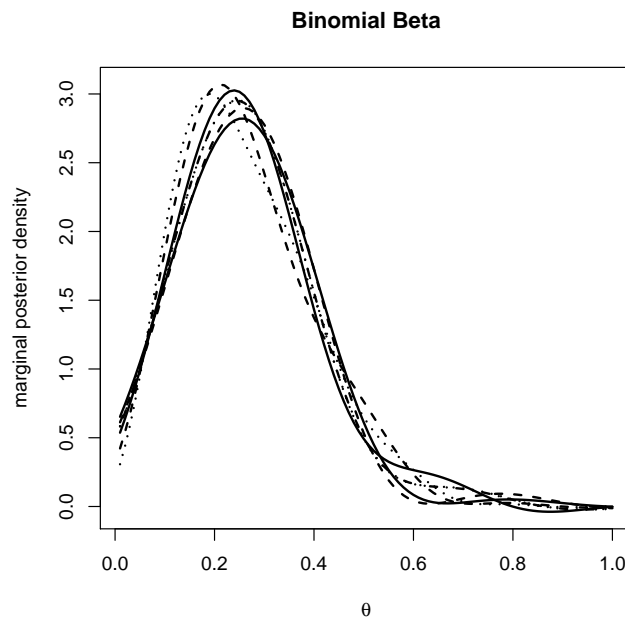
عملکرد تقریب حاصل از بسط جانسون با مرتبه $O(t^{-1})$ را نیز در این مثال، مورد بررسی قرار داده‌ایم

که با در نظر گرفتن $K = 1$ در (۳۸.۳) به دست آمده است. یعنی

$$p_t(\omega) \equiv \frac{dF_t(\omega)}{d\omega} = \phi(\omega) + \frac{d\gamma_j(\omega, x)}{d\omega} t^{-\frac{1}{4}} + O(t^{-1}).$$

در شکل می‌توانید مقادیر منفی حول $\theta = 0.7$ را برای تقریب تابع چگالی پسین ببینید. دلیل این پیامد می‌تواند به خاطر کوچک بودن اندازه نمونه باشد.

شکل ۲.۴ تقریب تابع چگالی پسین θ را با استفاده از (۳۴.۳) و با ۴۰، ۲۰، ۱۰، ۹، ۸، ۷، ۶، ۵، و با



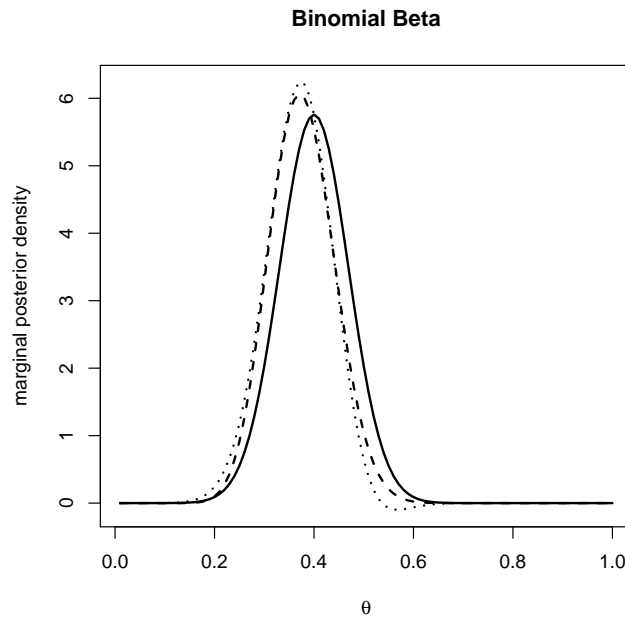
شکل ۲.۴: تابع چگالی پسین کناری θ در مدل دو جمله‌ای؛ نقطه-خط تیره: ۵ گشتاور، خط ممتد: ۶، ۷، ۸، ۹، ۱۰ گشتاور، خط تیره: ۲۰ گشتاور، نقطه چین: ۴۰ گشتاور

گشتاور Z_{tp} نشان می‌دهد. همان طور که انتظار می‌رفت، وقتی گشتاورهای بیشتری به کار گرفته می‌شوند، منحنی‌ها به چگالی پسین واقعی نزدیک‌تر می‌شوند.

برای بررسی بیشتر تقریب بسط جانسون، و درک بهتر تاثیر حجم نمونه، شبیه‌سازی را با حجم نمونه $t = 50$ ، برای $x = 20$ و ثابت نگه داشتن سایر شرایط، اجرا کردیم. شکل ۳.۴ منحنی‌های چگالی دقیق، تقریب نرمال و تقریب جانسون با $O(t^{-1})$ را نمایش می‌دهد. می‌توان مشاهده کرد که تقریب جانسون اصلاح شده است و تقریباً عملکردی مشابه با تقریب نرمال دارد.

۲.۴ مدل نرمال دو متغیره

در این پایان‌نامه $\hat{\theta}_t$ را به عنوان برآورد درست‌نمایی ماکسیمم تعریف کردیم. ولی شرط یکتایی MLE در شرایط نظم مورد نظر ما وجود ندارد. بنابراین مشتاق شدیم تا عملکرد (۳۴.۳) را در حالتی که MLE یکتا نیست و در واقع برای یک پارامتر چندین برآورد درست‌نمایی ماکسیمم وجود دارد، بررسی کنیم. برای این



شکل ۳.۴: تابع چگالی پسین کناری θ در مدل دوجمله‌ای؛ خط ممتد: توزیع دقیق، خط تیره: تقریب نرمال، نقطه چین: تقریب جانسون با $O(t^{-1})$

منظور، توزیع پسین ضریب همبستگی داده‌هایی از توزیع نرمال دومتغیره را که توسط موری (۱۹۷۷) معرفی شده است، در نظر گرفتیم. این توزیع پسین توسط تانر و ونگ (۱۹۸۷) نیز مورد توجه قرار گرفته است. مجموعه داده‌ها در جدول ۱.۴ آورده شده‌اند. این مجموعه داده شامل ۱۲ مشاهده است به طوری که فرض شده است دارای توزیع نرمال دومتغیره با $\mu_1 = \mu_2 = 0$ و ضریب همبستگی ρ و واریانس‌های σ_1^2 و σ_2^2 هستند. در این داده‌ها ۲ جفت داده، همبستگی ۱ و ۲ جفت همبستگی -1 دارند و ضریب همبستگی ۸ جفت دیگر گم شده هستند. اگر ماتریس کواریانس را با Γ نشان دهیم، مشابه تانر و ونگ (۱۹۸۷) فرض می‌کنیم توزیع پیشین Γ عبارتست از

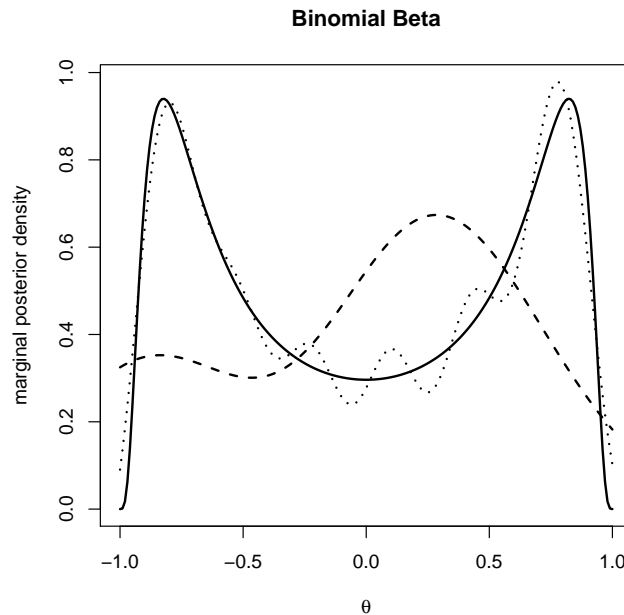
$$\xi(\Gamma) \propto |\Gamma|^{-\frac{(k+1)}{2}},$$

که در آن $k = 2$ بعد توزیع نرمال است.

برای پارامترهای $\theta = (\sigma_1^2, \sigma_2^2, \rho)$ ، دو برآورد ML وجود دارد: $\hat{\theta} = (2/67, 2/67, -0/5)$ و $\hat{\theta} = (2/67, 2/67, 0/5)$. از برآورد اولی به عنوان $\hat{\theta}_t$ در عبارت Z_t تعریف شده در (۶.۲) استفاده

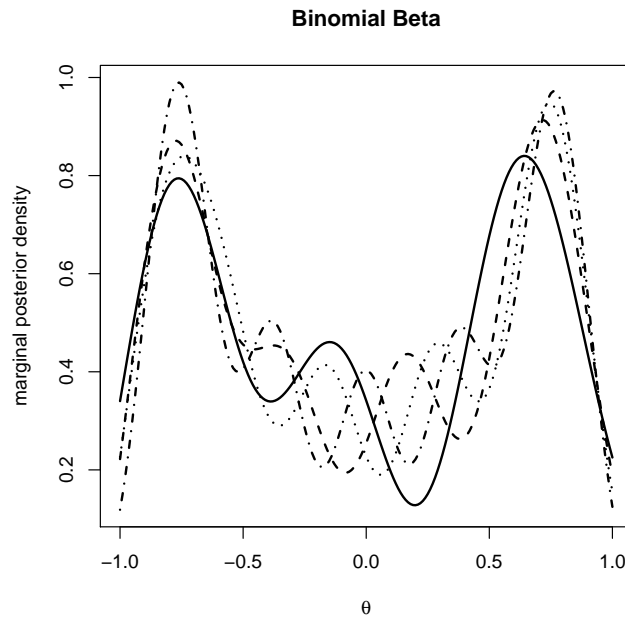
جدول ۱.۴: داده‌های توزیع نرمال دومتغیره (* نمایانگر مقادیری است که مشاهده نشده‌اند)

۱	۱	-۱	-۱	۲	۲	-۲	-۲	*	*	*	*
۱	-۱	۱	-۱	*	*	*	*	۲	۲	-۲	-۲



شکل ۴.۴: تابع چگالی پسین کناری ρ ؛ خط ممتد: توزیع دقیق، نقطه چین: تقریب با مرتبه $O(t^{-\frac{3}{2}})$ ، خط چین: تقریب با استفاده از ۱۰۰ گشتاور

می‌کنیم. در شکل ۴.۴ چگالی‌های پسین برآورد شده ρ با استفاده از (۳۴.۳) با اختیار $s = 2$ و $s = 33$ رسم شده‌اند. تقریب حاصل از $s = 33$ با تقریباً ۱۰۰ گشتاور اول Z_{tp} که با انتگرال‌گیری عددی برآورد شده‌اند، متناظر است. در این شکل منحنی تابع چگالی پسین واقعی ρ که متناسب است با $\frac{(1-\rho^2)^{4.5}}{(1.25-\rho^2)^8}$ (تانر و ونگ، ۱۹۸۷)، نیز رسم شده است. نتایج نشان می‌دهد که عملکرد تقریب حاصل از بسط با استفاده از ده‌ها گشتاور، حول دو مد کاملاً راضی‌کننده است. البته باید اشاره کرد که این تقریب در زیربازه‌ای از فضای پارامتر که دور از مدها است، دارای نوسان است. اما تقریب با مرتبه $O(t^{-\frac{3}{2}})$ با خطای زیادی همراه است. شکل ۵.۴ تقریب‌های حاصل از (۳۴.۳) با ۲۰، ۴۰، ۶۰ و ۸۰ گشتاور Z_{tp} را نمایش می‌دهد. همان‌طور که مشاهده می‌کنید نوسانات با افزایش تعداد گشتاورهای حاضر در تقریب، کاهش می‌یابد.



شکل ۵.۴: تابع چگالی پسین کناری ρ ؛ خط ممتد: ۲۰ گشتاور، نقطه چین: ۴۰ گشتاور، نقطه-خط تیره: ۶۰ گشتاور، خط تیره: ۸۰ گشتاور

۳.۴ نتیجه‌گیری و پیشنهادات برای آینده تحقیق

در این پایان‌نامه، با کمک اتحاد استاین، نشان دادیم توزیع مجانبی پسین توام پارامترهای یک مدل بیزی، نرمال است (فصل دوم). سپس برای توزیع‌های پسین کناری پارامترها یک بسط اچوورث را نمایش دادیم (فصل سوم). همچنین با دو مثال شبیه‌سازی (فصل چهارم)، نشان دادیم استفاده از بسط اچوورث پیشنهادی و انتگرال‌گیری عددی (برای گشتاورهای تبدیل Z_{tp})، حتی زمانی که حجم نمونه کوچک یا توزیع چندمدی است، تقریب‌های خوبی را نتیجه می‌دهد.

با توجه به مطالب نظری که در این پایان‌نامه مطرح شدند، می‌توان چند موضوع جالب مرتبط را به عنوان

آینده تحقیق بر شمرد:

الف) تبدیل Z_t که در این پایان‌نامه معرفی شد را می‌توان به صورت‌های دیگری تعریف کرد. به عنوان

مثال، مشابه تبدیل ریشه علامت‌دار^۱ بیکل و گوش (۱۹۹۰). با تعریف این تبدیل، تحت شرایط نظم مناسب، نمایشی که برای امید ریاضی پسین در (۲۳.۳) ارایه شد، برقرار خواهد ماند (ونگ، ۲۰۱۰). بنابراین بر اساس دو گزاره ۵.۳.۳ و ۶.۳.۳، می‌توان بسط اچوورث (۲۸.۳) را که بر مبنای چندجمله‌ای‌های هرمیتی است، نیز به‌دست آورد. اما با این تبدیل جدید Z_t ، تابع f_t در (۱۰.۲) متفاوت خواهد بود و در نتیجه مرتبه‌های مجانبی که برای $E_t^c[q_k(Z_{ti})]$ در فصل سوم، ارایه شدند، برقرار نیستند. بررسی مرتبه‌های مجانبی مرتبط با تعریف این نوع تبدیل‌های جدید می‌تواند به عنوان یک موضوع نظری جالب، مورد توجه علاقه‌مندان خود باشد.

ب) برآورد ناپارامتری تابع چگالی به یکی از زمینه‌های تحقیقاتی پرتعداد تبدیل شده است. اما امکان تعمیم نتایج معرفی‌شده در این پایان‌نامه برای برآوردگرهای تابع چگالی، واضح نیست و تاکنون انجام نشده است. یکی از مشکلات نظری برای چنین تعمیمی، پاسخ به این سوال است که آیا تابع چگالی پسین برآوردگر تابع چگالی را می‌توان به صورت (۱۰.۲) نوشت؟

ج) از آن‌جا که بسط پسین مبتنی بر سری تیلور قابل کاربرد برای مدل‌های بیزی با پیشین‌های ناهموار نیست، تعمیم نتایج مطرح‌شده به چنین مسایلی جالب توجه است. یک نقطه شروع ممکن برای چنین تعمیمی، تصحیح اتحاد استاین در لم ۱.۲.۳ برای توابع f_t است که به‌طور تکه‌ای هموار^۲ باشند.

د) حصول اطمینان از همگرایی روش‌های مبتنی بر نمونه‌گیری در مدل‌های بیزی، مانند الگوریتم‌های MCMC، از جمله مسایل مطرح و روز در استنباط‌های بیزی آماری است. شاید بتوان از این بسط‌های مجانبی برای بررسی همگرایی نتایج حاصل از روش‌های مذکور، استفاده کرد. ایده پایه‌ای این است که اگر نمونه تولیدشده به توزیع واقعی، که همان توزیع پسین است، همگرا شود، برآورد تابع چگالی نمونه تولیدشده باید با برآورد حاصل از (۳۴.۳) که با جایگذاری گشتاورهای تجربی نمونه به‌دست می‌آید، موافق باشد.

^۱Signed-root transformation

^۲Piecewise smooth

ه) در متون موجود، رهنمود یا رهیافت‌هایی که مشخص کنند از چند مولفه اول بسط باید در موقعیت‌های کاربردی استفاده کرد، وجود ندارند. این موضوع نیز به عنوان یک موضوع تحقیق جالب می‌تواند مطرح باشد.

و) ایده جالب دیگری که می‌توان آن را به عنوان یک موضوع تحقیق مورد بررسی قرار داد، تقریب توزیع پسین با استفاده از بسط مجانبی نقطه زینی^۳ و مقایسه آن با بسط اچوورث می‌باشد.

^۳Saddlepoint Approximation

پیوست آ

نمادها و تعاریف

در این پیوست، مفاهیمی که در پایان نامه مورد استفاده قرار گرفته‌اند به صورت تعریف و به شرح زیر فهرست کرده‌ایم. این تعاریف را از کتاب آنالیز حقیقی رودین (۱۹۸۷) و کتاب مبانی آمار ریاضی دکتر پارسیان اقتباس کرده‌ایم.

تعریف ۱.۰.۰. جفت $(\mathbb{X}, \mathcal{B})$ ، که \mathbb{X} یک مجموعه و \mathcal{B} یک σ -جبر^۱ از زیرمجموعه‌های \mathbb{X} است، فضای اندازه‌پذیر^۲ نامیده می‌شود.

تعریف ۲.۰.۰. فرض کنید $(\mathbb{X}, \mathcal{B})$ یک فضای اندازه‌پذیر باشد. تابع نامنفی $\mu(\cdot)$ که بر روی $A \in \mathcal{B}$ تعریف می‌شود، یک اندازه روی \mathcal{B} نامیده می‌شود اگر

$$1. \mu(\emptyset) = 0, \text{ که در آن } \emptyset \text{ نمایش مجموعه پوچ است.}$$

۲. اگر $\{A_n\}$ دنباله‌ای از مجموعه‌های اندازه‌پذیر دوه‌دو مجزا باشد، آن‌گاه

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

لم ۳.۰.۰. فرض کنید $(\mathbb{X}, \mathcal{B})$ یک فضای اندازه‌پذیر و $h(\cdot)$ تابعی حقیقی مقدار روی \mathbb{X} باشد. چهار حالت زیر معادلند:

^۱ Sigma algebra

^۲ Measurable

$$\{x : h(x) < \alpha\} \in \mathcal{B} . ۱$$

$$\{x : h(x) \leq \alpha\} \in \mathcal{B} . ۲$$

$$\{x : h(x) > \alpha\} \in \mathcal{B} . ۳$$

$$\{x : h(x) \geq \alpha\} \in \mathcal{B} . ۴$$

تعریف آ.۴.۰.۰. تابع حقیقی مقدار $h(\cdot)$ که در یکی از چهار حالت لم آ.۳.۰.۰ صدق می‌کند، تابع اندازه‌پذیر نامیده می‌شود.

تعریف آ.۵.۰.۰. نرم. فرض کنید X فضای برداری باشد، تابع $\|\cdot\| : X \rightarrow \mathbb{R}$ را یک نرم روی X می‌نامیم، هرگاه شرایط زیر برقرار باشند:

$$۱. \text{ به ازای هر } x \in X, \|x\| > ۰.$$

$$۲. \text{ به ازای هر } x \in X \text{ داشته باشیم } x = ۰ \text{ اگر و تنها اگر } \|x\| = ۰.$$

$$۳. \text{ به ازای هر } x \in X \text{ و هر اسکالر } \alpha \text{ داشته باشیم } \|\alpha x\| = |\alpha| \|x\|.$$

$$۴. \text{ به ازای هر } x, y \in X \text{ داشته باشیم } \|x + y\| \leq \|x\| + \|y\| \text{ (نامساوی مثلث).}$$

تعریف آ.۶.۰.۰. نرم اقلیدسی. اگر $\|\mathbf{X}\|_d$ روی بردار $\mathbf{X} = (X_1, \dots, X_p) \in \mathbb{R}^p$ به صورت

$$\|\mathbf{X}\|_d = (\mathbf{X}^T \mathbf{X})^{\frac{1}{2}} = \left(\sum_{i=1}^p X_i^2 \right)^{\frac{1}{2}},$$

تعریف شود، به آن نرم اقلیدسی اطلاق می‌شود.

تعریف آ.۷.۰.۰. نرم طیفی. $\|J\|^2 = \lambda_{\max}(J^T J)$ ، که در آن $\lambda_{\max}(A)$ بزرگترین مقدار ویژه ماتریس A

است، را نرم طیفی ماتریس J گویند.

تعریف آ.۸.۰.۰. همسایگی حول یک نقطه. فرض کنید X و θ عضو \mathbb{R}^p باشند. به ازای $\rho > ۰$ مجموعه

$$B_{\theta, \rho} = \{X; \|X - \theta\| \leq \rho\},$$

را یک همسایگی حول θ به شعاع ρ می‌نامیم. به این همسایگی یک گوی بسته نیز گفته می‌شود. اگر در این تعریف، علامت \leq به $<$ تبدیل شود، به آن یک گوی باز گفته می‌شود.

تعریف آ.۹.۰.۰. نقطه درونی مجموعه. فرض کنید $X \subseteq \mathbb{R}^p$ ، نقطه $\theta \in X$ را یک نقطه درونی مجموعه X می‌نامیم، در صورتی که گوی بازی مانند $B_{\theta, \rho}$ وجود داشته باشد به طوری که در X باشد. به عبارت دیگر

$$\forall \theta \in X; \quad \exists B_{\theta, \rho} \subseteq X.$$

تعریف آ.۱۰.۰.۰. مجموعه باز. مجموعه $X \subseteq \mathbb{R}^p$ را باز می‌نامیم، در صورتی که هر نقطه آن درونی باشد و X را بسته می‌نامیم در صورتی که $\mathbb{R}^p - X$ باز باشد.

تعریف آ.۱۱.۰.۰. بسط تیلور. برای تعریف بسط تیلور، ابتدا لازم است سری توانی را معرفی کنیم. یک سری به شکل

$$a_0 + a_1x + a_2x^2 + \dots = \sum_{n=0}^{\infty} a_n x^n,$$

که در آن a_0, a_1, \dots اعدادی ثابت هستند، یک سری توانی از x می‌باشد. در حالت کلی، سری توانی برای $|x| < R$ همگرا و برای $|x| > R$ واگراست. در اینجا R عدد ثابتی است که آن را شعاع همگرایی سری می‌نامند. برای $|x| = R$ ، سری می‌تواند همگرا یا واگرا باشد.

بسط تیلور، نمایش یک تابع به صورت مجموع بی‌نهایت جمله است که از مشتق‌های یک تابع در یک نقطه به دست می‌آیند. با استفاده از این بسط، توابع را می‌توان حول یک نقطه با تعداد متناهی از جملات تقریب زد. بسط تیلور تابع $f(x)$ با مقادیر حقیقی یا مختلط که در همسایگی نقطه حقیقی یا مختلط x_0 بی‌نهایت بار مشتق‌پذیر است، سری توانی زیر است:

$$f(x) = f(x_0) + \frac{(x - x_0)}{1!} f'(x_0) + \frac{(x - x_0)^2}{2!} f''(x_0) + \dots,$$

که می‌توانیم به صورت خلاصه‌تر زیر بنویسیم:

$$f(x) = \sum_{n=0}^{\infty} \frac{(x - x_0)^n}{n!} f^{(n)}(x_0).$$

تعریف آ.۱۲.۰.۰. بسط مک‌لورن. توزیع بسط تیلور حول صفر ($x_0 = 0$) را بسط مک‌لورن می‌نامیم.

تعریف آ.۱۳.۰. قضیه مقدار میانگین. هرگاه دامنه تابع f مجموعه $[a, b]$ باشد به طوری که بر (a, b)

مشتق پذیر باشد، در این صورت عددی چون c در (a, b) موجود است، به طوری که $f'(c) = \frac{f(b)-f(a)}{b-a}$.

تعریف آ.۱۴.۰. فرمول تبدیل فوریه معکوس. فرض کنید μ یک اندازه احتمال بوده و $\varphi(t) = \int_{\mathbb{R}} e^{itx} d\mu(x)$

اگر $x_1 < x_2$ ، آن گاه

$$\mu(x_1, x_2) + \frac{1}{4}\mu(x_1) + \frac{1}{4}\mu(x_2) = \lim_{T \rightarrow \infty} \frac{1}{4\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t) dt.$$

که در آن x_1 و x_2 متعلق به فضای برداری می باشند.

تعریف آ.۱۵.۰. ماتریس معین مثبت. ماتریس متقارن M معین مثبت است هرگاه $z^T M z$ برای هر بردار

ستونی غیر صفر z ، مثبت باشد.

تعریف آ.۱۶.۰. تجزیه چولسکی. فرض کنید X ماتریس متقارن معین مثبت باشد. تجزیه چولسکی X به

صورت $X = Q^T Q$ می باشد که در آن Q یک ماتریس بالا مثلثی است.

تعریف آ.۱۷.۰. مجموعه پوچ. فرض کنید N مجموعه اندازه پذیر در فضای اندازه پذیر X باشد. آن گاه N

پوچ است اگر اندازه آن صفر باشد.

تعریف آ.۱۸.۰. قضیه حد مرکزی. فرض کنید X_1, X_2, \dots دنباله ای از متغیرهای تصادفی مستقل و

هم توزیع با میانگین μ و انحراف معیار متناهی σ باشد. در این صورت

$$S_n = \sum_{i=1}^n X_i \xrightarrow{D} N(n\mu, n\sigma^2).$$

تعریف آ.۱۹.۰. توزیع گاما. اگر متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد

$$f_{\alpha, \lambda}(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0, \alpha > 0, \lambda > 0$$

گوئیم X دارای توزیع گاما با پارامترهای α و λ است و آن را با نماد $X \sim \Gamma(\alpha, \lambda)$ نشان می دهیم.

تعریف آ.۲۰.۰. توزیع بتا. اگر متغیر تصادفی X دارای تابع چگالی احتمال زیر باشد

$$f_{\alpha,\beta}(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 < x < 1, \quad \alpha > 0, \quad \beta > 0$$

گوییم X دارای توزیع بتا با پارامترهای α و β است و آن را با نماد $X \sim \text{Beta}(\alpha, \beta)$ نشان می‌دهیم.

تعریف آ.۲۱.۰. توزیع نرمال دو متغیره. متغیرهای تصادفی (X_1, X_2) دارای توزیع نرمال توام با پارامترهای

$(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ هستند هرگاه تابع چگالی توام آن‌ها به صورت

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\{-Q(x_1, x_2)/2\}, \quad |x_1| < \infty, \quad |x_2| < \infty,$$

باشد، که در آن $Q(x_1, x_2)$ تابعی درجه دو به صورت

$$\frac{1}{1-\rho^2} \left\{ \left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \right\}$$

و $|\mu_1| < \infty, |\mu_2| < \infty, \sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$ می‌باشند.

تعریف آ.۲۲.۰. پیوستگی مطلق. در فضای اندازه‌پذیر \mathbb{X} ، اندازه P نسبت به Q پیوسته مطلق می‌باشد، اگر

به‌ازای هر $A \in \mathbb{X}$ که $Q(A) = 0$ آن‌گاه $P(A) = 0$ و با نماد $P \ll Q$ نشان داده می‌شود.

تعریف آ.۲۳.۰. متقابلا پیوسته مطلق. P و Q را متقابلا پیوسته مطلق می‌نامند، هرگاه $P \ll Q$ و

$Q \ll P$.

پیوست ب

کدهای زبان برنامه‌نویسی R برای بازتولید نتایج شبیه‌سازی پایان‌نامه

کد مربوط به نمودار فصل اول

```
library(ttutils)

library(EQL)

n <- 3

df <- 2

x_eval_chi <- seq(0,10, by=0.05)

true <- dgamma(x=x_eval_chi, shape=df*n/2, scale=2/n)

plot(x_eval_chi, true, col=1, lty=1, type="l", xlab="x", ylab="density", lwd=2)

title(expression(paste("Density of the sample mean of three ", chi^2, "(2) variables")))

lines(x_eval_chi, dnorm(x_eval_chi, mean=df, sd=sqrt(2*df/n)), col=2, lwd=2, lty=4)

edge_2 <- edgeworth(x_eval_chi, n=n, mu=df, sigma2=2*df, rho3=sqrt(8/df),

                    rho4=12/df, type="mean", deg=2)$approx

edge_3 <- edgeworth(x_eval_chi, n=n, mu=df, sigma2=2*df, rho3=sqrt(8/df),
```

```
rho4=12/df, type="mean", deg=3)$approx
lines(x_eval_chi, edge_2, col=4, lty=3)
lines(x_eval_chi, edge_3, col=3, lty=2)
legend("topright", leg=c("True", "Edgeworth, degree 2", "Edgeworth, degree 3", "Normal"),
col=c(1,2,4,3), lty=c(2,3,2,4))
```

کدهای مربوط به نمودارهای فصل ۴

```
# Beta-Binomial
# Figures 1 and 2
k=40;
coeff=array(dim=c(k,k));
for (i in 1:k){
  for (j in 1:k){
    coeff[i,j]=0.0;
  }
}
coeff[1,1]=1;
coeff[2,1]=1; coeff[2,2]=-1;
for (i in 3:k){
  coeff[i,1]=1;
  if (floor(i/2) < i/2){
    a=(i+1)/2;
```

```
for (j in 2:a){
  coeff[i,j]=coeff[i-1,j]-coeff[i-1,j-1]*((i-1)-2*(j-2));
}
}

if (floor(i/2) == i/2){
  a=(i+2)/2;
  for (j in 2:a){
    coeff[i,j]=coeff[i-1,j]-coeff[i-1,j-1]*((i-1)-2*(j-2));
  }
}
}

# study binomial with beta prior
alpha=0.5;beta=4;n=5;x=2;
a_new=alpha+x; b_new=beta+n-x;
xbar=x/n;
lpp=function(theta){
  (-1)*(x/(theta^2)+(n-x)/((1-theta)^2));
}

lppp=function(theta){
  2*x/(theta^3)-2*(n-x)/((1-theta)^3);
}

lpppp=function(theta){
```

```

    (-1)*(6*x/(theta^4)+6*(n-x)/((1-theta)^4));
}

xi=function(theta){
    theta^(alpha-1) * (1-theta)^(beta-1);
} #up to proportional constant

xip=function(theta){
    (alpha-1)*theta^(alpha-2) * (1-theta)^(beta-1)-(beta-1)*theta^(alpha-1) * (1-theta)^(beta-2);
} #up to proportional constant

xipp=function(theta){
    (alpha-1)*(alpha-2)*theta^(alpha-3) * (1-theta)^(beta-1)-2*(alpha-1)*(beta-1)
    *theta^(alpha-2)*(1-theta)^(beta-2)+(beta-1)*(beta-2)*theta^(alpha-1) * (1-theta)^(beta-3);
} #up to proportional constant

xip2xi=function(theta){
    (alpha-1)/theta- (beta-1)/(1-theta);
}

xipp2xi=function(theta){
    (alpha-1)*(alpha-2)/(theta^2)-2*(alpha-1)*(beta-1)/(theta*(1-theta))+
    (beta-1)*(beta-2)/((1-theta)^2);
}

sigma=(n^2/x+n^2/(n-x))^(1/2); # actually sigma=(-lpp(xbar))^(1/2)

b_hat=(1/sqrt(n)) * sigma;

a3n_hat= (-1/n)*(1/(3*2))*lppp(xbar);

```

```

a4n_hat= (-1/n)*(1/(4*3*2))*lpppp(xbar);

c00=xi(xbar);

c01=xip(xbar)/b_hat;

c02=xipp(xbar)/(b_hat^2);

c10=a3n_hat*xi(xbar)/(b_hat^3);

c11=a4n_hat*xi(xbar)/(b_hat^4)+a3n_hat*xip(xbar)/(b_hat^4);

c20=(1/2)*(1/b_hat^6)*(a3n_hat^2)*xi(xbar);

c10toc00=a3n_hat/(b_hat^3);

c01toc00=xip2xi(xbar)/b_hat;

c20toc00=(1/2)*(1/b_hat^6)*(a3n_hat^2);

c11toc00=a4n_hat/(b_hat^4)+a3n_hat*xip2xi(xbar)/(b_hat^4);

c02toc00=xipp2xi(xbar)/(b_hat^2);

gam1=function(w){

    (-1)*dnorm(w)*c00^(-1)*(c10*(w^2+2)+c01);

}

gam1_p=function(w){

    w*((c10toc00*(w^2+2)+c01toc00)-2*c10toc00);

}

EZ=0;

for (i in 1:k){

    Zi=function(x){ ((x-xbar)^i)*dbeta(x,a_new,b_new); }

    EZ[i]=(sigma^i)*integrate(Zi,0,1)$value;

```

```

} #first 10 moments of Z

Eq=0;

Eq[1]=EZ[1];

Eq[2]=EZ[2]-1;

for (i in 3:k){

    Eq[i]=EZ[i];

    if (floor(i/2) == i/2){

        a=(i+2)/2;

        for (j in 2:(a-1)){

            b=i-2*(j-1);

            Eq[i]=Eq[i]+coeff[i,j]*EZ[b];

        }

        Eq[i]=Eq[i]+coeff[i,a];

    }

    if (floor(i/2) < i/2){

        a=(i+1)/2;

        for (j in 2:a){

            b=i-2*(j-1);

            Eq[i]=Eq[i]+coeff[i,j]*EZ[b];

        }

    }

}

```

```
herm=function(par){  
    i=par[1];  
    w=par[2];  
  
    Hermite=w^i;  
    if (floor(i/2) == i/2){  
        for (j in 2:(i/2)){  
            b=i-2*(j-1);  
            Hermite=Hermite+coeff[i,j]*w^b;  
        }  
        Hermite=Hermite+coeff[i,(i+2)/2];  
    }  
    if (floor(i/2) < i/2){  
        for (j in 2:((i+1)/2)){  
            b=i-2*(j-1);  
            Hermite=Hermite+coeff[i,j]*w^b;  
        }  
    }  
    Hermite;  
}  
  
marginal_1=function(par){  
    sig=par[1]
```

```
w=par[2]

Eq1=par[3]

Eq2=par[4]

Eq3=par[5]

Eq4=par[6]

main=1+w*Eq1 + (1/2)*(w^2-1)*Eq2;

for (i in 3:4){

    main=main+(1/factorial(i))*herm(c(i,w))*par[i+2];

}

sig*(1/(sqrt(2*pi)))*exp(-w^2/2)*main;

}

addition_1=function(kk,par){

    sig=par[1]

    w=par[2]

    main=0;

    for (i in 5:kk){

        main=main+(1/factorial(i))*herm(c(i,w))*par[i-2];

    }

    sig*(1/(sqrt(2*pi)))*exp(-w^2/2)*main;

}

xlevel=y_1=y_2=y_3=0; y_john1=0;

y_4=y_5=y_6=y_7=y_8=y_9=y_10=y_11=0;
```

```

for (s in 1:100){
  xlevel[s]=s/100;
  delta=sigma*(xlevel[s]-xbar);
  y_1[s]=dbeta(xlevel[s], a_new, b_new)
  y_2[s]=marginal_1(c(sigma,delta,Eq[1],Eq[2],Eq[3],Eq[4]));
  y_3[s]=y_2[s]+addition_1(40,c(sigma,delta,Eq[5:40]));
  y_4[s]=y_2[s]+addition_1(5,c(sigma,delta,Eq[5:5]));
  y_5[s]=y_2[s]+addition_1(6,c(sigma,delta,Eq[5:6]));
  y_6[s]=y_2[s]+addition_1(7,c(sigma,delta,Eq[5:7]));
  y_7[s]=y_2[s]+addition_1(8,c(sigma,delta,Eq[5:8]));
  y_8[s]=y_2[s]+addition_1(9,c(sigma,delta,Eq[5:9]));
  y_9[s]=y_2[s]+addition_1(10,c(sigma,delta,Eq[5:10]));
  y_10[s]=y_2[s]+addition_1(20,c(sigma,delta,Eq[5:20]));
  y_11[s]=y_2[s]+addition_1(30,c(sigma,delta,Eq[5:30]));
  y_john1[s]=sigma*dnorm(delta)*(1+(1/sqrt(n))*gam1_p(delta));
}
par(mfrow=c(2,1));
matplot(xlevel,cbind(y_1,y_5,y_3,y_john1),xlab=expression(theta),
ylab="marginal posterior density",main="Binomial Beta",lty=1:4,type="l")
matplot(xlevel,cbind(y_5,y_6,y_7,y_8,y_9,y_10,y_3),xlab=expression(theta),
ylab="marginal posterior density",main="Binomial Beta",lty=1:4,type="l")
# Beta Binomial

```


Λ\

```
# Figure 3
```

```
k=40;
```

```
coeff=array(dim=c(k,k));
```

```
for (i in 1:k){
```

```
  for (j in 1:k){
```

```
    coeff[i,j]=0.0;
```

```
  }
```

```
}
```

```
coeff[1,1]=1;
```

```
coeff[2,1]=1; coeff[2,2]=-1;
```

```
for (i in 3:k){
```

```
  coeff[i,1]=1;
```

```
  if (floor(i/2) < i/2){
```

```
    a=(i+1)/2;
```

```
    for (j in 2:a){
```

```
      coeff[i,j]=coeff[i-1,j]-coeff[i-1,j-1]*((i-1)-2*(j-2));
```

```
    }
```

```
  }
```

```
  if (floor(i/2) == i/2){
```

```
    a=(i+2)/2;
```

```
    for (j in 2:a){
```

```
      coeff[i,j]=coeff[i-1,j]-coeff[i-1,j-1]*((i-1)-2*(j-2));
```

```
    }  
  }  
}  
  
# study binomial with beta prior  
  
alpha=0.5;beta=4;n=50;x=20;  
  
a_new=alpha+x; b_new=beta+n-x;  
  
xbar=x/n;  
  
lpp=function(theta){  
  (-1)*(x/(theta^2)+(n-x)/((1-theta)^2));  
}  
  
lppp=function(theta){  
  2*x/(theta^3)-2*(n-x)/((1-theta)^3);  
}  
  
lpppp=function(theta){  
  (-1)*(6*x/(theta^4)+6*(n-x)/((1-theta)^4));  
}  
  
xi=function(theta){  
  theta^(alpha-1) * (1-theta)^(beta-1);  
}  
  
#up to proportional constant  
  
xip=function(theta){  
  
(alpha-1)*theta^(alpha-2) * (1-theta)^(beta-1)-(beta-1)*theta^(alpha-1)
```

```

    * (1-theta)^(beta-2);
}

#up to proportional constant

xipp=function(theta){

(alpha-1)*(alpha-2)*theta^(alpha-3) * (1-theta)^(beta-1)-2*(alpha-1)*(beta-1)*

theta^(alpha-2)*(1-theta)^(beta-2)+(beta-1)*(beta-2)*theta^(alpha-1)

* (1-theta)^(beta-3);

}

#up to proportional constant

xip2xi=function(theta){

(alpha-1)/theta- (beta-1)/(1-theta);

}

xipp2xi=function(theta){

(alpha-1)*(alpha-2)/(theta^2)-2*(alpha-1)*(beta-1)/(theta*(1-theta))

+(beta-1)*(beta-2)/((1-theta)^2);

}

sigma=(n^2/x+n^2/(n-x))^(1/2); # actually sigma=(-lpp(xbar))^(1/2)

b_hat=(1/sqrt(n)) * sigma;

a3n_hat= (-1/n)*(1/(3*2))*lppp(xbar);

a4n_hat= (-1/n)*(1/(4*3*2))*lpppp(xbar);

c00=xi(xbar);

c01=xip(xbar)/b_hat;

```

```

c02=xipp(xbar)/(b_hat^2);

c10=a3n_hat*xi(xbar)/(b_hat^3);

c11=a4n_hat*xi(xbar)/(b_hat^4)+a3n_hat*xip(xbar)/(b_hat^4);

c20=(1/2)*(1/b_hat^6)*(a3n_hat^2)*xi(xbar);

c10toc00=a3n_hat/(b_hat^3);

c01toc00=xip2xi(xbar)/b_hat;

c20toc00=(1/2)*(1/b_hat^6)*(a3n_hat^2);

c11toc00=a4n_hat/(b_hat^4)+a3n_hat*xip2xi(xbar)/(b_hat^4);

c02toc00=xipp2xi(xbar)/(b_hat^2);

gam1=function(w){

    (-1)*dnorm(w)*c00^(-1)*(c10*(w^2+2)+c01);

}

gam1_p=function(w){

    w*((c10toc00*(w^2+2)+c01toc00)-2*c10toc00);

}

EZ=0;

for (i in 1:k){

    Zi=function(x){ ((x-xbar)^i)*dbeta(x,a_new,b_new); }

    EZ[i]=(sigma^i)*integrate(Zi,0,1)$value;

} #first 10 moments of Z

Eq=0;

Eq[1]=EZ[1];

```

```

Eq[2]=EZ[2]-1;

for (i in 3:k){

    Eq[i]=EZ[i];

    if (floor(i/2) == i/2){

        a=(i+2)/2;

        for (j in 2:(a-1)){

            b=i-2*(j-1);

            Eq[i]=Eq[i]+coeff[i,j]*EZ[b];

        }

        Eq[i]=Eq[i]+coeff[i,a];

    }

    if (floor(i/2) < i/2){

        a=(i+1)/2;

        for (j in 2:a){

            b=i-2*(j-1);

            Eq[i]=Eq[i]+coeff[i,j]*EZ[b];

        }

    }

}

herm=function(par){

    i=par[1];

    w=par[2];

```

```
Hermite=w^i;

if (floor(i/2) == i/2){

  for (j in 2:(i/2)){

    b=i-2*(j-1);

    Hermite=Hermite+coeff[i,j]*w^b;

  }

  Hermite=Hermite+coeff[i,(i+2)/2];

}

if (floor(i/2) < i/2){

  for (j in 2:((i+1)/2)){

    b=i-2*(j-1);

    Hermite=Hermite+coeff[i,j]*w^b;

  }

}

Hermite;

}

marginal_1=function(par){

sig=par[1]

w=par[2]

Eq1=par[3]

Eq2=par[4]

Eq3=par[5]
```

```

Eq4=par[6]

main=1+w*Eq1 + (1/2)*(w^2-1)*Eq2;

for (i in 3:4){

    main=main+(1/factorial(i))*herm(c(i,w))*par[i+2];

    }

    sig*(1/(sqrt(2*pi)))*exp(-w^2/2)*main;

}

addition_1=function(par){

    sig=par[1]

    w=par[2]

    l=par[3]

    main=0;

    for (i in 5:1){

        main=main+(1/factorial(i))*herm(c(i,w))*par[i-2];

        }

        sig*(1/(sqrt(2*pi)))*exp(-w^2/2)*main;

    }

P=xbar; d=1;

l2=array(dim=c(d,d));

l3=array(dim=c(d,d,d));

l4=array(dim=c(d,d,d,d));

for (j in 1:d){

```

```

for (l in 1:d){
  l2[j,l]=-x/(P^2)-(n-x)/(1-P)^2;
  for (i in 1:d){
    l3[j,l,i]=(2*x)/(P^3)-2*(n-x)/(1-P)^3;
    for (q in 1:d){
      l4[j,l,i,q]=-(6*x)/(P^4)-6*(n-x)/(1-P)^4;
    }
  }
}
}

Hessian=l2;

det_Hessian=det(Hessian)

covariance=solve(-Hessian)

#The covariance matrix of \hat{\theta} is inverse of negtive Hessian.

xlevel=y_0=y_1=y_2=y_3=0; y_john1=0;

for (s in 1:100){
  xlevel[s]=s/100;
  delta=sigma*(xlevel[s]-xbar);
  y_0[s]=sigma*dnorm(delta);
  y_1[s]=dbeta(xlevel[s], a_new, b_new)
  y_2[s]=marginal_1(c(sigma,delta,Eq[1],Eq[2],Eq[3],Eq[4]));
  y_3[s]=y_2[s]+addition_1(c(sigma,delta,k,Eq[5:k]));
}

```



```

    y_john1[s]=sigma*dnorm(delta)*(1+(1/sqrt(n))*gam1_p(delta));
}

par(mfrow=c(1,1))

matplot(xlevel,cbind(y_0,y_1,y_john1),xlab=expression(theta),
ylab="marginal posterior density",main="Binomial Beta",lty=1:3,type="l")

```

مراجع

- [۱] اسحاقی، ا. (۱۳۹۲)، مدل‌های نیمه پارامتری تحلیل بقا برای داده‌های بازگشتی با روش هسته، پایان‌نامه کارشناسی ارشد، دانشگاه صنعتی شاهرود.
- [2] Basawa, I. V. & Rao, B. L. S. P. (1980), *Statistical Inference for Stochastic Processes*, Academic Press, London.
- [3] Blinnikov, S., & Moessner, R. (1998). Expansions for nearly Gaussian distributions. *Astronomy and Astrophysics Supplement Series*, 130(1), 193-205.
- [4] Brillinger, D. R. (1969), The Calculation of the cumulants via conditioning, *The Annals of the Institute of Statistical Mathematics*, **21**, 215-218.
- [5] Boas, M. L. (2006), *Mathematical Methods in Physical Sciences, 3rd Edition*, John Wiley & Sons, Inc, New Jersey.
- [6] Chen, C. F. (1985), On asymptotic normality of limiting density functions with Bayesian implications, *Journal of the Royal Statistical Society, Series B*, **47**, 540-546.
- [7] Chen, L. H. Y. (1975), Poisson approximation for dependent trials, *The Annals of Probability*, **3**, 534-545.
- [8] Courant, R. & Hilbert, D. (1953), *Methods of Mathematical Physics*, Interscience, Volume 1.
- [9] Cramer, H. (1999), *Mathematical methods of statistics*, Princeton University Press.
- [10] Cronish, E. A. & Fisher Ronald, A. (1938), Moments and cumulants in the specification of distributions, *Review of the Interational Statistical Institue*, **5**(4), 307-320.
- [11] Das Gupta, A. (2008), *Asymptotic Theory of Statistics and Probability*, Springer, New York.

- [12] David, A. P. (1970), On the limiting normality of posterior distribution, *Mathematical Proceedings of the Cambridge Philosophical Society*, **67**, 625-633.
- [13] Essen, C. (1945), Fourier Analysis of Distribution Functions, *Acta Mathematica*, **77**, 1-125.
- [14] Filho, R. G. B. & Rosenfeld, R. (2004), Testing option pricing with the Edgeworth expansion physica A: *Statistical Mechanics and Its Applications*, **344**, 484-490.
- [15] Fisher R. D. & Cornish, E. A. (1960). The percentile points of distributions having known cumulants American society for quality, *Journal of the American Statistical Association*, **2**(2), 209-225.
- [16] Ghosh, J. K., Sinha, B., and Joshi, S. (1982), Expansions for Posterior Probability and Integrated Bayes Risk, In Gupta, S. & Berger, J. (eds.), *Statistical Decision Theory and Related Topics III*, Academic Press, New York, **1**, 403-456.
- [17] George C. & Roger L. B. (2001), *Statistical Inference, 2nd Edition*, Springer, New York.
- [18] Gut A. (2005), *Probability: A Graduate Course*, Springer, New York.
- [19] Gentle, J. E. (1998), *Numerical linear algebra for applications in statistics*, Springer Verlag, New York.
- [20] Hald, A. (2000), The early history of the cumulants and the Gram-Charlier series, *International Statistical Review*, **68**(2), 137-153
- [21] Hall, P. (1992), *The bootstrap and Edgeworth expansion*, Springer, New York.
- [22] Heyde, C.C. & Johnstone, I.M. (1979), On asymptotic posterior normality for stochastic processes, *Journal of the Royal Statistical Society, Series B*, **41**, 184-189.
- [23] Johnson, R. (1967), An asymptotic expansion for posterior distributions, *The Annals of Mathematical Statistics*, **38**, 1899-1906.
- [24] Johnson, R. (1970), Asymptotic expansions associated with posterior distributions, *The Annals of Mathematical Statistics*, **141**, 851-864.
- [25] Lahiri, S. N. (2003), *Resampling methods for dependent data*, Springer-Verlag, New York.
- [26] Lindley D. V. (1961), The use of prior probability distributions in statistical inference and decisions, in *Proc. of the 4th Berkeley Symposium* **1**, 453-468.

- [27] Lindley, D. V. (1980). Approximate bayesian methods, *Bayesian Statistic*, University Press.
- [28] Lecam, L. (1953), On some asymptotic properties of maximum likelihood and related bayes estimates, *University of California Publications in Statistics*, **1**, 277-330.
- [29] Mosteller, F. & Wallace, D. L. (1964), *Inference and disputed authorship: The Federalis*, Reading, MA: Addison-Wesley Publishing Company.
- [30] Murray, G. D. (1977), Comment on maximum likelihood from incomplete data via the EM algorithm by A. P. Dempster, N. M. Laird & D. B. Rubin, *Journal of the Royal Statistical Society, Series B*, **39**, 27-28.
- [31] Stein, C. (1972), A bound for the error in the normal approximation to the distribution of a sum of dependent random variables, *In Proceeding of Sixty Berkeley Symposium on Mathematical Statistics and Probability*, **2**, 586-602.
- [32] Stein, C. (1981), Estimation of the Mean of a Multivariate Normal Distribution, *The Annals of Mathematical Statistics*, **9**, 1135-1151.
- [33] Stein, C. (1987), *Approximate Computation of expectations, 7th Edition*, Institute of Mathematical Statistics, Hayward, California.
- [34] Stuart, A. & Ord, J. K. (1994), *Kendall's Advanced Theory of Statistics, 6th Edition*, London.
- [35] Sweeting, T. J. (1992), On asymptotic posterior normality in the multiparameter case, *In Bayesian Statistics* (eds Bernardo, J. M., Berger, J. O., Dawid, A. P. & Smith, A. F. M.) University Press, (Eds), Oxford, 825-835.
- [36] Sweeting, T. J. & Adekola, A. O. (1987), Asymptotic posterior normality for stochastic processes revisited, *Journal of the Royal Statistical Society Series B*, **49**, 215-222.
- [37] Tanner, M. A. & Wong, W. H. (1987), The calculation of posterior distributions by data augmentation, *Journal of the American Statistical Association*, **82**, 528-540.
- [38] Van der vaart A. W. (1998), *Asymptotic Statistic, 1st Edition*, Cambridge University Press, New York.
- [39] Walter, R. (1987), *Real and Complex Analysis, 3rd Edition*, Mc Graw-Hill, Inc, New York.

- [40] Wallace, D. L. (1958), Asymptotic approximations to distribution, *The Annals of Mathematical Statistics*, **29**, 635-654.
- [41] Walker, A. M. (1969). On the asymptotic behaviour of posterior distributions, *Journal of the Royal Statistical Society, Series B*, **31**, 80-88.
- [42] Weber, H. J. & Arfken, G. B. (2004), *Essential Mathematical Methods for Physicists*, Elsevier Academic Press, San Diego.
- [43] Weng, R. C. (2003), On Stein's identity for posterior normality, *Statistica Sinica*, **13**, 495-506.
- [44] Weng, R.C. (2010), A Bayesian Edgeworth expansion by Stein's identity, *Bayesian Analysis*, **5**, 741-763.
- [45] Weng, R. C. & Lin, C.-J. (2010), A Bayesian approximation method for online Ranking, *Revision invited by Journal of Machine Learning Research*.
- [46] Weng, R. C. & Tsai, W.-C. (2008), Asymptotic posterior normality for multiparameter problems, *Journal of Statistical Planning and Inference*, **138**, 4068-4080.
- [47] Weng, R. C. & Woodroffe, M. (2000), Integrable expansions for posterior distributions for multiparameter exponential families with applications to sequential confidence levels, *Statistica Sinica*, **10**, 693-713.
- [48] Weng, R. C. & Hsu C. H (2013), A study of expansion of posterior distributions, *Statistica Sinica*, Volume 42.
- [49] Woodroffe, M. (1989), Very weak expansions for sequentially designed experiments: linear models, *The Annals of the Statistics*, **17**, 1087-1102.
- [50] Woodroffe, M. (1992), Integrable expansions for posterior distributions for One-parameter exponential families, *Statistica Sinica*, **2**, 91-111.
- [51] Woodroffe, M. & Coad, D. S. (1997), Corrected confidence sets for sequentially designed experiments, *Statistica Sinica*, **7**, 53-74.
- [52] Warren p. Johnson. (1996), A q-analogue of Fa'a Di Bruno's Formula, *The Journal of Combinatorial Theory, Series A*, **76**(2), 305-314
- [53] Robert, C. P. (2007), *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation, 2nd Edition*, Speriger-Verlag, New York.

-
- [54] Tierney, L. & Kadane, J. B. (1986), Accurate approximations for posterior moments and marginal densities, *Journal of the American Statistical Association*, **81**, 82-86.

Surname: Kheirgoo

Name: Rahim

Title: Approximating posterior distributions by using Edgeworth expansion and Stein's Identity

Supervisor: Dr. Hossein Baghishani

Advisor: Dr. Negar Eghbal

Degree: Master of Science

Subject: Statistics

Field: Mathematical Statistics

University of Shahrood

Faculty of Mathematical Sciences

Date: 2014

Number of pages: [96](#)

Keywords: Edgeworth expansion, Hermite polynomials, posterior distribution, Stein's Identity.

Abstract

In many applications, researchers are interested to approximate a probability distribution function. The Edgeworth expansion is a method for this aim. The Edgeworth expansion is a series that approximates a probability distribution in terms of its cumulants. One can derive it by first expanding the probability distribution in Hermite orthogonal functions and then collecting terms in powers of the sample size. In this thesis, by using a version of Stein's Identity and properties of Hermite polynomials, we derive an expansion for posterior distributions which possesses these features of an Edgeworth series. Then, we compare the performance of the proposed series with the previous ones from a Bayesian perspective. Finally, two examples are provided to illustrate the accuracy of the proposed expansion.



University of Shahrood
Faculty of Mathematical Sciences

Dissertation Submitted in Partial
Fulfillment of the Requirements for the
Degree of Master of Science in
Statistics

Approximating posterior distributions by using Edgeworth expansion and Stein's Identity

Supervisor

Dr. Hossein Baghishani

Advisor

Dr. Negar Eghbal

by

Rahim Kheirgoo

2014