بِسْمِ اللَّهِ الرَّحْمَٰنِ الرَّحِيمِ

دانشگاه پیام نور

دانشکده علوم پایه و کشاورزی

مرکز تهران

پایان نامه

برای دریافت مدرک کارشناسی ارشد

آمار ریاضی

گروه آمار

برآورد غیر مستقیم توزیع‌ها و فرآیندهای آلفا پایدار

یونس جوادی

استاد راهنما : دکتر پرویز نصیری

استاد مشاور : دکتر علی شادرخ

اردیبهشت ۱۳۹۰

تقدیم به مادرم و پدرم

که با مهر خویش مرا پروردند.

بدینوسیله از تمامی افرادی که به نحوی مرا در انجام این تحقیق یاری نمودند تشکر و قدردانی می نمایم .

چکیده

خانواده‌ی توزیع‌های α-پایدار یک کلاس غنی از توزیع های احتمالی هستند که از فرم عمـومی قضـیه حـد مرکزی نشأت می گیرند ،به طوریکه فرضیه‌ای با محدودیت خیلی کمتر درباره‌ی رفتار منظم دم های توزیـع بـا فرض واریانس متناهی در قضیه حد مرکزی جایگزین شده است. بـه عبـارتی دیگـر بـا در نظـر گـرفتن فـرض واریانس نامتناهی به دم های توزیع آزادی عمل بیشتری داده شده است.خانواده‌ی این توزیع ها الگوهای خیلـی جالبی از شکل توزیع ها را دارند به طوریکه با در نظر گرفتن عدم تقارن و دم های کلفـت و تغییرپـذیری زیـاد برای مدل‌بندی پدیده ها در زمینه های امور مالی، اقتصادی و فیزیکی مناسب می باشند.با آنکه چگـالی احتمـال متغیر تصادفی α-پایدار وجود دارد ولی فرم آنالیزی بسته‌ای برای توابع توزیع و چگالی این خـانواده از توزیـع ها تعریف نشده است و از آنجایی که مقادیر شبیه سازی شده از توزیع های α-پایدار را بطور صریح مـی تـوان بدست آورد روش های غیرمستقیم می تواند راهکاری مفید برای غلبه براین مشکل باشد از این رو دراین مقالـه به بررسی چنین روشی در بـرآورد پارامترهـای توزیـع α-پایـدار پرداختـه مـی شـود. ایـن روش بـه بـرآورد پارامترهای فرآیند آرمای آلفا پایدار نیز گسترش می یابد .

واژگان کلیدی : توزیع های پایدار ،برآورد غیر مستقیم ، واریانس نامتناهی

فهرست

فهرست اشکال و جداول

پیشگفتار

آنچه در کائنات هست بی شک خارج از علم ازلی خداوند نیست و در این بین پیروی تمـام عـالم از نظم و الگویی خاص مشهود است ،اما حقیقت طلبی و علم دوستی انسان ،او را به سمت کشـف ایـن الگوها می کشاند ولی هر رخدادی در زمان و مکان و در یک حالت خاص در مورد یک فرد یا شـئی خاص اتفاق می افتد.

به طوری که در زندگی روزمره شاهد پدیده های تصادفی هستیم که به هـر دلیلـی نمـی تـوانیم آن را کنترل و پیش بینی کنیم .اگر این پدیده ها با مشخصه‌های عددی مرتبط شود از این مقادیر بـه عنـوان متغیرهای تصادفی یاد می شود.

به منظور ایجاد یک نظریه‌ی اساسی برای مطالعه متغیرهای تصادفی ،معرفـی و رسـمی کـردن نظریـه احتمال برای تمام موقعیت های ممکن با کمک گرفتن از مدل های ریاضی لازم است. فی‌المثل بـا در نظر گرفتن یک تابع هموار $Y = f(X_۱, X_۲, \dots X_n)$

از متغیرهای تصادفی $X_۱, X_۲, \dots X_n$ که فعالیت های کوچک و مستقل بر روی سیستم تحت نظارت را نشان می دهد و اگر $f(۰, \dots, ۰) = ۰$ ، ساده ترین شکل این تابع همـوار ،یعنـی مجمـوع متغیـر های تصادفی

$$Y = \sum_{i=۱}^{n} c_i X_i$$

حاصل می شود.

یکی از موقعیت های مشابه هنگامی رخ می دهد که خطـای مشـاهدات(ناشـی از اخـتلاف مـدل بـا واقعیت) که در آزمایش‌ها اندازه‌گیری می شوند مورد تحلیل قرار گیرند.

لاپلاس و گاوس کسانی بودند که نظریه‌ی خطای مشاهدات را در آغاز قرن ۱۹ گسترش دادنـد . اگـر هر فرض صفری به غیر از کوچکی خطاها را داشته باشیم ،به هیچ وجه نظریـه ی مطلـوبی دربـاره ی توزیع مجموع این متغیرها به دست نمی دهد. و دقیقاً به ایـن دلیـل توزیـع خطـا بـا طـرح مجمـوع متغیرهای تصادفی وابسته می شود. با این حال اگر کسی فرض کند که خطاهـا مسـتقل انـد ناگهـان وضعیت تغییر می کند.

اولین نتایج درباره طرح مجموع متغیرهای تصادفی مستقل توسط ژاکوب برنولی در ۱۷۱۳ ارائه شـد . وی با در نظر گرفتن دنباله‌ای از مجموع نرمالیده که متغیرهای تصادفی مستقل X_i با احتمال p مقدار ۱ و با احتمال $p - ۱$ مقدار صفر را می گیرند. بیان می کند که برای هر $۰ < \varepsilon$ ثابت دلخواه کوچک احتمال زیر را داریم

$$P\left\{\left|\frac{۱}{n}\sum_{i=۱}^{n} X_i - p\right| > \varepsilon\right\} \to ۰ \qquad n \to \infty$$

برآورد آن که ،امروزه فرم برنولی از قانون اعداد بزرگ نامیده می شود ،سخت بود. اما قانون اعداد بزرگ و تعمیم های مختلف آن و شدت روابط آنها با هم استنباط از اطلاعات به دست آمده از آزمایش ها و مشاهدات را آسان‌تر نموده است.

دومین نتیجه ی معنی داری که در قرن ۱۸ از بسط قضیه برنولی به دست آمد قضیه مویور–لاپلاس بود. این یک حالت خاص (مربوط به متغیرهای تصادفی از قضیه ی برنولی) از قضیه حد مرکزی نظریه‌ی احتمال است.

تا به امروز قضیه حد مرکزی به شیوه های مختلفی بیان شده و گسترش یافته است . یکی از آنها به این منظور گسترش یافت که قضیه حد مرکزی نه فقط با استفاده از توزیع نرمال به عنوان تقریب حدی بلکه برخی دیگر از توزیع هایی که ساختار جبری معین دارند ،درک شود.

دنباله ی متغیرهای تصادفی مستقل هم توزیع $X_۱, X_۲, \dots$ بدون هیچ فرضی درباره توزیع اولیه آنها را در نظر بگیرید . با استفاده از دنباله ای از مقادیر ثابت حقیقی $a_۱, a_۲, \dots$ و ثابت های مثبت $d_۱, d_۲, \dots$ داریم

$$Z_n = \frac{\sum_{i=۱}^{n} X_i - a_n}{d_n}$$

حال فرض کنیم ثابت های a_n و d_n طوری انتخاب شوند که داشته باشیم ،تابع توزیع مربوطه برای هر x که نقطه‌ی پیوستگی تابع است به توزیع حدی $G(x)$ همگرای ضعیف باشد ؛

$$P\{Z_n < x\} \Rightarrow G(x) \qquad n \to \infty$$

حال به بررسی چگونگی گسترش کلاس این تابع توزیع می پردازیم .و بدین ترتیب وارد بحث درباره‌ی خانواده‌ی توزیع های α–پایدار می شویم.

برای اولین بار در سال ۱۹۲۰ به هنگام مطالعه پول لەوی[۱]؛ در مورد مجموع متغیرهای تصادفی مستقل هم توزیع ،توزیع‌های پایدار شناخته شدند. گنەدنکو و کلموگروف[۲] در سال (۱۹۵٤)، با جایگذاری فرضیه ای با محدودیت خیلی کمتر درباره رفتار منظم دم های توزیع به جای فرض واریانس متناهی در قضیه حد مرکزی اولین تعریف خانواده توزیع‌های α– پایدار را ارائه کردند. به عبارت دیگر، خانواده‌ی توزیع‌های α–پایدار یک کلاس غنی از توزیع‌های احتمالی هستند که از فرم

[۱] Paul Lévy
[۲] Gnedenko and Kolmogorov

عمومی قضیه حد مرکزی نشأت می گیرند ، به طوریکه با در نظر گرفتن فرض واریانس نامتناهی به دم های توزیع آزادی عمل بیشتری داده شده است.
با آنکه چگالی احتمال متغیر تصادفی α-پایدار وجود دارد ولی فرم جبری بسته‌ای برای توابع توزیع و چگالی این خانواده از توزیع ها تعریف نشده است بجز در موارد خاص مانند توزیع‌های نرمال ،کوشی و له‌وی. ولی خانواده‌ی این توزیع ها الگوهای خیلی جالبی از شکل توزیع ها را دارند به طوریکه با در نظر گرفتن عدم تقارن و دم های کلفت و تغییرپذیری زیاد برای مدل‌بندی پدیده ها مناسب می باشند.
به همین دلیل ،توزیع های α-پایدار کاربرد زیادی در مدل بندی پدیده ها در زمینه های امور مالی، اقتصادی و فیزیکی دارند .به عنوان مثال میدان جاذبه‌ی ستارگان ،توزیع دمای راکتور هسته‌ای ،قیمت های بازار سهام ، بارش سالیانه و ...

به طوری که مندلبروت[1] در سال (۱۹٦۳) در یک مقاله با توجهی خاص به توزیع‌های دم کلفت در مقایسه با توزیع تی استیودنت با نگرش اقتصادی پرداخت. بکار گیری توزیع α - پایدار در مدل های سری زمانی با بررسی ارتباط بین مدل های قارچ[2]و توزیعهایα– پایدار توسط دی وریس[3]در سال(۱۹۹۱)آغاز شد و مدل های قارچ مبتنی بر اغتشاشα - پایدار توسط مک کولاک[4](۱۹۸۵)و لیو و برورسن[5] (۱۹۹۵) پیشنهاد شده بود . برآورد فرآیندهای مدل سری زمانی با اغتشاش α - پایدار بر اساس توزیع گاوسی توسط میکوسک و همکاران[6] در سال (۱۹۹۵) مورد بررسی قرار گرفته است.

متاسفانه داشتن سختی‌های برآورد مانع گسترش کاربرد توزیع‌های α-پایدار در میان دست اندرکاران شده است.و به طور خاص نبود یک فرم بسته تابع چگالی مانع استفاده ازروش حداکثر راستنمایی[7] (ML) ونبود گشتاورهای مراتب بالاتر مانع استفاده از روش گشتاوری[8] (MM) شده است.
روش های برآوردی جایگزین عمدتاً برپایه چارک ها[1] یا تابع مشخصه تجربی[2] پیشنهاد شده ودرنوشته ها و مقالات بکاررفته است . در مورد روش های برآوردی جایگزین بر پایه چارک ها

[1] Mandelbrot
[2] GARCH
[3] De Vries
[4] McCulloch
[5] Liu &Brorsen

[6] Mikosch et al
[7] Maximum likelihood
[8] Method of Moments

بوسیله مک کولاک (۱۹۸٦) چارک های عددی نولان[3](۱۹۹۷) ، تابع مشخصه تجربی توسط کوگان و ویلیامز[4] در (۱۹۹۸) و کاراسکو و فلورنس[5] (۲۰۰۲) معرفی شده اند.به طوری که در سال (۲۰۰۲)کاراسکو و فلورنس برآورد گرهای کارا را بوسیله انطباق تابع مشخصه تجربی با تابع مشخصه نظری به دست آوردند و برای تعداد متناهی از گشتاورها ،روش گشتاوری تعمیم یافته[6] (GMM) را ارائه کردند که بررسی روش GMM مبتنی بر دنباله‌ی گشتاور از کارهای دیگری است که توسط ایشان و در این مقاله انجام شده است. در ادامه کاراسکو و فلورنس به بررسی کارایی مجانبی و ویژگی های مجانبی برآوردگر GMM و همچنین به بحث بر روی تابع مشخصه شرطی پرداختند.

ازآنجایی که مقادیرشبیه سازی شده از توزیع‌های α-پایدار را بطور صریح می توان بدست آورد روش های برآورد غیرمستقیم[7]می تواند راهکاری مفید برای غلبه براین مشکل باشد .

گوآریروکس و همکاران[8]در سال (۱۹۹۳) روش های غیرمستقیم برای برآورد پارامترها ارائه داده اند که در آن به بررسی "پارامتر کمکی و برآوردگر آن" و توزیع مجانبی این برآوردگرها پرداخته‌اند و آزمون فرض بر روی پارامتر مورد علاقه و کاربرد این روش در مثال های مختلف بررسی شده است.

میت نیک و همکاران[9]در (۱۹۹۹) روش غیر مستقیم با حداکثر راستنمایی تقریبی را مورد بررسی قرار داده اند.

در سال (۲۰۰٤)گارسیا و همکاران[10] به تشریح روش برآورد غیر مستقیم پارامترهای توزیع α– پایدار پرداخته و روش خود را با دو شیوه از روش های جایگزین گشتاور منطبق[11] مقایسه کرده اند.

روش غیرمستقیم مقید که به وسیله کالزولاری وهمکاران[1]در(۲۰۰٤)که تعمیمی است از EMM گالانت و تائوچن[2] (۱۹۹٦)،ارائه شده است. چمبرز وهمکاران[3]در(۱۹۷٦)و(۱۹۸۷) الگوریتمی برای شبیه سازی متغیرهای تصادفی پایدار با بررسی سرعت ودقت محاسبات ارائه کرده اند.

---

[1] quantiles
[2] empirical characteristic function
[3] Nolan
[4] Kogon and Williams
[5] Carrasco and Florens
[6] Generalized Moment Method
[7] Indired approaches
[8] Gouri´eroux et al
[9] Mittnik et al
[10] Garcia, R., E. Renault, and A. Veredas
[11] Alternative moment matching methods

در مورد توزیع چوله -t و مشـتقات مراتـب اول لگـاریتم تـابع راسـتنمایی (تـابع اسـکور) ،آزالینـی و کاپیتانیو[٤](۲۰۰۳) و(۲۰۰۵)که نتیجه حاصل از توسعه توزیع‌های چوله- نرمال آزالینی (۱۹۸۵) است به خوبی بحث کرده اند.جزئیات مربوط به خواص مطلوب جهت " فقط برای شناسایی"[٥] مـدل کمکـی در مقاله هاگلند فرایسی[٦] (۲۰۰٤) و بـرای "بـیش شناسـایی"[٧] انتخـاب مـدل کمکـی تولیـد کننـده برآوردگرهای گشتاوری کارا (EMM)[٨] در مقاله گالانت وتائوچن (۱۹۹٦)آورده شده است.

روش بیزی به دلیل افزایش توان محاسباتی بوسیله ی باکل[٩](۱۹۹٥) کیو؛ و راویشانکر[١٠](۱۹۹۸) مـورد بررسی قرار گرفته اند ، فرناندز و استیل[١١] در سال (۱۹۹۸) با بهره گیری از رویکـرد بیـزی در یـافتن مدل کمکی برای توزیع α - پایدار با استفاده از توزیع t اسـتیودنت و شـبیه سـازی مـدل کمکـی بـا روش MCMC نسبت به توزیع پیشین هر یک از ٤ پارامتر توزیـعα – پایـدار، پارامترهـا را بـرآورد کرده‌اند.
لمباردی[١٢](۲۰۰۷)با استفاده از روش بیزی به استنباط در مورد توزیع‌های α– پایدار مـی پـردازد و بـا استفاده از روش MCMC پارامترهای مدل را برآورد می کند.

در سال (۲۰۰۹) اندروز و همکاران [١٣] به برآورد پارامترهـای مـدل اتورگرسـیو علّـی و غیـر علّـی بـا اغتشاش α-پایدار غیرنرمال پرداخته و به دلیل اینکه فرم حدی توزیع برآوردگر بسته نبـود از ایـن رو بررسی شکل توزیع برآوردگر با روش بوت استرپ انجام شده است .

لمباردی و کالزولاری در سال (۲۰۰۸) با ارائه روش برآورد غیـر مسـتقیم توزیـع‌هـای α– پایـدار و استفاده از توزیع چوله–t به عنوان مدل کمکی ،بحث بر روی تابع اتصال[١٤] بـه بـرآورد غیـر مسـتقیم

---

[١] Calzolari et al
[٢] Gallant and Tauchen
[٣] Chambers et al
[٤] Azzalini and Capitanio.
[٥] just-identified
[٦] Heggland and Frigessi .
[٧] over-identified
[٨] Efficient Method of Moments
[٩] Buckle
[١٠] Qiou and Ravishanker
[١١] Fernandez, C. and M. Steel
[١٢] Lombardi

[١٣] Andrews, B., M. Calder, and R. Davis
[١٤] binding function

مقید دست یافتند و به مقایسه روش خود با حداکثر راستنمایی تقریبی در روش مبتنی بر FFT میت نیک (۱۹۹۹) پرداختند و به این نتیجه رسیدند که هر دو روش نسبتاً به نتایج مشابهی منتهی می شوند. و در نهایت با برآورد پارامترهای مدل ARMA(۱,۱) با خطای چوله-t و نرمال به نتیجه عدم اختلاف زیاد تاثیر این دو توزیع در برآورد پارامترهای مدل دست یافتند.

از این رو دراین پایانامه به بررسی چنین روشی در برآورد پارامترهای توزیع α-پایدار پرداخته خواهد شد. و به برآورد پارامترهای فرآیند آرمای آلفا پایدار نیز گسترش می یابد. روش غیر مستقیم تبیین‌کننده‌ی یک رویکرد استنباطی مناسب برای شرایطی است که برآورد مدل آماری مطلوب ، سخت می باشد ، و برای تولید مقادیر شبیه سازی شده از مدل مشابه ،که دقیقاً هم ارز توزیع‌اصلی است مناسب می باشد.

۱– قضیه‌ی حد مرکزی

مقدمه

قضیه‌ی حد مرکزی در آمار از اهمیت بسیاری برخوردار است و در عبارت جبری آن که لیندبرگ و له‌وی ارائه داده اند ،بیان می شود که :فرض کنیم یک دنباله‌ی n تایی از متغیرهای تصادفی مستقل هم توزیع با واریانس متناهی داشته باشیم ،در این صورت وقتی n به بی نهایت میل می کند مجموع استاندارد شده‌ی متغیرهای تصادفی بدون در نظر گرفتن توزیع اولیه متغیرها ، به توزیع نرمال میل می کند.

اغلب پدیده های تصادفی ممکن است به صورت مجموع سهم عوامل کوچکتر ملاحظه شوند. مثلاً اگر فرض کنیم اغتشاشات در مدل های رگرسیون و سری های زمانی ناشی از تعدد زیاد اثرات کوچک با واریانس متناهی هستند ،در نتیجه توزیع حاصل باید نرمال باشد .

۱–۱–قضیه‌ی حد مرکزی کلاسیک

در اینجا عبارت جبری و اثبات قضیه حد مرکزی در شکل "کلاسیک" آن ارائه می شود ،با اینکه برخی نتایج به خوبی شناخته شده هستند اما یادآوری آنها بجا خواهد بود.

قضیه۱.۱-(لیندبرگ-له‌وی[۱]*).* فرض کنید$\{X_i\}$، $i = ۱,۲,\ldots$ دنباله ای از متغیرهای تصادفی مستقل هم توزیع با میانگین μ و واریانس متناهی $\sigma^۲$ باشند ، آنگاه وقتی n به بی نهایت میل می کند

$$S_n = \frac{۱}{\sqrt{n}}\sum_{i=۱}^{n}\frac{X_i-\mu}{\sigma} \qquad (۱.۱)$$

به توزیع نرمال همگرا است.(لمباردی ،۲۰۰٤ :۱)

اثبات : فرض کنید Z_i مقدار استاندارد شده ی X_i با میانگین صفر و واریانس ۱ به صورت زیر می باشد:

$$Z_i = \frac{X_i-\mu}{\sigma}$$

---
[۱] Lindeberg- Levy

بدیهی است که Z_i ها متغیرهای تصادفی مستقل هم توزیـع مـی باشـند بنـابراین از تـابع مشخصـه یکسان برخوردارند که با $\phi_Z(t)$ نمایش داده می شود . در نتیجه تابع مشخصه S_n از رابطه ی زیـر به دست می آید:

$$\phi_{S_n}(t) = \prod_{i=۱}^{n} e^{itZ_i n^{-۱/۲}} = \left[\phi_Z(\frac{t}{\sqrt{n}})\right]^n$$

با بسط سری مک لورن از تابع مشخصه Z_i ها داریم؛

$$\phi_Z\left(\frac{t}{\sqrt{n}}\right) = ۱ + i\frac{t}{\sqrt{n}}\,E(Z_i) + i^۲\frac{t^۲}{۲n}E(Z_i{}^۲) + o$$

بنابراین تابع مشخصه مجموع ، S_n برابر است با

$$\phi_{S_n}(t) = \left[۱ - \frac{t^۲}{۲n} + o\right]^n \;\; ;$$

به طوریکه $\lim_{n\to\infty}\left(۱ + \frac{a}{n}\right)^n = e^a$ در نتیجه

$$\lim_{n\to\infty}\phi_{S_n}(t) = e^{\frac{-t^۲}{۲}}$$

که تابع مشخصه توزیع $\mathcal{N}(۰,۱)$ می باشد.

□

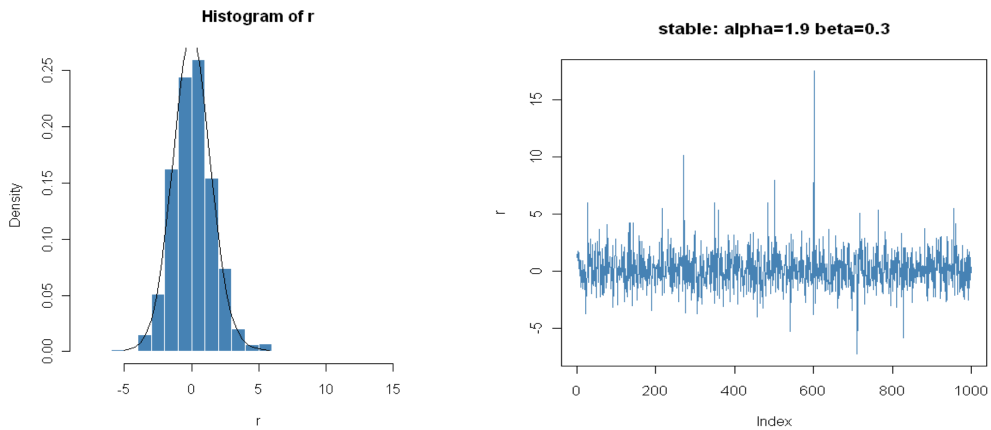
این تنها یکی از فرم های متعدد قضیه حد مرکزی می باشد و دلیل انتخاب آن ایجـاد زمینـه بـرای بـه دست آوردن فرم تعمیم یافته ی قضیه حد مرکزی است.

۱-۲- فرآیندهای با واریانس نامتناهی

بعضاً در یافته های تجربی مواردی یافت می شود که با فرضیات تئوریـک و آنچـه کـه مـورد انتظـار است، در تضاد باشد .مثلاً ممکن است مشاهده شود کـه مانـده هـای بـرآورد (اغتشاشـات)،دم هـایی کلفت‌تر از توزیع نرمال داشته باشند. این بدان معناست که فرض اینکـه اغتشاشـات بـه وسـیله نقـش تعداد زیادی از عوامل که واریانس متناهی دارند ، تبیین می شود باید اشتباه باشد. در واقع این کار بر

روی این فرض متمرکز می باشد که پدیده های تصادفی آماری و تغییرات آنهـا بـه صـورت مجمـوع سهم عوامل کوچک نباشد . بلکه به لحاظ تئوری واریانس پدیده ها ،نامتناهی در نظر گرفته می شود. به عنوان مثال نمونه ای از اغتشاشات را در نظر بگیرید که در شکل ۱.۱ نشان داده شده است.

شکل 1.1 بافت نگارو مدل اغتشاشات α-پایدار



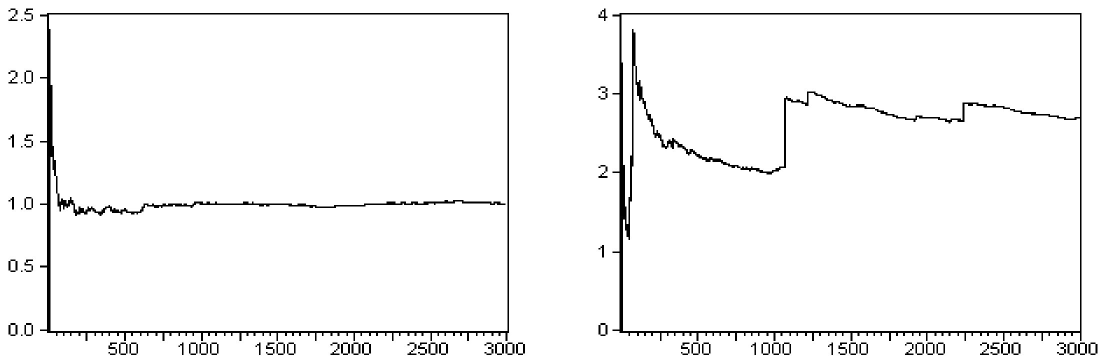
علیرغم اینکه به نظر می رسد این اغتشاشات از تجمع تعداد زیادی از اجزاء باشد ، ولی بافت‌نگـار دم ها را نسبت به توزیع نرمال کلفت‌تر نمایش می دهد و این می تواند نشانه ای مبنی بر نامتناهی بـودن واریانس اجزای این اغتشاشات باشد. در اینجا ابزار توصیفی مفیدی برای سنجش واریانس متناهی یـا نامتناهی یک بردار تصادفی ارائه می شود.

تغییرنگار یک دنباله تصادفی i.i.d برای بردار V با k ورودی به صورت زیر تعریف می شود

$$V_k = \sum_{i=1}^{k} \frac{(X_i-\bar{X}_k)^2}{n-1} \qquad (۲.۱)$$

که $\bar{X}_k$ نشان دهنده میانگین نمونه ای اولین k مشاهده است، V_k واریانس نمونه ای اولین k مشـاهده از مجموعه داده ها می باشد . بر اساس استنباط های مبتنی بر قضـیه حـد مرکـزی واریـانس نمونـه ، برآوردگر سازگار واریانس جامعه می باشد بنابراین وقتی تغیر نگار برای نمونه های به اندازه ی کافی بزرگ تهیه شود ،باید یک الگوی همگرا را نمایش دهد . وقتی چنین نشود در نتیجه "جهشی" ناشی از مشاهدات که در انتهای دم ها (به دلیل داشتن چگالی احتمال کمتر) ی توزیع قرار دارند بوجود می آید که نشانه ای از نا متناهی بودن واریانس می باشد. به عنوان مثال در شکل ۲.۱ یک الگوی واگـرا و یک الگوی همگرا نشان داده شده است.که در آن محور افقی حجم نمونه و محور عمـودی واریـانس مربوطه می باشد.

شکل ۲.۱ تغییر نگار همگرا و واگرا



۱–۳– قضیه‌ی حد مرکزی تعمیم یافته

در این بخش با مقدمه ای بر قضیه حد مرکزی تعمیم یافته به شناسایی خانواده‌ای از توزیع‌های حدی که نرمال حالت خاص آن است می پردازیم.

تعریف ۱.۱ (پایداری): تابع توزیع $F(x)$ را پایدار گوئیم اگر برای هر عدد مثبت c_1 و c_2 وهرعدد حقیقی d_1 و d_2 و به طور مشابه $c > 0$ و d داشته باشیم؛(گنه‌دنکو و کولموگروف،۱۹۶۷:۱۶۲)

$$F(c_1x + d_1).F(c_2x + d_2) = F(cx + d) \qquad (۳.۱)$$

اولین پایه هایی که علت وجودی قضیه حد مرکزی تعمیم یافته بوده اند در قضیه بعد آورده شده است؛ توزیع‌های پایدار تنها توزیع حدی ممکن برای مجموع نرمالیده از نوع رابطه ی زیر می باشد؛

$$S_n = \frac{X_1 + X_2 + \cdots + X_n}{C_n} - D_n$$

این نتیجه اولین بار توسط له وی در سال ۱۹۲٤ به صورت فرموله در قضیه زیر بیان شد:

قضیه ۲.۱–(له وی): فرض کنید$\{X_i\}$، دنباله ای از متغیرهای تصادفی مستقل هم توزیع باشد ،شرط لازم و کافی برای آنکه $F(x)$ توزیع حدی S_n ؛

$$S_n = \frac{\sum_{i=1}^{n} X_i}{C_n} - D_n \qquad (٤.۱)$$

باشد آن است که $F(x)$ پایدار باشد. (یوخایکین و زولوتارو[1]،۱۹۹۹:۳۳)

[1] Uchaikin, Vladimir V. Zolotarev, Vladimir M

اثبات: فقط شرط لازم اثبات می شود برای اثبات شرط کافی به (گنه دنکو و کولموگروف ،۱۹٦۷ :۱٦۳) مراجعه شود.

فرض کنید که S_n به توزیع حدی معین $F(x)$ همگرا باشد ،اثبات می شود که $F(x)$ پایدار است.

بر اساس لم صفحه ۱٤٦ گنه دنکو و کولموگروف(۱۹٦۷) ، اگر X یک توزیع مناسب داشته باشد عامل مقیاس بندی C_n به صورت زیر تعیین می شود؛

$$\lim_{n\to\infty} C_n = \infty$$

$$\lim_{n\to\infty} \frac{C_{n+1}}{C_n} = 1$$

دو مقدار ثابت و مثبت c_1 و c_2 را طوری در نظر بگیرید که $c_1 < c_2$ ؛بنابراین ممکن است برای هر ε و $l \geq l_\varepsilon$ با انتخاب m داشته باشیم؛

$$0 \leq \frac{C_m}{C_l} - \frac{c_2}{c_1} < \varepsilon$$

حال دو مقدار ثابت حقیقی d_1 و d_2 را در نظر می گیریم و تعریف می کنیم ؛ $C_n = {C_l}/{c_1}$ و

$$D_n = {[C_l D_l + C_m D_m + d_1 C_l + d_2 C_m]}\Big/{C_n}$$

با نوشتن مجدد مجموع رابطه (٤.۱) داریم ؛

$$\frac{C_l}{C_n}\left[\frac{X_1 + \cdots + X_l}{C_l} - D_l - \delta_1\right] + \frac{C_m}{C_n}\left[\frac{X_{l+1} + \cdots + X_{l+m}}{C_m} - D_m - \delta_2\right]$$

بنا به فرض اولیه S_n به $F(x)$ همگراست پس دو جمع فوق (بنا به گنه دنکو و کولموگروف ، ٤۲:۱۹٦۷ قضیه ۲) به ترتیب به $F(c_1^{-1}x + d_1)$ و $F((c_2^{-1}x + d_2)$ همگرا هستند ،از طرف دیگر چون توزیع حدی مجموع ،$F(x)$ است پس با پیچش تابع توزیع تجمعی دو مجموع فوق داریم ؛

$$F(x) = F(c_1^{-1}x + d_1) \,.\, F((c_2^{-1}x + d_2)$$

در نتیجه رابطه(۳.۱) اثبات شد و $F(x)$ پایدار است.

□