دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

برآورد اندازه طیفی بردارهای تصادفی پایدار

پایان‌نامه کارشناسی ارشد آمار محض

فاطمه گلستانه

استاد راهنما

دکتر صفیه محمودی

۱۳۸۷

دانشگاه صنعتی اصفهان

دانشکده علوم ریاضی

پایان‌نامه کارشناسی ارشد آمار محض خانم فاطمه گلستانه

تحت عنوان

برآورد اندازه طیفی بردارهای تصادفی پایدار

در تاریخ ۲۰ اسفندماه ۱۳۸۷ توسط کمیته تخصصی زیر مورد بررسی و تصویب نهائی قرار گرفت.

۱ــ استاد راهنمای پایان‌نامه	دکتر صفیه محمودی
۲ــ استاد مشاور پایان‌نامه	دکتر سعید پولادساز
۳ــ استاد داور ۱	دکتر علیرضا نعمت‌اللهی (دانشگاه شیراز)
۴ــ استاد داور ۲	دکتر افشین پرورده (دانشگاه اصفهان)

سرپرست تحصیلات تکمیلی دانشکده   دکتر رسول نصر اصفهانی

فهرست مطالب

چکیده:

مجموع متغیرهای تصادفی مستقل و هم‌توزیع بعد از تغییر مکان و مقیاس مناسب دارای یک توزیع حدی تحت عنوان توزیع پایدار می‌شود. این خانواده توزیع، توزیع نرمال را نیز دربر می‌گیرد و به‌جز در حالت نرمال دارای واریانس نامتناهی و دم‌های کلفت می‌باشد. این ویژگی و همچنین تنوع پارامترها این توزیع را برای مدل‌بندی آماری بسیاری از پدیده‌ها مناسب ساخته است. از نظر تئوری بررسی‌های زیادی در مورد برآورد پارامترهای پایدار یک‌متغیره صورت گرفته است، اما در مورد توزیع‌های پایدار چندمتغیره نیاز به بررسی‌های بیشتری وجود دارد. پارامترها در این حالت شامل شاخص پایداری، مکان و اندازه طیفی می‌باشد. مساله‌ی اصلی برآورد، برآورد اندازه طیفی است که یک اندازه متناهی روی کره واحد S_d در $\mathbb{R}^d$ می‌باشد. در این پایان‌نامه به اختصار به معرفی برآوردگرهای پارامترها می‌پردازیم و به طور مفصل برآوردگرهای پیشنهاد شده برای اندازه طیفی را بررسی می‌کنیم. همچنین با شبیه‌سازی و استفاده از داده‌های واقعی به طور شهودی آن‌ها را مقایسه خواهیم کرد.

رده بندی موضوعی: اولیه ٦٢G٠٥، ٦٢M١٥. ثانویه ٦٢G٣٢، ٦٢H١٢.

کلمات کلیدی: توزیع‌های چندمتغیره پایدار، اندازه طیفی، تابع مشخصه، تصویر، شبیه‌سازی.

فصل ۱

مقدمه

توزیع‌های $\alpha-$ پایدار[۱] کلاس بسیار وسیعی از توزیع‌ها را دربر می‌گیرند و به علت جرم موجود در دم تابع چگالی، این گروه از توزیع‌ها، به توزیع‌های دم‌سنگین معروفند. جرم احتمال موجود در دم‌های بالایی و پایینی توزیع به عدد α، مقداری بین صفر و دو، وابسته است به طوری که هرچقدر α کمتر باشد جرم احتمال در دم‌ها بیشتر است. نظریه توزیع‌های پایدار در دهه‌های ۱۹۲۰ و ۱۹۳۰ توسط پل لوی[۲] و الکساندر یائولویچ خینچین [۳] بیان شد. آثار ندنکو و کولموگرف [۱۶] و فلر [۱۴] از جمله آثار قدیمی هستند که به بررسی جزئیات این توزیع پرداخته‌اند و اخیراً نیز زولوتارو [۳۸] به بررسی این موضوع پرداخته است.

در چند دهه‌ی اخیر داده‌هایی با دم‌های سنگین در زمینه‌های مختلف اقتصادی، ارتباطات، فیزیک و غیره جمع آوری شده‌اند که استفاده از توزیع‌های پایدار غیرگوسی به عنوان مدل‌های ممکن برای آن‌ها پیشنهاد می‌شود. چنین مدل‌هایی انعطاف‌پذیری و تغییرپذیری بیشتری نسبت به توزیع‌های گوسی دارند. با وجود اینکه توزیع‌های گوسی به طور کامل به وسیله میانگین و کوواریانس خود مشخص می‌شوند اما توزیع‌های پایدار غیرگوسی به پارامترهای بیشتری برای تشخیص نیاز دارند. همچنین توزیع گوسی همواره نسبت به میانگین متقارن است، اما توزیع‌های غیرگوسی پایدار می‌توانند درجه‌ای دلخواه از چولگی را اختیار کنند.

---

[۱] $\alpha-$stable

[۲] Paul Levy

[۳] Aleksander Yakovlevich Khinchine

چگالی احتمال متغیرهای تصادفی $\alpha-$ پایدار موجود و پیوسته است، اما به‌جز در چند مورد بقیه فرم مشخص بسته‌ای ندارند. همین امر روش‌های معمول آماری از جمله روش‌های برآورد کردن پارامترها را در مورد این توزیع‌ها چه در حالت یک‌متغیره و چه در حالت چندمتغیره تحت تأثیر قرار داده و بسیاری از آن‌ها را بی‌اعتبار ساخته است. برای مثال برای محاسبه ماکزیمم درستنمایی از تقریب‌های عددی سنگین باید استفاده کرد. این یک مانع بزرگ برای استفاده از توزیع‌های پایدار به وسیله محققان است. برنامه کامپیوتری معتبری برای محاسبه چگالی‌های پایدار و چندک‌های آن وجود ندارد. وجود یک‌چنین برنامه‌ای امکان استفاده از مدل‌های پایدار را در سطح وسیعی فراهم می‌کند. تنها فرم بسته‌ی موجود در مورد این توزیع‌ها فرم تابع مشخصه‌ی آن‌ها است که البته توسط آماردانان مختلف به شیوه‌های متفاوتی پارامتربندی شده است.

روش‌های مختلفی برای برآورد پارامترها ارایه شده است که از آن جمله می‌توان روش برآورد بر اساس تابع مشخصه توسط پائولسون و همکاران [۳۱] در سال ۱۹۷۵ و کوتروولِیس [۱۷] در سال ۱۹۸۰، روش برآورد از طریق محاسبه‌ی چندک‌ها در سال ۱۹۸۶ توسط مک‌کلاچ [۲۰] و روش درستنمایی ماکزیمم توسط نولان [۲۶] در سال ۲۰۰۱ را نام برد. از دیگر کارهای آماری صورت گرفته در این زمینه می‌توان ارایه فرمول صریح برای شبیه‌سازی داده‌های پایدار توسط چمبر و همکاران در سال ۱۹۷۵ و اصلاح آن در سال ۱۹۸۷ ([۶])، بررسی رفتار دم و مد توزیع‌های پایدار توسط نولان [۲۵] در سال ۱۹۹۸ و فوفاک و نولان [۱۵] در سال ۱۹۹۹، و محاسبه چگالی با استفاده از نمایش انتگرال زولوتارو توسط نولان [۲۴] در سال ۱۹۹۷ و از طریق معکوس تابع مشخصه توسط میتنیک و همکاران [۲۲] در سال ۱۹۹۹ را نام برد.

در مورد توزیع‌های پایدار یک‌متغیره تحقیقات بسیار زیادی انجام شده است اما مسایل زیادی در حالت چندمتغیره باقی مانده است. توجه کنید که توزیع‌های پایدار چندمتغیره‌ی غیرگوسی به سادگی توزیع‌های چندمتغیره‌ی گوسی نیستند. توزیع گوسی N بعدی به طور کامل توسط ماتریس کواریانس $N \times N$ اش مشخص می‌شود که به طور مناسبی مختصاتش تحت تغییرات خطی تبدیل می‌شود. در واقع چون نرمال چندمتغیره را می‌توان به عنوان جمعی از متغیرهای نرمال یک‌متغیره مستقل مشخص کرد، می‌توان با ماتریس‌های متعامد و قطری (به دست آمده از مقادیر ویژه و بردارهای ویژه) یک پایه برای $\mathbb{R}^N$ پیدا کرد، این را آنالیز مولفه‌های اصلی می‌نامیم. در مورد توزیع پایدار چندمتغیره موقعیت بسیار پیچیده‌تر است. حاشیه‌ها دارای واریانس متناهی نیستند و نمی‌توان ماتریسی مانند ماتریس کواریانس برای آن‌ها تعریف کرد. در این مورد نمادی متفاوت با کواریانس تعریف می‌شود که نمی‌توان آن را تحت تغییر مختصات تبدیل کرد. همچنین چیزی به نام آنالیز مولفه‌های اصلی وجود ندارد. ساختار توزیع پایدار در $\mathbb{R}^d$ شامل یک اندازه معروف به اندازه طیفی است که روی کره S_d تعریف می‌شود. این اندازه اصولاً نامتناهی‌البعد است به طوری که هیچ ماتریس $N \times N$ ای برای ارایه آن مناسب نمی‌باشد.

هدف ما در این پایان‌نامه معرفی برآوردگرهای مختلف پارامترها، به ویژه برآوردگر اندازه طیفی و بررسی آن‌ها می‌باشد. برآورد اندازه طیفی از آن‌جا حایز اهمیت است که در اکثر روابط مربوط به توزیع‌های چندمتغیره ظاهر می‌شود و بنابراین محاسبه و شناخت آن برای انواع کاربردهای عملی ضروری است. در سال ۱۹۷۲ پرس سعی کرد فرم ساده‌تری از تابع مشخصه توزیع‌های پایدار چندمتغیره را ارایه دهد. اما تنها توانست خانواده‌ای خاص از توزیع‌های پایدار چندمتغیره را معرفی کند. او با تعمیم روش برآوردی که در حالت یک‌متغیره ارایه داده بود پارامترهای این خانواده را برآورد کرد ([۳۳] و [۳۲]). راچو و زین [۳۴] در سال ۱۹۹۳ با در نظر گرفتن فرم تابع مشخصه در مختصات قطبی برآوردگری برای برآورد اندازه طیفی و α در حالت دومتغیره ارایه کردند که در سال ۱۹۹۵ توسط چنگ و راچو [۷] به حالت چندمتغیره تعمیم داده شد. این برآورگرها بر اساس دم توزیع و برای بردارهایی که در دامنه جذب یک توزیع پایدار هستند، ارایه شد. در سال ۱۹۹۹، داویدف و پائولاسکا [۸] نیز برای بردارهایی که در دامنه جذب یک توزیع پایدار هستند و بر اساس آماره‌های مرتب برآوردگرهایی را معرفی کردند. سپس در سال ۲۰۰۰، مک‌کلاچ [۱۹] برآوردگری برای برآورد اندازه طیفی در حالت دومتغیره مطرح کرد که بر اساس تصویر داده‌ها در جهت‌های مختلف پایه‌گذاری شده بود. در این روش به جای استفاده از داده‌های دومتغیره از حالت ساده‌تر یک‌متغیره برای برآورد استفاده می‌شود. نولان و همکاران [۲۸] نیز در سال ۲۰۰۱، با همین ایده اما با شیوه‌ای متفاوت، روشی در حالت کلی برای برآورد اندازه طیفی مبتنی بر تصویر یک‌بعدی داده‌ها و همچنین تابع مشخصه تجربی معرفی کردند.

در زمینه توزیع‌های پایدار چندمتغیره، مدرس و نولان [۲۳] در سال ۱۹۹۴ روشی برای شبیه‌سازی با اندازه طیفی گسسته بیان کردند که توسط نولان برای حالت‌های خاص دیگری مانند زیرگوسی بیان شد. همچنین روش‌هایی برای تقریب و محاسبه‌ی عددی چگالی بردارهای تصادفی پایدار توسط بایزکووسکی و همکاران [۵] در سال ۱۹۹۳، نولان و راجپوت [۲۹] در سال ۱۹۹۵ و عبدالحمید و نولان [۱] در سال ۱۹۹۸ مطرح شد. نولان و پانرسکا نیز [۲۷] در سال ۱۹۹۷ روش‌هایی برای تحلیل داده‌ها برای نمونه‌های دم‌سنگین ارایه کردند. به عنوان ساختارهای وابسته به توزیع‌های پایدار چندمتغیره مانند امید ریاضی شرطی می‌توان از [۳۷] استفاده کرد.

ادامه این پایان‌نامه به صورت زیر تدوین شده است. در فصل دوم به بیان مختصری از تعاریف، قضایا و خواص توزیع های پایدار در حالت یک‌بعدی و چندبعدی می‌پردازیم. در فصل سوم برآورد براساس روش گشتاوری را شرح می‌دهیم. درفصل چهارم دو برآوردگر معرفی شده برای بردارهای در دامنه جذب یک توزیع پایدار، که بر اساس دم توزیع بیان شده است را معرفی می‌کنیم. در فصل چهارم برآورد بر اساس روش تصویر ارایه شده در حالت دومتغیره، و برآورد در حالت کلی معرفی شده توسط نولان و همکاران را بیان می‌کنیم. در فصل ششم روش‌های معرفی شده را با شبیه‌سازی و همچنین استفاده از

داده‌های واقعی مورد بررسی قرار می‌دهیم و همچنین به طور شهودی به مقایسه برآوردگرها می‌پردازیم. بنابراین مباحث اصلی این پایان‌نامه در فصل پنجم و ششم ارایه شده است.

فصل ۲

پیش نیازها

۲−۱ مقدمه

در این فصل به بررسی برخی مفاهیم اولیه که در این پایان‌نامه مورد نیازاست، می‌پردازیم. ابتدا مروری بر مفاهیم اولیه‌ی توزیع‌های $\alpha-$ پایدار خواهیم داشت. سپس به معرفی بردارهای تصادفی پایدار خواهیم پرداخت. اثبات گزاره‌ها و قضایای این فصل را می‌توان در مرجع [۳۷] یافت.

۲−۲ معرفی توزیع $\alpha-$ پایدار

توزیع $\alpha-$ پایدار توسط چهار پارامتر مشخص می‌شود که به اختصار آن را به صورت زیر نمایش می‌دهیم

$$X \sim S_\alpha(\sigma, \beta, \mu),$$

به طوری که α شاخص پایداری توزیع و $\alpha \in (۰, ۲]$، σ پارامتر مقیاس و $\sigma \in [۰, \infty)$، β پارامتر چولگی و $\beta \in [-۱, ۱]$ و μ پارامتر مکان و $\mu \in (-\infty, \infty)$ می‌باشد.

تعریف ۱.۲ عبارت‌های زیر تعریف‌های معادل از یک متغیر تصادفی پایدار X هستند.

الف) فرض کنید a و b دو عدد حقیقی مثبت، و $X_۱$ و $X_۲$ متغیرهای تصادفی مستقل و هم‌توزیع با X

باشند. مقادیر $c \in \mathbb{R}^+$ و $d \in \mathbb{R}$ وجود دارد به طوری که

$$aX_1 + bX_2 \overset{d}{=} cX + d. \qquad (۱.۲)$$

(نماد $\overset{d}{=}$ نشان دهنده هم‌توزیعی است.) معادل a و b مقدار $0 < \alpha < 2$ وجود دارد به طوری که $c^\alpha = a^\alpha + b^\alpha$.

ب) فرض کنید n یک عدد صحیح مثبت، $n \geq 2$ و $X_1, ..., X_n$ کپی‌های مستقل از X باشند، در این صورت $c_n \in \mathbb{R}^+$ و $d_n \in \mathbb{R}$ وجود دارد به طوری که اگر $\alpha \neq 1$

$$X_1 + X_2 + ... + X_n \overset{d}{=} n^{1/\alpha} X + \mu(n - n^{1/\alpha}),$$

و اگر $\alpha = 1$

$$X_1 + X_2 + ... + X_n \overset{d}{=} nX + \frac{2}{\pi}\sigma\beta n \ln n.$$

ج) X دارای دامنه جذب است. یعنی یک دنباله از متغیرهای تصادفی مستقل و هم‌توزیع $\{Y_i\}_{i\in\mathbb{N}}$، یک دنباله حقیقی مثبت $\{a_i\}_{i\in\mathbb{N}}$ و یک دنباله حقیقی $\{b_i\}_{i\in\mathbb{N}}$ وجود دارد به طوری که

$$\frac{1}{a_n}\sum_{i=1}^{n} Y_i - b_n \xrightarrow{d} X.$$

(نماد $\xrightarrow{d}$ به معنای همگرایی در توزیع است.)

د) دارای تابع مشخصه‌ای به صورت زیر می‌باشد

$$\phi_X(t) = E(\exp itX) = \begin{cases} \exp\Big\{ -\sigma^\alpha |t|^\alpha (1 - i\beta \text{sign}(t)\tan(\frac{\pi\alpha}{2})) + i\mu t\Big\} & \alpha \neq 1, \\ \exp\Big\{ -\sigma|t|(1 + i\beta\frac{2}{\pi}\text{sign}(t)\ln|t|) + i\mu t\Big\} & \alpha = 1, \end{cases} \qquad (۲.۲)$$

که

$$\text{sign}(t) = \begin{cases} -1 & t < 0, \\ 0 & t = 0, \\ 1 & t > 0. \end{cases}$$

تا کنون پارامتربندی‌های مختلفی برای این توزیع توسط افراد مختلف مطرح شده است و از آن‌جا که فرم بسته‌ای برای تابع چگالی این گروه از توزیع‌ها وجود ندارد، تنوع پارامتربندی مطالعه آن‌ها را راحت‌تر می‌سازد. به عنوان مثال می‌توان از پارامتربندی S° نام برد که توسط زولوتارو [۳۸] مطرح شده است. فرض کنید $X^\circ \sim S^\circ_\alpha(\sigma, \beta, \mu^\circ)$ در این صورت

$$E(\exp itX^{\circ}) = \begin{cases} \exp\Big\{-\sigma^{\alpha}|t|^{\alpha}[1+i\beta\mathrm{sign}(t)\tan(\frac{\pi\alpha}{2})((\sigma|t|)^{1-\alpha}-1)]+i\mu^{\circ}t\Big\} & \alpha\neq 1,\\ \exp\Big\{-\sigma|t|[1+i\beta\frac{2}{\pi}\mathrm{sign}(t)\ln(\sigma|t|)]+i\mu^{\circ}t\Big\} & \alpha=1.\end{cases}$$

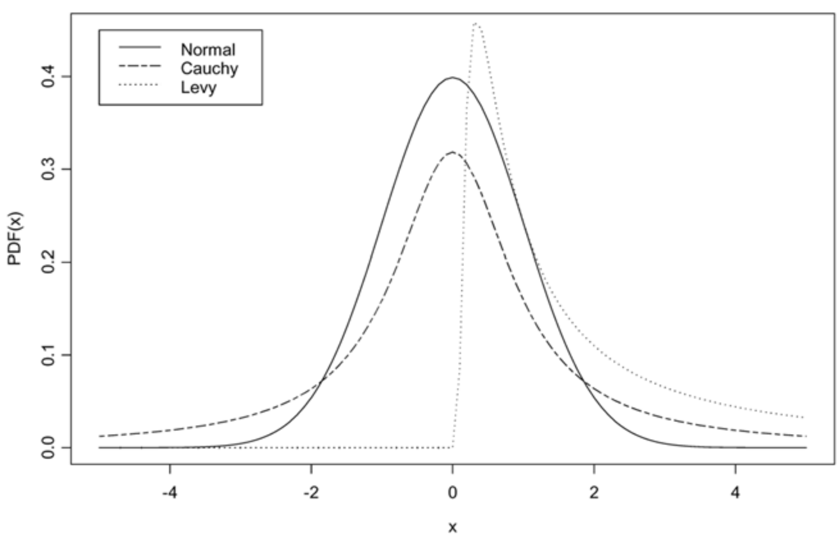
در این پارامتربندی تابع مشخصه و بنابراین تابع چگالی و تابع توزیع در هر چهار پارامتر تواماً پیوسته می‌باشند.

پارامترهای α، β و σ در هر دو پارامتربندی S و S° معنای یکسان دارند، در حالی که پارامتر مکان آن‌ها دارای رابطه زیر می‌باشد

$$\begin{cases} \mu=\mu^{\circ}-\beta(\tan\frac{\alpha\pi}{2})\sigma & \alpha\neq 1,\\ \mu=\mu^{\circ}-\beta\frac{2}{\pi}\sigma\ln\sigma & \alpha=1.\end{cases}$$

در این پایان‌نامه پارامتربندی که توسط ساموردنیسکی و تاهو [۳۷] استفاده شده است (پارامتربندی S) را به کار می‌بریم.

چگالی احتمال متغیرهای تصادفی $\alpha-$ پایدار موجود و پیوسته است اما به جز تعداد بسیار کمی از این توزیع‌ها، بقیه فرم مشخص بسته‌ای ندارند ([۳۸]). چند توزیع معروف که در این کلاس از توزیع‌ها دارای چگالی احتمال مشخص می‌باشند عبارتند از توزیع گوسی یا نرمال با $S_2(\sigma,0,\mu)=N(\mu,2\sigma)$، توزیع کشی[۱] با $S_1(\sigma,0,\mu)=\mathrm{C}(\sigma,\mu)$، توزیع لوی[۲] با $S_{1/2}(\sigma,1,\mu)=\mathrm{L}(\sigma,\mu)$ و توزیع تباهیده در μ با $S_{\alpha}(0,0,\mu)$، $0<\alpha<2$.



شکل ۲—۱: نمودار چگالی $\mathrm{L}(1,0)$، $\mathrm{C}(1,0)$ و $N(0,1)$.

[۱] Cauchy
[۲] Levy

تعریف ۲.۲ متغیر تصادفی X دارای توزیع دم سنگین از طرف راست است اگر عدد $\alpha > ۰$ وجود داشته باشد به طوری که

$$P[X > x] = x^{-\alpha} L(x),$$

که در آن L تابعی با وردش تدریجی است به این معنا که برای هر $x > ۰$،

$$\lim_{t \to \infty} \frac{L(tx)}{L(t)} = ۱. \qquad (۳.۲)$$

(تعریف توزیع دم سنگین برای دم سمت چپ و دوطرفه به طور مشابه می‌باشد.)

تذکر ۱.۲ متغیر تصادفی $X \sim S_\alpha(\sigma, \beta, \mu)$ متقارن است اگر $\beta = ۰$ و $\mu = ۰$، و متقارن در اطراف μ است اگر $\beta = ۰$.

توزیع $S_\alpha(\sigma, \beta, \mu)$ چوله به راست گفته می‌شود اگر $\beta > ۰$ و چوله به چپ گفته می‌شود اگر $\beta < ۰$ باشد. اگر $\beta = ۱$ باشد این توزیع کاملاً چوله به راست یا مثبت است و اگر $\beta = -۱$ باشد این توزیع کاملاً چوله به چپ یا منفی است.

۲–۳ برخی از خواص توزیع $\alpha-$ پایدار

در این بخش برخی از خواص توزیع‌های پایدار که در این پایان‌نامه مورد نیاز هستند مطرح می‌شود.

گزاره ۱.۲ اگر $X \sim S_\alpha(\sigma, \beta, \mu)$ و a یک عدد ثابت غیر صفر و حقیقی باشد، آن‌گاه

$$\begin{aligned} aX &\sim S_\alpha(|a|\sigma, \operatorname{sign}(a)\beta, a\mu), && \alpha \neq ۱, \\ aX &\sim S_\alpha(|a|\sigma, \operatorname{sign}(a)\beta, a\mu - \frac{۲}{\pi}\, a(\ln|a|)\sigma\beta), && \alpha = ۱. \end{aligned}$$

گزاره ۲.۲ اگر $X_۱$ و $X_۲$ متغیرهای تصادفی مستقل با $X_i \sim S_\alpha(\sigma_i, \beta_i, \mu_i)$، آن‌گاه $X_۱ + X_۲ \sim S_\alpha(\sigma, \beta, \mu)$ به طوری که

$$\sigma = (\sigma_۱^\alpha + \sigma_۲^\alpha)^{\frac{۱}{\alpha}}, \quad \beta = \frac{\beta_۱\sigma_۱^\alpha + \beta_۲\sigma_۲^\alpha}{\sigma_۱^\alpha + \sigma_۲^\alpha}, \quad \mu = \mu_۱ + \mu_۲.$$

گزاره ۳.۲ فرض کنید $X \sim S_\alpha(\sigma,\beta,\mu)$ با $0 < \alpha < 2$ باشد، آن‌گاه

$$E|X|^p < \infty, \quad \forall p, \quad 0 < p < \alpha,$$
$$E|X|^p = \infty, \quad \forall p, \quad p \geq \alpha.$$

متغیرهای تصادفی $\alpha-$ پایدار با $\alpha < 2$، دارای گشتاور دوم نامتناهی هستند، بنابراین بسیاری از تکنیک‌هایی که در حالت گوسی معتبر است برای آن‌ها کاربردی ندارد. همچنین زمانی که $\alpha \leq 1$ چون $E|X| = \infty$، عملاً استفاده از امید ریاضی غیرممکن می‌باشد. پس قابل ذکر است که پارامترهای σ و μ انحراف استاندارد و میانگین نمی‌باشند.

گزاره ۴.۲ فرض کنید $X \sim S_\alpha(\sigma,\beta,\mu)$ با $0 < \alpha < 2$ باشد، آن‌گاه

$$\lim_{x\to\infty} x^\alpha P(X > x) = C_\alpha(1+\beta)\sigma^\alpha, \tag{۴.۲}$$

و

$$\lim_{x\to\infty} x^\alpha P(X < -x) = C_\alpha(1-\beta)\sigma^\alpha,$$

به طوری که

$$C_\alpha = \left(2\int_0^\infty x^{-\alpha}\sin x\, dx\right)^{-1}.$$

رابطه (۴.۲) با $\sigma = 1$ و $\mu = 0$ را در نظر بگیرید. برای x به اندازه‌ی کافی بزرگ، با مشتق گرفتن از آن می‌توان نشان داد

$$f(x;\alpha,\beta) \sim \alpha(1+\beta)C_\alpha x^{-(1+\alpha)}.$$

بنابراین با در نظر گرفتن توزیع پارتو به صورت

$$f_{\text{Pareto}}(x;\alpha,\beta) = \alpha(1+\beta)C_\alpha x^{-(1+\alpha)}, \qquad \overline{F}_{\text{Pareto}}(x;\alpha,\beta) = (1+\beta)C_\alpha x^{-\alpha},$$

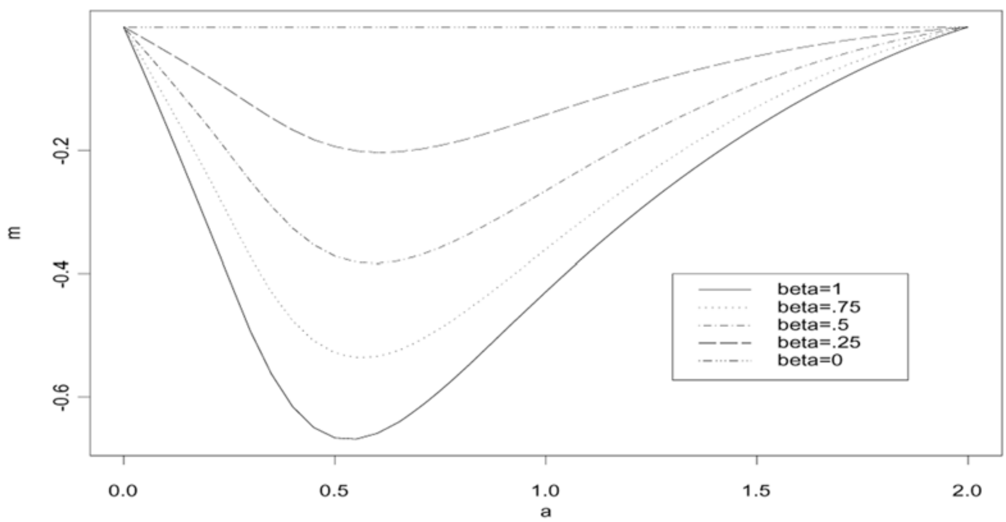
با مقیاس در نظر گرفته شده، مشاهده می‌کنیم وقتی $x \to \infty$، $f(x;\alpha,\beta) \sim f_{\text{Pareto}}(x;\alpha,\beta)$. (نتایج به طور مشابه برای دم چپ، با توجه به این که $P(X < -x) \sim C_\alpha(1-\beta)x^{-\alpha}$، برقرار می‌باشد.)

حال توزیع نرمال $N(0,2) = S_2(1,0,0)$ را در نظر بگیرید، وقتی $x \to \infty$،

$$P(X > x) = P(X < -x) \sim \frac{\exp(-x^2/4)}{2x\sqrt{\pi}}.$$

پس وقتی x به اندازه کافی بزرگ باشد، دم‌های توزیع نرمال با سرعت $\frac{\exp(-x^{۲})}{|x|}$ کاهش می‌یابند، در صورتی که وقتی $\alpha < ۲$، دم‌ها خیلی کندتر با سرعت یک چندجمله‌ای به صورت $|x|^{-\alpha}$ کاهش می‌یابد.

تذکر ۲.۲ توزیع‌های پایدار تک‌مدی هستند. به طور مثال برای یک توزیع پایدار متقارن $(\beta = ۰)$ در هر پارامتربندی مد در $x = ۰$ قرار دارد. هر چند وقتی $\beta \neq ۰$ مد توزیع‌های پایدار جابه‌جا می‌شود. به طور عددی مشاهده شده است زمانی که $\beta > ۰$، $P(Z > m(\alpha,\beta)) > P(Z < m(\alpha,\beta))$، و زمانی که $\beta < ۰$، $P(Z > m(\alpha,\beta)) < P(Z < m(\alpha,\beta))$. مد یک توزیع $S_{\alpha}^{۰}(۱,\beta,۰)$ به طور یکنواخت کراندار است، در حقیقت $|m(\alpha,\beta)| \leq ۱$ (شکل ۲—۲). اما مد $S_{\alpha}(۱,\beta,۰)$ در رابطه $m(\alpha,\beta) + \beta \tan\frac{\alpha\pi}{۲}$ صدق می‌کند، که برای α های نزدیک به یک خیلی بزرگ می‌شود. (برای اطلاع از مد در پارامتربندی‌های مختلف می‌توان به مرجع [۲۵] رجوع کرد.)



شکل ۲—۲: مکان مد در تابع چگالی $S_{\alpha}^{۰}(۱,\beta,۰)$.

۲—۴ معرفی برخی روش‌های برآورد پارامترهای توزیع یک‌متغیره پایدار

چندین روش برای برآورد پارامترهای پایدار ارایه شده است که با توجه به لزوم استفاده از آن‌ها در فصل‌های بعدی در این بخش اشاره مختصری به برخی از آن‌ها خواهیم داشت.

ساده‌ترین روش برای برآورد شاخص پایداری α، رسم تابع توزیع تجربی داده‌های مشاهده شده روی یک مقیاس لگاریتم—لگاریتم می‌باشد. رابطه (۴.۲) را در نظر بگیرید. برای x به اندازه کافی بزرگ، با

گرفتن لگاریتم از دو طرف رابطه، معادله‌ی خطی

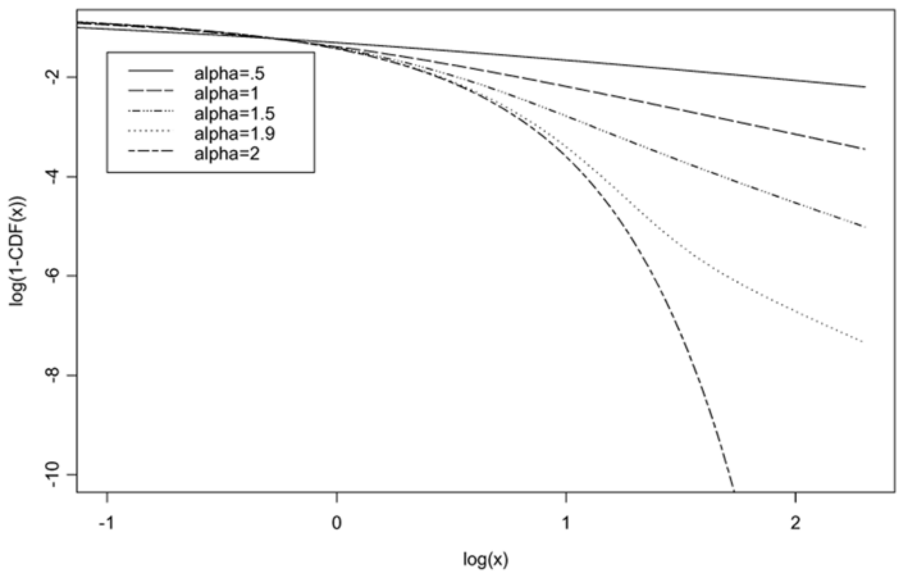
$$\log P(X > x) \sim \log(C_\alpha(۱ + \beta)\sigma^\alpha) - \alpha \log x,$$

با شیب منفی α به دست می‌آید (شکل ۲—۳). پس با در نظر گرفتن

$$y_i = \frac{\text{تعداد}(X_i > x_i)}{n} \simeq P(X > x_i),$$

با اجرای یک رگرسیون خطی ساده، می‌توان α را از رابطه‌ی زیر به دست آورد

$$\hat{\alpha} = -\frac{\sum_{i=۱}^{k} \log x_i(\log y_i - \overline{\log y_i})}{\sum_{i=۱}^{k}(\log x_i - \overline{\log x_i})^۲}.$$



شکل ۲—۳: نمودار دم توابع توزیع تجمعی $\alpha-$ پایدار با $S_\alpha(۱, ۰, ۰)$ و α های مختلف در مقیاس لگاریتم—لگاریتم.

این روش با وجود سادگی و سرراست بودن، در عمل معتبر نیست. مشکل اصلی این است که مشخص نیست رفتار پارتویی دم چه موقع اتفاق می‌افتد و چه موقع در موقعیت دم هستیم. فوفاک و نولان [۱۵] نشان دادند که این مقدار تابع پیچیده‌ای از α و β می‌باشد. به عنوان مثال وقتی $\alpha \to ۲$، کمتر از ۰/۱ از مقادیر انتهایی توزیع دارای رفتار حدی پارتویی می‌باشد. این روش برای نمونه‌های بسیار بزرگ که انتهایی‌ترین مقادیر آن‌ها در نظر گرفته شده می تواند جواب قابل قبولی داشته باشد.

روش دیگری که برای برآورد پارامترهای پایدار ارایه شده است بر اساس چندک‌های توزیع می‌باشد که در سال ۱۹۷۱، در حالت متقارن و با محدودیت‌هایی روی α توسط فاما و رول معرفی، و در سال ۱۹۸۶

توسط مک‌کلاچ [۲۰] به حالت کلی‌تر آن تعمیم داده شد. مک‌کلاچ برآوردگرهایی از هر چهار پارامتر را برای $۰/۶ \leq \alpha \leq ۲$ معرفی کرد.

چون تابع مشخصه توزیع‌های پایدار دارای فرم مشخصی است چندین محقق برآوردگرهایی را بر این اساس بنا نهادند. به نظر می‌رسد پرس [۳۳] اولین کسی باشد که این کار را انجام داده باشد. چندین برآوردگر دیگر نیز بر این اساس به دست آمده‌اند که از آن جمله می‌توان به کار پائولسون و همکاران [۳۱] و کوتروولیس [۱۷] اشاره کرد که اولی بر اساس مینیمم کردن رابطه

$$I = \int_{-\infty}^{\infty} |\hat{\phi}(u) - \phi(u)|^{۲} \ e^{-u^{۲}} \ du,$$

و دومی بر اساس روش رگرسیونی، پارامترها را برآورد می‌کند.

برآوردگرهایی نیز بر اساس روش درستنمایی ماکزیمم ارایه شده است. اولین تلاش برای برآورد پارامترهای پایدار را می‌توان در کار دوموشل [۱۲] یافت که در آن داده‌ها را گروهبندی و از میانگین‌های آن‌ها برای محاسبه چگالی استفاده کرده است. در این روش ابتدا چگالی با استفاده از معکوس تابع مشخصه [۲۲] یا از طریق نمایش انتگرال زولوتارو [۲۴] به دست می‌آید. سپس از تقریب‌های عددی برای ماکزیمم کردن رابطه درستنمایی استفاده می‌شود. در مورد روش‌های دیگر در این زمینه می‌توان از کار براسن و یانگ [۴] نام برد که در حالت متقارن با استفاده از فرمول چگالی ارایه شده توسط زولوتارو [۳۸]، یک فرمول مستقیم برای برآورد ارایه کردند. برانت [۳] نیز یک روش برای تقریب درستنمایی با استفاده از تابع مشخصه معرفی کرده است.

۲—۴—۱ روش چندک‌های مک‌کلاچ

فرض کنید $X_{۱}, ..., X_n$ یک نمونه مستقل و هم‌توزیع از $S_\alpha(\sigma, \beta, \mu)$، x_p، p امین چندک جامعه یعنی $S_\alpha(\sigma, \beta, \mu)(x_p) = p$ و $\hat{x}_p$ چندک نمونه مطابق با آن یعنی $F_n(\hat{x}_p) = p$ باشد. برآورد چندکی از $۱۰۰p$ درصد مشاهدات مرتب شده نمونه، اگر این مقدار در نمونه باشد، به دست می آید. در غیر این صورت برآوردگر با درون‌یابی خطی به دست می‌آید. مقادیر x_i توسط $\hat{x}_{q(i)}$ که $q(i) = \frac{۲i-۱}{۲n}$، مشخص می‌شوند و برای $q(i) < p < q(i+۱)$، $\hat{x_p}$ از دو مقدار متناظر با $q(i)$ مجاورش درون‌یابی می‌شود.

در این روش

$$\upsilon_\alpha = \frac{x_{۰/۹۵} - x_{۰/۰۵}}{x_{۰/۷۵} - x_{۰/۲۵}} \ , \ \ \upsilon_\beta = \frac{x_{۰/۹۵} + x_{۰/۰۵} - ۲x_{۰/۵۰}}{x_{۰/۹۵} - x_{۰/۰۵}},$$

به عنوان توابعی از α و β معرفی می‌شوند که مستقل از σ و μ می‌باشند. با معکوس کردن توابع، α و β بر حسب υ_α و υ_β به دست می آیند

$$\alpha = \psi_۱(\upsilon_\alpha, \upsilon_\beta) \ \ , \ \ \beta = \psi_۲(\upsilon_\alpha, \upsilon_\beta).$$

پس می‌توان با مقادیر $\hat{\upsilon}_\alpha$ و $\hat{\upsilon}_\beta$ به دست آمده از مقادیر نمونه‌ای، برآوردهای $\hat{\alpha}$ و $\hat{\beta}$ را به دست آورد.

روابط زیر را در نظر بگیرید

$$\upsilon_\sigma = \frac{x_{۰/۷۵} - x_{۰/۲۵}}{\sigma},$$

$$\upsilon_\zeta = \frac{\zeta - x_{۰/۵}}{\sigma},$$

به طوری که

$$\zeta = \begin{cases} \mu + \beta\sigma \tan\frac{\pi\alpha}{۲} & \alpha \neq ۱, \\ \mu & \alpha = ۱. \end{cases} \tag{۵.۲}$$

با در نظر گرفتن برآوردگرهای $\hat{\alpha}$، $\hat{\beta}$، $\hat{x}_{۰/۲۵}$ و $\hat{x}_{۰/۷۵}$ می‌توان σ را به صورت زیر برآورد کرد

$$\hat{\sigma} = \frac{\hat{x}_{۰/۷۵} - \hat{x}_{۰/۲۵}}{\hat{\upsilon}_\sigma}.$$

پس داریم

$$\hat{\zeta} = \hat{x}_{۰/۵} + \hat{\sigma}\hat{\upsilon}_\zeta,$$

که به وسیله آن می‌توان μ را از رابطه (۵.۲) برآورد کرد.

مک‌کلاچ [۲۰] جداولی که α و β را به عنوان تابعی از υ_α و υ_β نشان می‌دهد تنظیم کرده است. با استفاده از α و β به دست آمده از این جداول می توان مقادیر υ_σ و υ_ζ مرتبط با σ و μ را از جداول مربوطه‌اش به دست آورد. این مقاله همچنین شامل واریانس و کواریانس مجانبی برآوردگرهای معرفی شده می‌باشد.

۲-۴-۲ روش گشتاوری پرس

نمونه تصادفی مستقل و هم‌توزیع $X_۱, ..., X_n$ را در نظر بگیرید، تابع مشخصه نمونه به صورت زیر تعریف می‌شود

$$\hat{\phi}_X(t) = \frac{۱}{n}\sum_{j=۱}^{n} e^{itX_j}.$$