

توزیع لامبدای تعمیم یافته و ویژگی های آن

الهه کدخدای^۱، مرتضی محمدی^۲، غلامرضا محتشمی برزادران^۳

تاریخ دریافت: ۱۳۹۴/۱۱/۲۱

تاریخ پذیرش: ۱۳۹۶/۱/۳۰

چکیده:

توزیع لامبدای تعمیم یافته^۴ (GLD)، تعمیمی از توزیع تک پارامتری توکی است که انعطاف پذیری بسیار زیادی در مدل بندی اطلاعات و داده های آماری دارد. در این مقاله، نخست دو شکل پارامتری متفاوت از توزیع GLD را ارائه و ویژگی های این توزیع را بررسی خواهیم کرد. سپس چهار روش گشتاوری، صدکی، starship و ماکسیمم درست نمایی را برای برآورد پارامترهای توزیع GLD ارائه می کنیم. در انتها به کمک آزمون کولموگروف-اسمیرنوف به مقایسه دو شکل پارامتری توزیع GLD و چهار روش برآورد پارامترهای این توزیع می پردازیم و این توزیع را به داده های بورس اوراق بهادار تهران برازش می دهیم.

واژه های کلیدی: توزیع لامبدای تعمیم یافته، گشتاورها، برآورد پارامترها.

۲ توزیع لامبدای تعمیم یافته

۱ مقدمه

خانواده توزیع لامبدای تعمیم یافته (GLD) به دلیل تطبیق پذیری زیاد مورد توجه قرار گرفته است.

این توزیع قادر است تا با دقت زیاد بر توزیع های پر کاربرد پیوسته نظیر نرمال، لگ نرمال، گاما، یکنواخت و وایبل منطبق شود. انعطاف پذیری این توزیع علاوه بر برآورد توزیع های پیوسته باعث انطباق آن بر بافت نگاشت داده ها و برآورد نوع توزیع آنها نیز می شود. از این رو، به عنوان ابزار قدرتمندی برای تحقیق در حوزه های مختلف نظیر برآورد پارامتر، انطباق توزیع ها بر داده ها و همچنین تحقیقات شبیه سازی که مبتنی بر تولید داده هستند، کاربرد دارد.

توزیع لامبدای تعمیم یافته (GLD) تعمیمی از توزیع تک پارامتری است [۸]. اولین روش پارامتری کردن این تعمیم چهار پارامتری را که به RS^۵ مشهور است پیشنهاد کردند و با تابع چندکی اش مشخص می شود که برابر با

$$Q_{RS}(u) = \lambda_1 + \frac{u^{\lambda_2} - (1-u)^{\lambda_2}}{\lambda_2} \quad (1)$$

برازش یک توزیع احتمال به داده ها عملی مهم در تحلیل داده های آماری است. داده هایی که قرار است مدل بندی شوند ممکن است شامل رخدادهای مشاهده شده مانند سری زمانی مالی یا نتایج شبیه سازی شده مانند مجموعه ای از مقادیر سود و زیان برای سهام باشد. هنگام برازش به داده ها در ابتدا باید یک رده یا یک خانواده عمومی از توزیع ها را انتخاب کنیم و سپس مقادیر پارامترهای توزیعی که بهترین تطبیق با داده های مشاهده شده را دارند، بیابیم.

از طرفی چون توزیع لامبدای تعمیم یافته اشکال مختلفی را در بر می گیرد (تک مدی، یکنوا، S-شکل و U-شکل)، در مدل بندی گستره وسیعی از داده های آماری انعطاف پذیری زیادی دارد. به هر حال به دلیل تطبیق پذیری زیاد GLD، دست یابی به پارامترهای مناسب برای آن می تواند مسئله ای چالش برانگیز باشد. در این مقاله به مقایسه چند روش برآورد پارامتر و آزمون نیکویی برازش برای ارزیابی پارامترهای به دست آمده می پردازیم.

^۱ دانش آموخته کارشناسی ارشد آمار، دانشگاه زابل، ایران

^۲ عضو هیئت علمی گروه آمار، دانشگاه زابل، ایران

^۳ عضو هیئت علمی گروه آمار، دانشگاه فردوسی مشهد، ایران

^۴ generalized lambda distribution

^۵ Ramberg and Schmeiser

برقرار باشد [۳]. اگر X دارای توزیع $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ با $\lambda_4 > -\frac{1}{4}$ و $\lambda_3 > -\frac{1}{4}$ باشد، میانگین، واریانس، چولگی و کشیدگی RS به ترتیب به صورت

$$\alpha_1 = \mu = E(X) = \lambda_1 + \frac{A}{\lambda_2} \quad (۶)$$

$$\alpha_2 = \sigma^2 = E((X - \mu)^2) = \frac{B - A^2}{\lambda_2^2} \quad (۷)$$

$$\alpha_3 = \frac{E((X - E(X))^3)}{\sigma^3} = \frac{C - 3AB + 2A^3}{\lambda_2^3 \sigma^3} \quad (۸)$$

$$\alpha_4 = \frac{E((X - E(X))^4)}{\sigma^4} = \frac{D - 4AC + 6A^2B - 3A^4}{\lambda_2^4 \sigma^4} \quad (۹)$$

به دست می آیند [۳]. که در آنها

$$A = \frac{1}{1 + \lambda_3} - \frac{1}{1 + \lambda_4}, \quad (۱۰)$$

$$B = \frac{1}{1 + 2\lambda_3} + \frac{1}{1 + 2\lambda_4} - 2\beta(1 + \lambda_3, 1 + \lambda_4), \quad (۱۱)$$

$$C = \frac{1}{1 + 3\lambda_3} - \frac{1}{1 + 3\lambda_4} - 3\beta(1 + 2\lambda_3, 1 + \lambda_4) + 3\beta(1 + \lambda_3, 1 + 2\lambda_4), \quad (۱۲)$$

$$D = \frac{1}{1 + 4\lambda_3} + \frac{1}{1 + 4\lambda_4} - 4\beta(1 + 3\lambda_3, 1 + \lambda_4) + 6\beta(1 + 2\lambda_3, 1 + 2\lambda_4) - 4\beta(1 + \lambda_3, 1 + 3\lambda_4) \quad (۱۳)$$

می باشد. چهار گشتاور اولیه FMKL نیز از عبارات

$$\alpha_1 = \lambda_1 - \frac{1}{\lambda_2 \lambda_3} + \frac{1}{\lambda_2 \lambda_4} + \frac{1}{\lambda_2} \left(\frac{1}{\lambda_3(\lambda_3 + 1)} - \frac{1}{\lambda_4(\lambda_4 + 1)} \right),$$

$$\alpha_2 = \frac{1}{\lambda_2^2} (\nu_2 - \nu_1^2),$$

$$\alpha_3 = \frac{\nu_3 - 3\nu_1 \nu_2 + 2\nu_1^3}{(\nu_2 - \nu_1)^3},$$

$$\alpha_4 = \frac{\nu_4 - 4\nu_1 \nu_3 + 6\nu_1^2 \nu_2 - 3\nu_1^4}{(\nu_2 - \nu_1)^4}$$

^۶Freimer, Mudholkar Kollia and Lin (FMKL)

است و تابع چگالی آن در نقطه $x = F^{-1}(u)$ از رابطه

$$f_{RS}(x) = \frac{\lambda_2}{\lambda_3 u^{\lambda_3 - 1} + \lambda_4 (1 - u)^{\lambda_4 - 1}}, \quad u \in [0, 1] \quad (۲)$$

به دست می آید. [۷] ذکر کردند که ترکیبات معلومی از λ_3 و λ_4 هستند که تابع توزیع فوق در آنها یک توزیع احتمالی معتبر نمی باشد. برای اطلاعات بیشتر، به [۳] مراجعه کند. برای توسیع تکیه گاه (λ_3, λ_4) روی همه صفحه (λ_3, λ_4) ، [۱] شکل پارامتری متفاوتی را برای GLD مطرح کردند که آن را با FMKL^۶ نشان می دهیم و به صورت

$$Q_{FMKL}(u) = \lambda_1 + \frac{1}{\lambda_2} \left(\frac{u^{\lambda_3} - 1}{\lambda_3} - \frac{(1 - u)^{\lambda_4} - 1}{\lambda_4} \right) \quad (۳)$$

تعریف می شود و به ازای همه مقادیر روی صفحه (λ_3, λ_4) قابل تعریف است و تابع چگالی آن برابر با

$$f_{FMKL}(x) = \frac{\lambda_2}{u^{\lambda_3 - 1} + (1 - u)^{\lambda_4 - 1}}, \quad 0 \leq u \leq 1 \quad (۴)$$

است [۱]. اشکال به دست آمده از (۳) را به صورت زیر دسته بندی کردند:

رده ۱: $(\lambda_3 < 1, \lambda_4 < 1)$: که چگالی های تک مدی با دم های پیوسته را شامل می شود.

رده ۲: $(\lambda_3 > 1, \lambda_4 < 1)$: چگالی های یکنواخت (مانند توزیع نمایی یا خی دو) که دم چپ توزیع در آنها بریده می باشد.

رده ۳: $(1 < \lambda_3 < 2, 1 < \lambda_4 < 2)$: چگالی های U -شکل که هر دو دم آنها بریده شده است.

رده ۴: $(\lambda_3 > 2, 1 < \lambda_4 < 2)$: چگالی های S -شکل که هر دو دم آنها بریده شده است.

رده ۵: $(\lambda_3 > 2, \lambda_4 > 2)$: چگالی های تک مدی که هر دو دم آنها بریده شده است.

۳ گشتاورهای GLD

گشتاور مرتبه k ، $GLD(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$ وجود دارد اگر و تنها اگر رابطه

$$\min(\lambda_3, \lambda_4) > -\frac{1}{k} \quad (۵)$$

به دست می‌آیند، که در آنها

$$\hat{\alpha}_1 = \bar{X} = \sum_{i=1}^n \frac{x_i}{n},$$

$$\hat{\alpha}_2 = \sum_{i=1}^n \frac{(x_i - \bar{x})^2}{n},$$

$$\hat{\alpha}_3 = \sum_{i=1}^n \frac{(x_i - \bar{x})^3}{n\hat{\sigma}_2^3},$$

$$\hat{\alpha}_4 = \sum_{i=1}^n \frac{(x_i - \bar{x})^4}{n\hat{\sigma}_2^4}.$$

حال برای به دست آوردن $\lambda_1, \lambda_2, \lambda_3$ و λ_4 در هر دو روش RS و FMKL دستگاه معادلات زیر را تشکیل می‌دهیم:

$$\alpha_i = \hat{\alpha}_i, \quad i = 1, 2, 3, 4.$$

در معادلات (۱۰)، (۱۱)، (۱۲) و (۱۳) مشاهده شد که A, B, C و D به λ_1 و λ_2 بستگی ندارند. لذا برای به دست آوردن پارامترهای RS با قرار دادن A, B, C و D حل دستگاه آسان می‌شود و در نتیجه α_3 و α_4 فقط به λ_3 و λ_4 بستگی دارند. در این صورت اگر بتوان λ_3 و λ_4 را از حل دستگاه معادلات

$$\alpha_3 = \hat{\alpha}_3, \quad \alpha_4 = \hat{\alpha}_4 \quad (18)$$

که شامل دو معادله بر حسب دو متغیر λ_3 و λ_4 می‌باشد، به دست آورد، با استفاده از عبارت‌های (۷) و (۶)، برآورد پارامترهای λ_1 و λ_2 به دست می‌آید.

به روش مشابه، برای به دست آوردن پارامترهای FMKL با توجه به معادلات (۱۴)، (۱۵)، (۱۶) و (۱۷) مشاهده می‌شود که $\alpha_1, \alpha_2, \alpha_3$ و α_4 به λ_1 و λ_2 بستگی ندارند و در نتیجه α_3 و α_4 فقط به λ_3 و λ_4 بستگی دارند. در این صورت اگر بتوان λ_3 و λ_4 را از حل دستگاه (۱۸) به دست آورد، λ_1 و λ_2 به ترتیب با استفاده از عبارت‌های زیر محاسبه می‌شوند:

$$\lambda_2 = \frac{(\nu_2 - \nu_1^2)^{\frac{1}{2}}}{\hat{\sigma}},$$

$$\lambda_1 = \hat{\mu} + \frac{1}{\lambda_2} \left(\frac{1}{\lambda_3 + 1} - \frac{1}{\lambda_4 + 1} \right).$$

۲.۴ روش صدک‌ها

در روش صدک‌ها به جای تطبیق گشتاورهای نمونه‌ای و گشتاورهای جامعه‌ای، از صدک‌های نمونه و جامعه استفاده می‌کنیم. این روش، بخش بیشتری از خانواده GLD را برای برازش به داده‌ها در دسترس قرار می‌دهد و مشکلات محاسباتی را

$$\nu_1 = \frac{1}{\lambda_3(\lambda_3 + 1)} - \frac{1}{\lambda_4(\lambda_4 + 1)}, \quad (14)$$

$$\nu_2 = \frac{1}{\lambda_3^2(2\lambda_3 + 1)} + \frac{1}{\lambda_4^2(2\lambda_4 + 1)} - \frac{2}{\lambda_3\lambda_4}\beta(\lambda_3 + 1, \lambda_4 + 1), \quad (15)$$

$$\nu_3 = \frac{1}{\lambda_3^3(3\lambda_3 + 1)} - \frac{1}{\lambda_4^3(3\lambda_4 + 1)} - \frac{3}{\lambda_3^2\lambda_4}\beta(2\lambda_3 + 1, \lambda_4 + 1) + \frac{3}{\lambda_3\lambda_4^2}\beta(\lambda_3 + 1, 2\lambda_4 + 1), \quad (16)$$

$$\nu_4 = \frac{1}{\lambda_3^4(4\lambda_3 + 1)} + \frac{1}{\lambda_4^4(4\lambda_4 + 1)} + \frac{6}{\lambda_3^2\lambda_4^2}\beta(2\lambda_3 + 1, 2\lambda_4 + 1) - \frac{4}{\lambda_3^3\lambda_4}\beta(3\lambda_3 + 1, \lambda_4 + 1) - \frac{4}{\lambda_3\lambda_4^3}\beta(\lambda_3 + 1, 3\lambda_4 + 1). \quad (17)$$

می‌باشد.

۴ روش‌های برآورد پارامتر

همان‌طور که اشاره شد هدف، برازش GLD به یک مجموعه داده است. برای این منظور در ابتدا باید با استفاده از مقادیر نمونه پارامترهای توزیع را برآورد کنیم. روش گشتاورها، صدک‌ها، starship و روش ماکسیمم درست‌نمایی را ارائه خواهیم کرد.

۱.۴ روش گشتاورها

[۶] روش گشتاورها را برای برآورد پارامترهای GLD مطرح کردند، بدین ترتیب که برای برازش یک GLD به یک مجموعه داده، در ابتدا با برابر قرار دادن $\alpha_1, \alpha_2, \alpha_3$ و α_4 با $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ و $\hat{\alpha}_4$ آماره‌های متناظر نمونه‌ای و حل معادلات، مقادیر $\lambda_1, \lambda_2, \lambda_3$ و λ_4 را به دست می‌آوریم.

برای داده‌های X_1, X_2, \dots, X_n ، گشتاورهای متناظر نمونه‌ای برای $\alpha_1, \alpha_2, \alpha_3$ و α_4 را با $\hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3$ و $\hat{\alpha}_4$ نشان می‌دهیم که عبارت‌اند از:

با توجه به GLD تعریف شده در رابطه (۱)، آماره های متناظر با $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$ و $\hat{\rho}_4$ را تعریف می کنیم:

$$\begin{aligned}\rho_1 &= Q\left(\frac{1}{\lambda_1}\right) \\ &= \lambda_1 + \frac{\left(\frac{1}{\lambda_1}\right)^{\lambda_2} - \left(\frac{1}{\lambda_1}\right)^{\lambda_2}}{\lambda_2}, \\ \rho_2 &= Q(1-u) - Q(u) \\ &= \frac{(1-u)^{\lambda_2} - u^{\lambda_2} + (1-u)^{\lambda_2} - u^{\lambda_2}}{\lambda_2}, \\ \rho_3 &= \frac{Q\left(\frac{1}{\lambda_1}\right) - Q(u)}{Q(1-u) - Q\left(\frac{1}{\lambda_1}\right)} \\ &= \frac{(1-u)^{\lambda_2} - u^{\lambda_2} + \left(\frac{1}{\lambda_1}\right)^{\lambda_2} - \left(\frac{1}{\lambda_1}\right)^{\lambda_2}}{(1-u)^{\lambda_2} - u^{\lambda_2} + \left(\frac{1}{\lambda_1}\right)^{\lambda_2} - \left(\frac{1}{\lambda_1}\right)^{\lambda_2}}, \\ \rho_4 &= \frac{Q\left(\frac{1}{\lambda_1}\right) - Q\left(\frac{1}{\lambda_2}\right)}{\rho_2} \\ &= \frac{\left(\frac{1}{\lambda_1}\right)^{\lambda_2} - \left(\frac{1}{\lambda_2}\right)^{\lambda_2} + \left(\frac{1}{\lambda_1}\right)^{\lambda_2} - \left(\frac{1}{\lambda_2}\right)^{\lambda_2}}{(1-u)^{\lambda_2} - u^{\lambda_2} + (1-u)^{\lambda_2} - u^{\lambda_2}}.\end{aligned}$$

از تعاریف ρ_1, ρ_2, ρ_3 و ρ_4 نتیجه می شود که $-\infty \leq \rho_1 \leq \infty$ ، $\rho_2 \geq 0$ ، $\rho_3 \geq 0$ و $\rho_4 \leq 1$. برای برآورد پارامترهای GLD باید دستگاه معادلات

$$\hat{\rho}_i = \rho_i, \quad i = 1, 2, 3, 4 \quad (19)$$

را بر حسب $\lambda_1, \lambda_2, \lambda_3$ و λ_4 حل کنیم. پس از تشکیل دستگاه (۱۹) مشاهده می کنیم که زیردستگاه

$$\hat{\rho}_3 = \rho_3, \quad \hat{\rho}_4 = \rho_4$$

فقط بر حسب λ_3 و λ_4 است، لذا اگر بتوان مقادیر λ_3 و λ_4 را از حل این زیردستگاه به دست آورد، از معادلات

$$\begin{aligned}\lambda_2 &= \frac{(1-u)^{\lambda_2} - u^{\lambda_2} + (1-u)^{\lambda_2} - u^{\lambda_2}}{\hat{\rho}_2}, \\ \lambda_1 &= \hat{\rho}_1 - \frac{\left(\frac{1}{\lambda_1}\right)^{\lambda_2} - \left(\frac{1}{\lambda_1}\right)^{\lambda_2}}{\lambda_2}\end{aligned}$$

مقادیر λ_1 و λ_2 به دست می آیند.

۳.۴ روش starship

روش starship یک روش برآورد کامپیوتری پر قدرت برای برآورد شکل است که در [۵] معرفی شده و برای توزیع هایی که از تبدیل یک توزیع پایه به دست آمده اند مورد استفاده قرار می گیرد. چون GLD یک تبدیل از توزیع یکنواخت است می توان از این روش برای برآورد پارامترهای آن استفاده کرد.

که در حل معادلات روش گشتاورها با آن مواجه بودیم، آسان تر می کند.

با توجه به رابطه (۵) تعداد زیادی از توزیع های GLD وجود دارند که گشتاورهای کمتری دارند و در هنگام استفاده از روش برآورد گشتاورها نادیده گرفته می شوند. در مواقعی که گشتاورها وجود ندارند، هنوز صدک ها برای برآورد پارامتر و به دست آوردن GLD قابل استفاده اند.

[۲] برای یک مجموعه داده X_1, X_2, \dots, X_n فرض کنید $\hat{\pi}_p$ نشان دهنده صدک $(100p)$ ام داده ها باشد. در ابتدا مقادیر r و $\frac{a}{b}$ را از عبارت $(n+1)p = r + \frac{a}{b}$ به دست می آوریم که r یک عدد صحیح مثبت و $\frac{a}{b}$ یک کسر در بازه $[0, 1)$ است. حال اگر Y_1, Y_2, \dots, Y_n آماره های ترتیبی داده ها باشند، از رابطه

$$\hat{\pi}_p = Y_r + \left(\frac{a}{b}\right)(Y_{r+1} - Y_r)$$

به دست می آید. آماره های نمونه ای $\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3$ و $\hat{\rho}_4$ به صورت

$$\hat{\rho}_1 = \hat{\pi}_{.05},$$

$$\hat{\rho}_2 = \hat{\pi}_{1-u} - \hat{\pi}_u,$$

$$\hat{\rho}_3 = \frac{\hat{\pi}_{.05} - \hat{\pi}_u}{\hat{\pi}_{1-u} - \hat{\pi}_{.05}},$$

$$\hat{\rho}_4 = \frac{\hat{\pi}_{.05} - \hat{\pi}_{.25}}{\hat{\rho}_2}$$

هستند که در آنها u عددی دلخواه از ۰ تا ۰/۲۵ می باشد (برای سهولت در محاسبات قرار می دهیم $u = 0.1$). این آماره ها به شرح زیر هستند:

- $\hat{\rho}_1$ میانه نمونه است.
- $\hat{\rho}_2$ دامنه میان دهکی است؛ یعنی اختلاف بین صدک ۱۰ام و صدک ۹۰ام.
- $\hat{\rho}_3$ نسبت وزن دم چپ و راست^۷ است (فاصله بین میانه و صدک ۱۰ام در صورت کسر و فاصله بین صدک ۹۰ام و میانه در مخرج کسر).

- $\hat{\rho}_4$ عامل وزنی دم^۸ یا نسبت دامنه میان چارکی به دامنه میان صدکی است که همواره کمتر از ۱ می باشد و وزن دم را اندازه می گیرد (مقادیر نزدیک به ۱ نشان می دهد که توزیع در دم هایش زیاد پخش نشده و مقادیر نزدیک به صفر نشان دهنده آن است که توزیع، دم های طولانی دارد).

^۷left-right tail-weight ratio

^۸tail-weight factor

الگوریتم روش ماکسیمم درست‌نمایی در ۴ گام خلاصه می‌شود:

گام ۱: محدوده‌ای از مقادیر ورودی برای λ_3 و λ_4 و تعداد مقادیر ورودی انتخاب‌شده را تعیین کنید. در نرم‌افزار R از دنباله سوبول برای تولید اعداد شبه تصادفی استفاده می‌شود. به طور ورودی تعداد مقادیر پیش‌فرض ۱۰۰۰۰ در نظر گرفته می‌شود.

گام ۲: λ_1 و λ_2 را برای هر کدام از λ_3 و λ_4 محاسبه کنید. مجموعه مقادیر ورودی که در پارامتری کردن GLD صدق نکنند یا همه فضای مجموعه داده‌ها را پوشش ندهند، غیر قابل قبول تلقی کرده، آنها را حذف کنید. در میان مجموعه نقاط شروع باقیمانده مجموعه‌ای را بیابید که کمترین مقدار $\sqrt{(\hat{\rho}_3 - \rho_3)^2 - (\hat{\rho}_4 - \rho_4)^2}$ و $\sqrt{(\hat{\alpha}_3 - \alpha_3)^2 - (\hat{\alpha}_4 - \alpha_4)^2}$ را تولید کند. حال این مجموعه مقادیر برای فرایند بهینه‌سازی مورد استفاده قرار می‌گیرد.

گام ۳: با جایگذاری مقادیر ورودی در رابطه‌های (۱) و (۳) صدک‌های u_i را برای RS و FMKL بیابید.

گام ۴: u_i به دست آمده در گام ۳ را در فرمول‌های لگاریتم عددی درست‌نمایی (روابط (۲۰) و (۲۱)) جایگذاری کنید.

$$MLRS = \sum_{i=1}^n \log \left[\frac{\lambda_2}{\lambda_3 u_i^{\lambda_3 - 1} + \lambda_4 u_i^{\lambda_4 - 1}} \right] \quad (20)$$

$$MLFMKL = \sum_{i=1}^n \log \left[\frac{\lambda_2}{u_i^{\lambda_3 - 1} + u_i^{\lambda_4 - 1}} \right] \quad (21)$$

گام ۵: مقدار بهینه با استفاده از الگوریتم سیمپلکس نلدر-مید یا هر الگوریتم مینیمم‌سازی عددی دیگر به دست می‌آید.

نتایج برازش به دست آمده در گام ۵، روش صدک‌های تجدید نظر شده RS تحت برآوردگر ماکسیمم درست‌نمایی RPRS.ML و روش گشتاورهای تجدید نظر شده FMKL تحت برآوردگر ماکسیمم درست‌نمایی RMFMKL.MLV نامیده می‌شوند.

مثال ۱.۴. برای مقایسه چهار روش برآورد، تعداد ۱۰۰۰ داده از توزیع نرمال استاندارد شبیه‌سازی کرده و به روش‌های گشتاورها، صدک‌ها، *starship* و ماکسیمم درست‌نمایی برآورد پارامترها را به دست آورده‌ایم و سپس آزمون کولموگوروف-اسمیرنوف

روش *starship* برای GLD توسط [۴] ارائه شده است، به طور مستقیم کیفیت برازش را از طریق ارزیابی مقادیر پارامترهای منتخب به وسیله آزمون نیکویی برازش بررسی می‌کند. این روش برای همه فضای پارامتر و هر دو پارامتری کردن GLD به کار می‌رود و در ساده‌ترین صورت در سه گام خلاصه می‌شود که عبارت‌اند از:

(i) ناحیه‌ای را در فضای چهاربعدي به گونه‌ای انتخاب کنید که محدوده چهار پارامتر $\lambda_1, \lambda_2, \lambda_3$ و λ_4 را به روشی مناسب پوشش دهد.

(ii) در ناحیه انتخاب شده در گام i یک شبکه مستطیلی چهاربعدي قرار دهید.

(iii) نقاط شبکه ایجاد شده در گام ii را از طریق اجرای آزمون نیکویی برازش روی توزیع متناظر، ارزیابی کنید. هنگامی که مقدار آماره آزمون در ناحیه رد قرار نگرفت، عملیات را متوقف کنید. در غیر این صورت با نقطه بعدی در شبکه ادامه دهید (همچنین می‌توان همه نقاط شبکه را بررسی و نقطه‌ای با بهترین اندازه نیکویی برازش را انتخاب کرد).

بنا بر این روش *starship* به جای مینیمم کردن یک تابع دو متغیره از λ_3 و λ_4 یک جستجوی شبکه‌ای چهار بعدی را انجام می‌دهد که در آن فقط یک معیار برای توقف وجود دارد که بر پایه نتیجه آزمون نیکویی برازش است. در نتیجه این روش محدوده وسیع تری از پارامترها را پوشش می‌دهد و به دلیل آنکه الگوریتم شامل نیکویی برازش است برازش خوبی را ارائه می‌کند. همچنین چون در این روش، جست‌وجو در فضای چهار بعدی انجام می‌شود، وقت گیر است که می‌توان این زمان را با انتخاب دقیق تر ناحیه شبکه‌ای، تا حدی کاهش داد. برای مطالعه دقیق تر، خواننده را به [۴] ارجاع می‌دهیم.

۴.۴ روش ماکسیمم درست‌نمایی

در این روش که در [۹] ارائه شده است، از اعداد شبه تصادفی و روش صدک‌ها برای RS و روش گشتاورها برای FMKL-GLD به منظور یافتن مقادیر ورودی استفاده می‌کنیم. سپس این مقادیر ورودی را برای ماکسیمم کردن لگاریتم عددی درست‌نمایی به کار می‌گیریم تا پارامترهای GLD مناسب را بیابیم.

را ۱۰۰۰ بار اجرا کرده ایم که نتایج در جدول های ۱ و ۲ آمده است. پارامترهای پیش فرض در ناحیه تعریف شده هر دو روش پارامتری کردن RS و $FMKL$ در نظر گرفته شده اند. با توجه به نتایج جدول های ۱ و ۲ می توان اظهار داشت که برای نواحی ای که شامل RS و $FMKL$ هستند، بر اساس معیارهای میانگین توان دوم خطاها و میانگین قدر مطلق خطاها، روش پارامتری کردن $FMKL$ بهتر از روش RS عمل می کند.

جدول ۱. برآورد پارامترها در روش $FMKL$

| گشتاورها | صدکها | starship | ماکسیمم درست نمایی | |
|----------|--------|----------|--------------------|-------------|
| ۰/۰۱۶۶ | ۰/۰۱۲۸ | ۰/۰۱۴۴ | ۰/۱۰۷۰ | λ_1 |
| ۱/۴۹۷۷ | ۱/۳۶۴۹ | ۱/۴۳۸۸ | ۰/۱۷۲۷ | λ_2 |
| ۰/۰۹۲۵ | ۰/۱۶۷۷ | ۰/۱۲۴۸ | ۰/۱۲۸۱ | λ_3 |
| ۰/۱۴۴۹ | ۰/۲۱۳۰ | ۰/۱۷۲۲ | ۰/۱۰۲۶ | λ_4 |
| ۰/۸۳۲ | ۰/۸۵۶ | ۰/۸۳۰ | ۰/۸۱۲ | P_{KS} |

جدول ۲. برآورد پارامترها در روش RS

| گشتاورها | صدکها | starship | ماکسیمم درست نمایی | |
|----------|--------|----------|--------------------|-------------|
| ۰/۰۱۶۶ | ۰/۰۱۲۸ | ۰/۰۱۴۴ | ۰/۱۰۷۰ | λ_1 |
| ۱/۴۹۷۷ | ۱/۳۶۴۹ | ۱/۴۳۸۸ | ۰/۱۷۲۷ | λ_2 |
| ۰/۰۹۲۵ | ۰/۱۶۷۷ | ۰/۱۲۴۸ | ۰/۱۲۸۱ | λ_3 |
| ۰/۱۴۴۹ | ۰/۲۱۳۰ | ۰/۱۷۲۲ | ۰/۱۰۲۶ | λ_4 |
| ۰/۷۵۸ | ۰/۸۷۳ | ۰/۸۲۹ | ۰/۹۱۴ | P_{KS} |

در این مثال با استفاده از آزمون کولموگوروف-اسمیرنوف،

برازش توزیع GLD در هر دو روش پارامتری کردن RS و $FMKL$ به داده های شبیه سازی شده از توزیع نرمال استاندارد را بررسی کرده ایم. سطر آخر جدول های ۲ و ۳ نشان دهنده این مطلب است که از ۱۰۰۰ بار اجرای آزمون کولموگوروف-اسمیرنوف چند مرتبه توزیع GLD در هر دو روش پارامتری کردن با پارامترهای برآورد شده در هر کدام از روش های گشتاورها، صدکها، starship و ماکسیمم درست نمایی، به خوبی به داده های شبیه سازی شده از توزیع نرمال استاندارد برازش یافته است. با توجه به نتایج به دست آمده در جدول های ۱ و ۲ نتیجه می گیریم که GLD با پارامترهایی که از روش برآورد ماکسیمم درست نمایی در روش پارامتری RS به دست آمده است از سایر روش ها بهتر می باشد؛ زیرا در ۱۰۰۰ بار تکرار آزمون نیکویی برازش، مشاهده کرده ایم که ۹۱۴ مرتبه برازش توزیع GLD به داده های شبیه سازی شده از توزیع نرمال استاندارد مناسب بوده است. در واقع مقدار P_{KS} جدول های ۱ و ۲ با استفاده از رابطه

$$P_{KS} = \frac{\text{تعداد دفعاتی که مقدار } D_n \text{ کمتر از مقدار بحرانی باشد}}{\text{تعداد تکرارها}}$$

$$D_n = \max_i |F_n(x_{(i)}) - F(x_{(i)})|$$

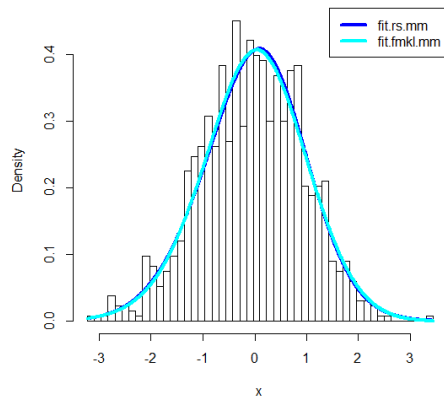
و $F_n(x_i) = \frac{i}{n}$ تابع توزیع تجمعی تجربی می باشد. هر دو روش پارامتری کردن RS و $FMKL$ توزیع GLD را با استفاده از چهار روش برآورد پارامتر به بافت نگاشت داده های شبیه سازی شده از توزیع نرمال استاندارد برازش داده ایم، که شکل ۱ این نمودارها را نشان می دهد.

شکل های ۱ تا ۵ بافت نگاشت و احتمال تجمعی داده ها را نشان می دهند. تقریباً در همه روش های برآورد، توزیع GLD به خوبی به متغیرهای تصادفی تولید شده از توزیع نرمال استاندارد برازش یافته است.

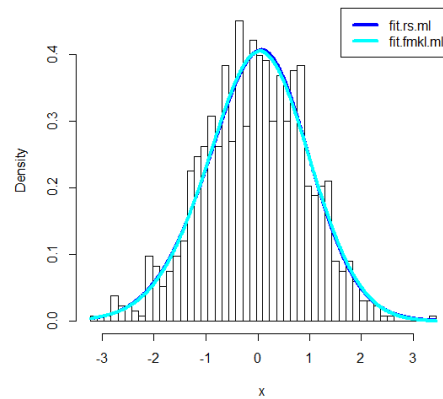
مثال ۲.۴. برای سنجش میزان دقت برآورد پارامترهای GLD ، تعداد ۱۰۰۰ داده از توزیع $GLD(0/1, 0/1, 0/1, 0/1)$ شبیه سازی کرده، مقدار اریبی، میانگین توان دوم خطاها و میانگین قدر مطلق خطاها را برای RS در روش برآورد ماکسیمم درست نمایی و برای $FMKL$ در روش صدکها محاسبه کرده ایم (با توجه به نتایج جدول ۲). دلیل انتخاب این پارامترها این است که اطمینان

به نتایج جدول ۳ می‌توان اظهار داشت که در کل برای نواحی‌ای که شامل RS و $FMKL$ هستند، بر اساس معیارهای میانگین توان دوم خطاها و میانگین قدر مطلق خطاها، روش پارامتری کردن $FMKL$ بهتر از روش RS عمل می‌کند.

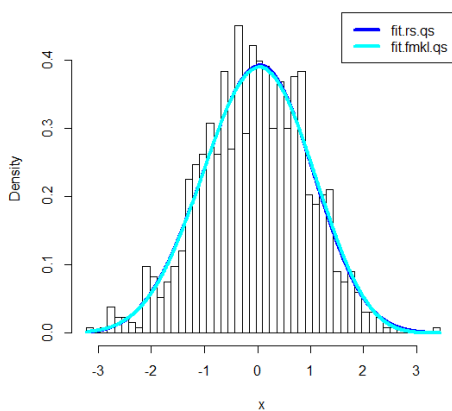
داشته باشیم توزیع GLD حاصل در رده چگالی‌های تک‌مدی قرار می‌گیرد؛ چون هدف، مقایسه پارامترهای برآورد شده متغیرهای تصادفی شبیه‌سازی شده از توزیع نرمال استاندارد است. پارامترهای پیش فرض در ناحیه تعریف شده هر دو روش پارامتری کردن RS و $FMKL$ در نظر گرفته شده‌اند. با توجه



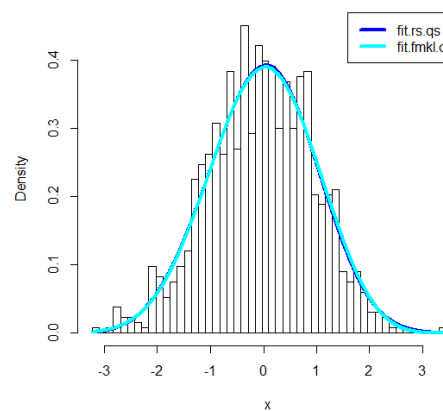
(ب) روش برآورد گشتاورها



(آ) روش ماکسیم درست‌نمایی



(د) روش starship



(ج) روش برآورد صدکها

شکل ۱. بافت‌نگاشت داده‌ها در چهار روش برآورد برای هر دو روش پارامتری RS و $FMKL$

که در این صورت توزیع حاصل به صورت

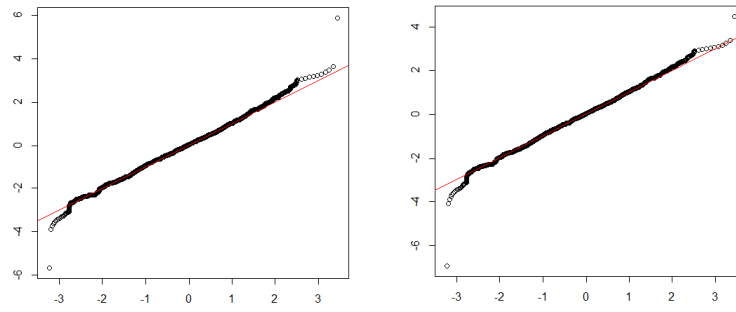
$$GLD_{RS}(48588/14, -0/000005, -0/042516, -0/29694)$$

$$GLD_{FMKL}(70021/87, 0/000009, 1/800929, 0/18697)$$

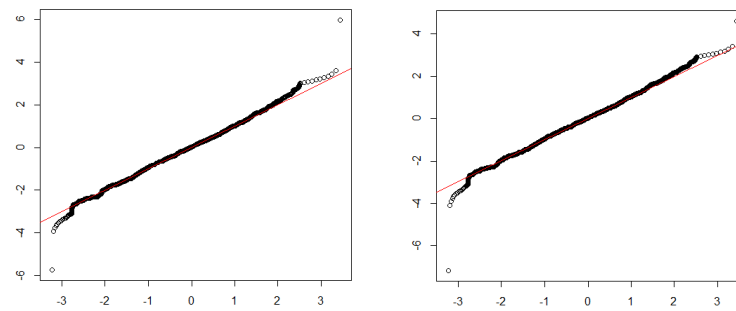
خواهد بود.

۵ مثال کاربردی

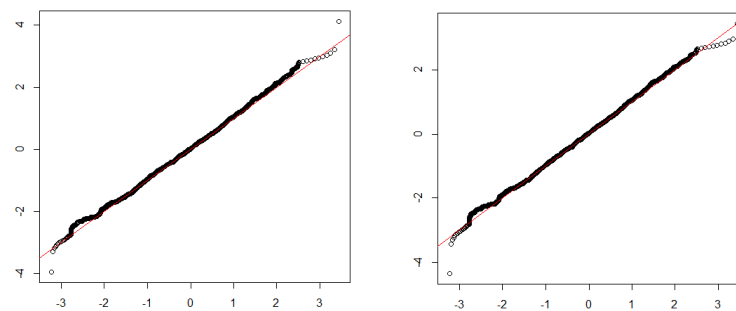
توزیع لامبدای تعمیم یافته را به داده‌های بورس اوراق بهادار شرکت‌های تهران که از ابتدای سال ۱۳۸۷ تا پایان سال ۱۳۹۱ فعالیت مستمر داشته‌اند برآزش می‌دهیم. با توجه به نتایج به دست آمده در جدول ۲، از روش برآورد ماکسیم درست‌نمایی برای RS و از روش برآورد صدکی برای GLD استفاده می‌کنیم،



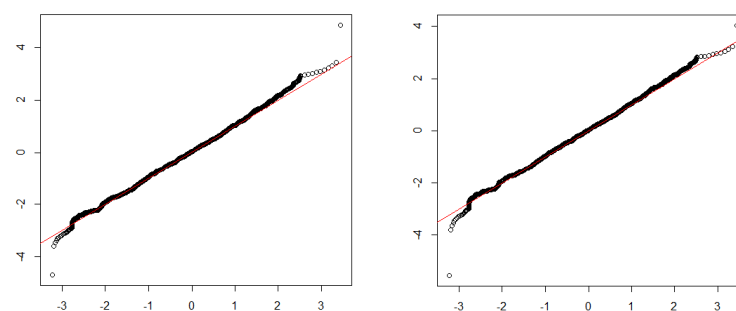
شکل ۲. نمودار احتمال تجمعی در روش برآورد ماکسیمم درست‌نمایی برای هر دو روش پارامتری RS و FMKL



شکل ۳. نمودار احتمال تجمعی در روش برآورد گشتاورها برای هر دو روش پارامتری RS و FMKL



شکل ۴. نمودار احتمال تجمعی در روش برآورد صدک‌ها برای هر دو روش پارامتری RS و FMKL



شکل ۵. نمودار احتمال تجمعی در روش برآورد starship برای هر دو روش پارامتری RS و FMKL

۶ نتیجه گیری

به مقایسه آنها پرداختیم. نتایج مقایسه نشان داد که برای برآورد پارامترهای RS روش ماکسیم درست‌نمایی و برای FMKL روش برآورد صدکی بهتر است و در کل برای نواحی‌ای که شامل RS و FMKL هستند، روش پارامتری کردن FMKL بهتر از روش RS عمل می‌کند. همچنین نتایج به‌دست‌آمده را برای داده‌های بورس اوراق بهادار تهران محاسبه کردیم.

در این مقاله توزیع لامبدای تعمیم‌یافته و خواص آن را ارائه کردیم. سپس دو شکل پارامتری RS و FMKL مربوط به توزیع GLD و چهار روش متفاوت برای برآورد پارامترهای این توزیع را ارائه کرده، توسط آزمون کولموگوروف-اسمیرنوف و معیارهای میانگین توان دوم خطاها و میانگین قدر مطلق خطاها

جدول ۳. مقایسه نتایج شبیه‌سازی روش‌های RS و FMKL بر اساس تکرارهای مختلف

| MAE | | MSE | | اریبی | | n |
|--------|--------|--------|--------|---------|---------|------|
| RS | FMKL | RS | FMKL | RS | FMKL | |
| ۱/۶۸۲۱ | ۱/۲۹۶۱ | ۴/۷۸۹۴ | ۲/۷۱۶۸ | -۰/۰۷۳۹ | ۰/۰۸۹۱ | ۱۰۰ |
| ۱/۸۵۸۵ | ۱/۳۵۳۲ | ۵/۵۱۵۸ | ۲/۹۱۶۹ | -۰/۰۴۴۷ | -۰/۱۶۱۹ | ۵۰۰ |
| ۱/۶۷۷۴ | ۱/۲۶۱۹ | ۴/۵۳۲۴ | ۲/۵۸۳۴ | -۰/۲۶۰۵ | -۰/۰۲۲۱ | ۱۰۰۰ |

مراجع

- [1] Freimer, M., Mudholkar, S., Kollia, G. and Lin, C. (1988), A study of the generalized Tukey lambda family, *Communication in Statistics, Theory and Methods*, **17(10)**, 3547-3567.
- [2] Karian, Z. A., and Dudewicz, E. J. (1999), Fitting the generalized lambda distribution (GLD) system by a method of percentiles, II: Tables, *American Journal of Mathematical and Management Sciences*, **19**, 1-73.
- [3] Karian, Z. A. and Dudewicz, E. J. (2000), *Fitting statistical distributions to data: generalized lambda distribution and generalized bootstrap methods*, CRC Press, New York.
- [4] King, R. and MacGillivray, H. (1999), A Starship estimation method for the generalized lambda distributions, *Australian and New Zealand Journal of Statistics*, **41(3)**, 353-374.
- [5] Owen, D. B. (1988), The starship, *Communications in Statistics, Simulation and Computation*, **17(2)**, 315-323.
- [6] Ramberg J. S., and Schmeiser B. W. (1974), An approximate method for generating symmetric random variables. *Communication ACM*, **17(2)**, 78-82.
- [7] Ramberg, J. S., Dudewicz, E.J., Tadikamalla P. and Mykytka, E. (1979). A probability distribution and its uses in fitting data, *Technometrics*, **21(2)**, 201-214.
- [8] Tukey, J. W. (1960), The practical relation between the common transformation of percentages of counts and of amounts. *Technical Report No 36, Statistical Techniques Research Group, Princeton University*.
- [9] Su, S. (2007), Numerical maximum log likelihood estimation for generalized lambda distributions. *Computational statistics & data analysis*, **51(8)**, 3983-3998.