مجله علوم آماری، پاییز و زمستان ۱۳۹۱
جلد ۶، شماره ۲، ص ۱۱۹-۱۳۴

تحلیلی از برآوردگرهای اندازه وابستگی دمی بالا

محمد امینی، هادی جباری نوقابی، مهلا قاسم‌نژاد فرسنگی
گروه آمار، دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۹۱/۴/۲۴ تاریخ آخرین بازنگری: ۱۳۹۱/۱۱/۹

چکیده : در این مقاله سه نوع برآوردگر جدید به روش ناپارامتری برای اندازه وابستگی دمی بالا به‌دست آورده و نشان داده می‌شود که برآوردگرهایی سازگار و به‌طور مجانبی نااریب هستند. سپس با شبیه‌سازی مونت‌کارلو از سه مفصل متفاوت، این سه برآوردگر با هم مقایسه شده و با به‌کارگیری داده‌های واقعی روشی جدید برای انتخاب بهترین برآوردگر ارائه می‌شود.

واژه‌های کلیدی : تابع مفصل، اندازه وابستگی دمی بالا، اندازه وابستگی دمی پایین، برآوردگر، نا اریبی، سازگاری، شبیه‌سازی مونت کارلو.

۱ مقدمه

امروزه در شاخه‌های مختلف علمی، اندازه‌گیری وابستگی بین متغیرهای مختلف، حائز اهمیت است و ضرایب وابستگی مانند ρ پیرسون، ρ_s اسپیرمن، τ کندال، γ جینی و غیره هر کدام بنا به موقعیت و ویژگی‌های منحصربه‌فردی که دارند، مورد استفاده قرار می‌گیرند. امبرتس و همکاران (۲۰۰۲) نشان دادند که همیشه نمی‌توان از اندازه‌های وابستگی معمولی برای اندازه‌گیری وابستگی بین متغیرها استفاده کرد.

آدرس الکترونیک مسئول مقاله: مهلا قاسم‌نژاد فرسنگی، ma_gh272@yahoo.com
کد موضوع‌بندی ریاضی (۲۰۱۰): ۶۲H۲۰

در واقع زمانی که توزیع جامعه نامشخص باشد یا زمانی که وابستگی بین متغیرها در یک برهه زمانی خاص (مانند زمان رخداد بحران‌ها) مورد نظر باشد، به نوع دیگری از اندازه‌های وابستگی نیاز است. اندازه‌ای که هم بتوان آن را بر حسب تابع مفصل بیان کرد و هم برای اندازه‌گیری میزان وابستگی در پیشامدهای فرین مناسب باشد. سیببویا (۱۹۶۰) نوعی از اندازه‌های وابستگی را معرفی کرد که دارای ویژگی‌های فوق بوده و اندازه وابستگی دمی (قوی)[۱] نام گرفته است. این اندازه به دو نوع اندازهٔ وابستگی دمی بالا و پایین تفکیک می‌شود که هر یک در فاصله [۰, ۱] قرار دارند. نزدیک شدن آن‌ها به یک بیانگر وابستگی قوی بوده و محدوده صفر، محدوده استقلال متغیرهاست. دوبریک و اسمیت (۲۰۰۵) سه نوع برآوردگر ناپارامتری برای اندازه وابستگی دمی پایین معرفی کرده و برخی از ویژگی‌های این برآوردگرها از قبیل سازگاری و نااریبی مجانبی را مورد بررسی قرار دادند. در این تحقیق سه نوع برآوردگر برای اندازه وابستگی دمی بالا به‌دست آورده و نشان داده می‌شود که این برآوردگرها، سازگار و به‌طور مجانبی نااریب هستند. همچنین با شبیه‌سازی مونت‌کارلو رفتار مجانبی این سه برآوردگر در سه زمینه تحت بررسی (مفصل گامبل، ترکیب محدب مفصل‌های حاصل‌ضرب ($\prod(u,v)$) و کران بالای فرشه ($M(u,v)$) و مفصل نرمال دومتغیره) مورد مقایسه قرار می‌گیرد. در بخش ۲ تعاریف مورد نیاز ارائه می‌شود. بخش ۳ شامل معرفی برآوردگرهای جدید مربوط به اندازه وابستگی دمی بالا و طریقه به‌دست آوردن آن‌هاست. خواص برآوردگرها از جمله خواص سازگاری و نااریبی مجانبی آن‌ها در بخش ۴ ارائه می‌شود. در این بخش به دو روش نموداری و محاسباتی، خواص برآوردگرهای مربوط به اندازه وابستگی دمی بالا مورد بررسی قرار می‌گیرد. در بخش ۵ روشی جدید برای انتخاب بهترین برآوردگر از بین سه برآوردگر مربوط به اندازه وابستگی دمی بالا، ارائه می‌شود. تمام شبیه‌سازی‌ها و محاسبات در محیط نرم افزار R انجام شده است.

---
[۱] Strong tail dependence measure

۲ اندازه های وابستگی دمی

اندازه‌های وابستگی دمی برای اندازه‌گیری میزان وابستگی بین متغیرها در پیشامدهای نادر مانند بحران‌های اقتصادی به‌کار می‌روند. همچنین کلایول و گاگن (۲۰۰۵)، با استفاده از این اندازه‌ها تأثیر متقابل بازارهای تجاری کشورهای تایلند، مالزی و اندونزی در بحران اقتصادی سال ۱۹۹۰ را مورد بررسی قرار دادند. اندازه‌های وابستگی دمی که توسط سیبویا (۱۹۶۰) معرفی شدند، میزان وابستگی را در گوشه بالای سمت راست مربع I^{2} (اندازه وابستگی دمی بالا) و گوشه پایین سمت چپ آن (اندازه وابستگی دمی پایین) اندازه می‌گیرند.

تعریف ۱ : فرض کنید X و Y متغیرهایی تصادفی با توابع توزیع F و G باشند. در صورتی که حدود ذکر شده موجود باشند، اندازه‌های وابستگی دمی بالا و پایین به صورت

$$\lambda_u = \lim_{t\to 1^-} P[Y > G^{-1}(t) \mid X > F^{-1}(t)], \qquad (۱)$$

$$\lambda_l = \lim_{t\to 0^+} P[Y \leq G^{-1}(t) \mid X \leq F^{-1}(t)] \qquad (۲)$$

نمایش داده می‌شوند. با نزدیک شدن t به ۱، اندازه دمی بالا برابر است با احتمال Y بزرگتر از صدک t ام توزیع G، به شرط این که X بزرگ‌تر از صدک tام توزیع F باشد. λ_l نیز به طریقی مشابه بیان می‌شود. اسکلار (۱۹۵۹) در قضیه معروف خود نشان داد اگر X و Y دو متغیر تصادفی با تابع توزیع توأم H و توابع توزیع حاشیه‌ای F و G باشند، آنگاه مفصل C وجود دارد به قسمی که $H(x,y) = C(F(x), G(y))$، که در آن اگر توابع توزیع F و G مطلقاً پیوسته باشند، مفصل C یکتاست. اگر C تابع مفصل بین X و Y باشد، در صورت وجود حدود موجود در (۱) و (۲) داریم

$$\lambda_u = 2 - \lim_{t\to 1^-} \frac{1 - C(t,t)}{1 - t}\ , \quad \lambda_l = \lim_{t\to 0^+} \frac{C(t,t)}{t}. \qquad (۳)$$

بنابراین طبق قضیه اسکلار (۱۹۵۹)، اندازه‌های وابستگی دمی بدون در دست داشتن تابع توزیع توأم موجود بین دو متغیر نیز قابل محاسبه هستند.

۳ معرفی برآوردگرهای λ_u

با توجه به این‌که مبنای محاسبه برآوردگرهای معرفی شده در این بخش تابع مفصل تجربی است، ابتدا این تابع معرفی می‌شود.

تعریف ۲ : اگر $(X_1, Y_1), \ldots, (X_n, Y_n)$ نمونه‌ای تصادفی به حجم n از جامعه‌ای با تابع توزیع توأم پیوسته $H(x, y)$ باشد، آن‌گاه تابع

$$C_n(\frac{i}{n}, \frac{j}{n}) = \frac{\text{تعداد زوج های } (x,y) \text{ نمونه که } x \leq x(i)\ ,\ y \leq y(j)}{n} \qquad (۴)$$

را مفصل تجربی نامند، که در آن $x(i)$ و $y(j)$ $(1 \leq i, j \leq n)$ آماره‌های مرتب نمونه هستند. مفصل‌های تجربی اولین بار توسط دی هیولز (۱۹۷۹) با عنوان توابع وابستگی تجربی معرفی شدند. در ادامه این برآوردگرها برای مقادیر k نزدیک به n تعریف می‌شوند.

برآوردگر اول: اولین برآوردگر با استفاده از رابطه (۱) و با به‌کار گیری قضیه اسکلار به‌صورت

$$\begin{aligned}
\hat{\lambda}_u &= \lim_{i \to n^-} P[Y > G^{-1}(\frac{i}{n}) \mid X > F^{-1}(\frac{i}{n})] \\
&= \lim_{i \to n^-} \frac{P(X > F^{-1}(\frac{i}{n}), Y > G^{-1}(\frac{i}{n}))}{P(X > F^{-1}(\frac{i}{n}))} \\
&= \lim_{i \to n^-} \frac{\bar{H}(F^{-1}(\frac{i}{n}), G^{-1}(\frac{i}{n}))}{1 - \frac{i}{n}} \\
&= 2 - \lim_{i \to n^-} \frac{1 - C_n(\frac{i}{n}, \frac{i}{n})}{1 - \frac{i}{n}},
\end{aligned}$$

به‌دست می‌آید. بنابراین

$$\hat{\lambda}^{(1)}_{u_{n,i}} = 2 - \lim_{i \to n^-} \frac{1 - C_n(\frac{i}{n}, \frac{i}{n})}{1 - \frac{i}{n}}. \qquad (۵)$$

حد موجود در (۵) برای داده‌های جامعه واقعی قابل محاسبه نیست. در نتیجه $\hat{\lambda}_u$ به صورت

$$\hat{\lambda}^{(1)}_{u_{n,k}} = 2 - \frac{1 - C_n(\frac{k}{n}, \frac{k}{n})}{1 - \frac{k}{n}} \qquad (۶)$$

معرفی می‌شود، که در آن $k \in N$ تابعی از n است و در داده‌های واقعی به روش کلایول و گاگن (۲۰۰۵) انتخاب می‌شود.

برآوردگر دوم: دومین برآوردگر پس از تعیین مدل رگرسیونی

$$C_n(\frac{i}{n},\frac{i}{n}) = ۱ - (۱ - \frac{i}{n})(۲ - \lambda_u) + \varepsilon_i \ , \ \ i = k, \dots, n$$

با جمله خطای ε_i و از روش کمترین توان‌های دوم به‌صورت

$$\hat{\lambda}^{(۲)}_{u_{n,k}} \ = \ ۲ - \frac{\sum_{i=k}^{n}(۱ - \frac{i}{n})[۱ - C_n(\frac{i}{n},\frac{i}{n})]}{\sum_{i=k}^{n}(۱ - \frac{i}{n})^۲}, \qquad (۷)$$

به‌دست می‌آید. این رابطه رگرسیونی با توجه به تعریف λ_u به‌صورت

$$\lambda_u = ۲ - \lim_{u \to ۱} \frac{۱ - C(u,u)}{۱ - u}$$
$$(۲ - \lambda_u)(۱ - u) = ۱ - C(u,u) + o(۱ - u)$$
$$C(u,u) = ۱ - (۲ - \lambda_u)(۱ - u) + \varepsilon_i$$

نتیجه می‌شود و برآوردگر مورد نظر با مشتق‌گیری از مجموع توان‌های دوم

$$L = \sum_{i=k}^{n} {\varepsilon_i}^۲ = \sum_{i=k}^{n} [C_n(\frac{i}{n},\frac{i}{n}) - [۱ + (۲ - \lambda_u)(۱ - \frac{i}{n})]]^۲,$$

و حل معادله

$$-۲\sum_{i=k}^{n}(۱ - \frac{i}{n})[C_n(\frac{i}{n},\frac{i}{n}) - [۱ + (۲ - \lambda_u)(۱ - \frac{i}{n})]] = ۰$$

به‌صورت

$$\hat{\lambda}^{(۲)}_{u_{n,k}} \ = \ ۲ - \frac{\sum_{i=k}^{n}(۱ - \frac{i}{n})[۱ - C_n(\frac{i}{n},\frac{i}{n})]}{\sum_{i=k}^{n}(۱ - \frac{i}{n})^۲},$$

به‌دست می‌آید. با توجه به تعریف λ_u در (۳) برای به‌دست آوردن برآوردگرها مقادیر i نزدیک به n $(i = k, \dots, n)$ در نظر گرفته می‌شود.

برآوردگر سوم: برای محاسبه سومین برآوردگر، مفصل C با ترکیب خطی محدب از مفصل‌های M و W (کران‌های بالا و پایین فرشه)، به‌صورت

$$C(u,v)=\lambda_u M(u,v)+(1-\lambda_u)W(u,v) \tag{8}$$

بازنویسی می‌شود. در واقع در بازنویسی مفصل C با یک ترکیب خطی از مفصل‌های C_1 و C_2 و با در نظر گرفتن تعریف λ_u، تنها در صورتی یک عبارت همواره صحیح حاصل خواهد شد که به جای مفصل‌های C_1 و C_2 به‌ترتیب از مفصل‌های M و W استفاده شود. پس از آن باید در نقطه $(\frac{i}{n},\frac{i}{n})$، فاصله مفصل C در رابطه (۸) و مفصل تجربی موجود در رابطه (۴) مینیمم شود. یعنی $\hat{\lambda}^{(3)}_{u_{n,k}}$ ، با استفاده از رابطه

$$L=\min\sum_{i=k}^{n}\{C_n(\frac{i}{n},\frac{i}{n})-[\lambda_u\frac{i}{n}+(1-\lambda_u)\max(2\frac{i}{n}-1,0)]\}^2$$

به‌دست می‌آید. با توجه به تعریف λ_u، برآوردگرهای $\hat{\lambda}^{(i)}_{u_{n,k}}$ به‌ازای مقادیر kی نزدیک به n و در فاصله $\frac{n}{2}<k<n-\sqrt{n}$، $(n>4)$، تعریف شدند، بنابراین $\max[2\frac{i}{n}-1,0]=2\frac{i}{n}-1$ و در نتیجه

$$\frac{\partial L}{\partial\lambda_u}=\frac{\partial\sum_{i=k}^{n}\{C_n(\frac{i}{n},\frac{i}{n})-[\lambda_u(1-\frac{i}{n})+(2\frac{i}{n}-1)]\}^2}{\partial\lambda_u}=0$$

که از آن برآوردگر

$$\hat{\lambda}^{(3)}_{u_{n,k}}=\frac{\sum_{i=k}^{n}(1-\frac{i}{n})[C_n(\frac{i}{n},\frac{i}{n})-(2\frac{i}{n}-1)]}{\sum_{i=k}^{n}(1-\frac{i}{n})^2}, \tag{9}$$

حاصل می‌شود.

۴ خواص برآوردگرها

در این بخش، خواص مجانبی برآوردگرهای λ_u با استفاده از قضایا و تحلیل‌های نموداری برگرفته از شبیه‌سازی مونت‌کارلو بیان می‌شوند.

قضیه ۱ : اگر $C(u,v)$، یک مفصل مطلقاً پیوسته و $\{(X_n,Y_n),\ n\geq 1\}$، یک دنباله از بردارهای تصادفی مستقل، هم توزیع و دارای مفصل متناظر C باشد، آنگاه در صورت برقراری $\lim_{n\to\infty}\frac{k}{n}=1$، با بزرگ شدن n داریم

$$\lim_{n\to\infty} P(|\hat{\lambda}^{(i)}_{u_{n,k}}-\lambda_u|\geq\epsilon)=0,\quad \lim_{n\to\infty} E(\hat{\lambda}^{(i)}_{u_{n,k}})=\lambda_u\quad i=1,2,3$$

به عبارت دیگر هر سه برآوردگر مربوط به λ_u، برآوردگرهایی سازگار و به‌طور مجانبی ناریب هستند.

برهان در ابتدا سازگاری برآوردگر نوع (۱) بررسی می‌شود.

$$\begin{aligned}\hat{\lambda}^{(1)}_{u_{n,k}} &= \frac{1-\frac{2k}{n}+C_n(\frac{k}{n},\frac{k}{n})}{1-\frac{k}{n}}\\ &= \frac{1-\frac{2k}{n}+C(\frac{k}{n},\frac{k}{n})}{1-\frac{k}{n}}+\frac{C_n(\frac{k}{n},\frac{k}{n})-C(\frac{k}{n},\frac{k}{n})}{1-\frac{k}{n}}\\ &= \frac{1-\frac{2k}{n}+C(\frac{k}{n},\frac{k}{n})}{1-\frac{k}{n}}+\frac{\sqrt{n}}{n-k}\sqrt{n}[C_n(\frac{k}{n},\frac{k}{n})-C(\frac{k}{n},\frac{k}{n})]\\ &= [1]+[2].\end{aligned}$$

که در آن k، هم‌زمان با افزایش n به n نزدیک می‌شود. سرعت نزدیک شدن k به n کمتر از سرعت نزدیک شدن n به بی‌نهایت است و همواره رابطه $k<n-\sqrt{n}$، برقرار است. چون $\lim_{n\to\infty}k=n$، داریم

$$[1]=\lim_{u\to 1}\frac{1-2u+C(u,u)}{1-u}=\lambda_u.$$

با افزایش n، جمله [۲] به صفر همگراست. زیرا بنابر $\lim_{n\to\infty}k=n$ و قضیه حدی تابعی[۲] (فرمانین و همکاران، ۲۰۰۴) برای توابع مفصل داریم

$$\sqrt{n}[C_n(\frac{k}{n},\frac{k}{n})-C(\frac{k}{n},\frac{k}{n})]\longrightarrow G_c(1,1).$$

که در آن $G_c(1,1)$ یک متغیر تباهیده در نقطه یک است. از طرف دیگر چون $k<n-\sqrt{n}$، بنابراین

$$\lim_{n\to\infty}\frac{\sqrt{n}}{n-k}=0.$$

[۲] Functional limit theorem

برای برآوردگر نوع (۲) داریم

$$\begin{aligned}\hat{\lambda}^{(2)}_{u_{n,k}} &= \frac{2\sum_{i=k}^{n}(1-\frac{i}{n})^{2} - \sum_{i=k}^{n}(1-\frac{i}{n}) + \sum_{i=k}^{n}(1-\frac{i}{n})C_n(\frac{i}{n},\frac{i}{n})}{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}} \\ &= \frac{\sum_{i=k}^{n}(1-\frac{i}{n})[2(1-\frac{i}{n})-1+C_n(\frac{i}{n},\frac{i}{n})]}{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}}.\end{aligned}$$

چون $k \le i \le n$ و $\lim_{n\to\infty} k = n$، بنابراین با افزایش n، مقدار i به n نزدیک می‌شود. به‌علاوه چون

$$\lim_{i\to n}\frac{1-\frac{2i}{n}+C_n(\frac{i}{n},\frac{i}{n})-\lambda_u(1-\frac{i}{n})}{1-\frac{i}{n}} = \lambda_u - \lambda_u = 0,$$

داریم

$$1-\frac{2i}{n}+C_n(\frac{i}{n},\frac{i}{n})-\lambda_u(1-\frac{i}{n}) = o(1-\frac{i}{n}).$$

بنابراین باجایگذاری

$$C_n(\frac{i}{n},\frac{i}{n}) = \lambda_u(1-\frac{i}{n}) - 1 + \frac{2i}{n} + o(1-\frac{i}{n}) \qquad (10)$$

در $\hat{\lambda}^{(2)}_{u_{n,k}}$ داریم

$$\begin{aligned}\hat{\lambda}^{(2)}_{u_{n,k}} &= \frac{\sum_{i=k}^{n}(1-\frac{i}{n})[2(1-\frac{i}{n})-1+[\lambda_u(1-\frac{i}{n})-1+\frac{2i}{n}+o(1-\frac{i}{n})]]}{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}} \\ &= \lambda_u + \frac{\sum_{i=k}^{n}(1-\frac{i}{n})o(1-\frac{i}{n})}{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}} \\ &= \lambda_u + \frac{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}(\frac{o(1-\frac{i}{n})}{1-\frac{i}{n}})}{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}},\end{aligned}$$

که جمله دوم با افزایش n، به صفر همگراست. برای برآوردگر نوع (۳) نیز به طریقی مشابه با جایگذاری رابطه (۱۰) در $\hat{\lambda}^{(3)}_{u_{n,k}}$، داریم

$$\begin{aligned}\hat{\lambda}^{(3)}_{u_{n,k}} &= \frac{\sum_{i=k}^{n}(1-\frac{i}{n})[C_n(\frac{i}{n},\frac{i}{n})+(1-\frac{2i}{n})]}{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}} \\ &= \frac{\sum_{i=k}^{n}(1-\frac{i}{n})[[\lambda_u(1-\frac{i}{n})-1+\frac{2i}{n}+o(1-\frac{i}{n})]+(1-\frac{2i}{n})]}{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}} \\ &= \lambda_u + \frac{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}(\frac{o(1-\frac{i}{n})}{1-\frac{i}{n}})}{\sum_{i=k}^{n}(1-\frac{i}{n})^{2}}.\end{aligned}$$

که در اینجا نیز جمله دوم با افزایش n به صفر همگراست.

با توجه به این که تمامی برآوردگرهای معرفی شده سازگار هستند، $\hat{\lambda}^{(i)}_{u_{n,k}}$ در احتمال به λ_u همگراست. از طرفی چون اندازه‌های وابستگی دمی در فاصله $[۰,۱]$ واقع می‌شوند، برآوردگرهای آن‌ها نیز در این فاصله قرار دارند. بنابراین $|\hat{\lambda}^{(i)}_{u_{n,k}} - \lambda_u| \leq ۱$ و در نتیجه $(\hat{\lambda}^{(i)}_{u_{n,k}} - \lambda_u)$، انتگرال پذیر یکنواخت است و طبق روابط همگرایی داریم

$$\lim_{n\to\infty} P(|\hat{\lambda}^{(i)}_{u_{n,k}} - \lambda_u| \geq \epsilon) = ۰, \ ۰ \leq \lambda_u \leq ۱ \ \ a.e\ , \ \ i = ۱, ۲, ۳$$

بنابراین $E(\hat{\lambda}^{(i)}_{u_{n,k}}) \to \lambda_u$، یعنی این برآوردگرها به‌طور مجانبی ناااریب هستند.

۱.۴ بررسی خواص برآوردگرهای $\hat{\lambda}^{(i)}_{u_{n,k}}$

برای بررسی خواص برآوردگرهای $\hat{\lambda}^{(i)}_{u_{n,k}}$ $\{i = ۱, ۲, ۳\}$، از طریق شبیه‌سازی مونت‌کارلو در نمونه‌های متناهی، سه مجموعه متفاوت خانواده مفصل گامبل (گامبل، ۱۹۶۰) ، خانواده ترکیب محدب M و $\prod$ و خانواده مفصل نرمال با توابع مفصل زیر در نظر گرفته می‌شود.

خانواده مفصل گامبل:

$$C_{۱\theta}(u,v) = \exp\{-[(-\log(u))^{\theta} + (-\log(v))^{\theta}]^{۱/\theta}\}\ , \ \ \theta \geq ۱$$

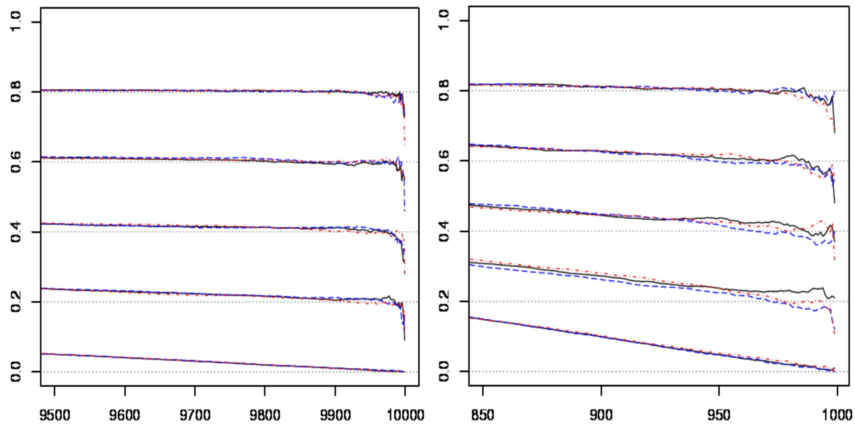
خانواده ترکیب محدب $\prod$ و M:

$$C_{۲\theta}(u,v) = \theta M(u,v) + (۱-\theta)\prod(u,v)\ , \ \ ۰ \leq \theta \leq ۱$$

خانواده مفصل نرمال:

$$C_{\rho}(u,v) = \int_{-\infty}^{\Phi^{-۱}(u)} \int_{-\infty}^{\Phi^{-۱}(v)} \frac{۱}{۲\pi\sqrt{۱-\rho^{۲}}} \exp\left(-\frac{s^{۲} - ۲\rho st + t^{۲}}{۲(۱-\rho^{۲})}\right) ds\ dt.$$

در شکل ۱ برای تحلیل اریبی برآوردگرهای $\hat{\lambda}^{(i)}_{u_{n,k}}$، با شبیه‌سازی مونت کارلو از مفصل گامبل مقادیر $E(\hat{\lambda}^{(i)}_{u_{n,k}})$ در مقابل k رسم شده‌اند. تعداد دفعات شبیه‌سازی

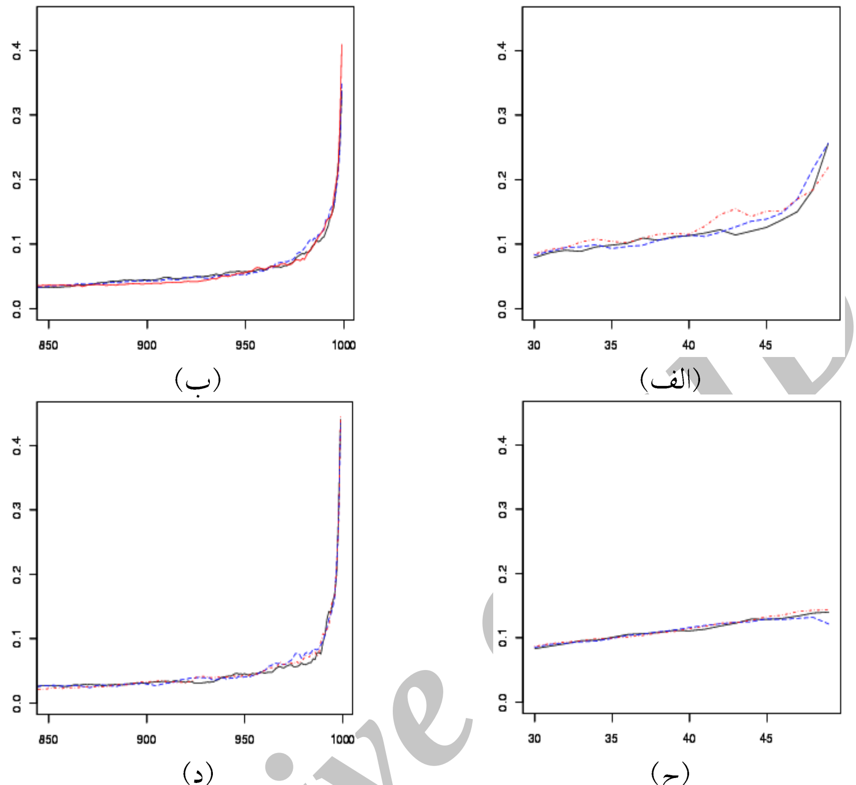


(ب) (الف)

شکل ۱: میانگین‌های ۱۰۰ بار شبیه‌سازی مونت کارلو از $\lambda_{u_{n,k}}^{(i)}$, $(i = ۱,۲,۳)$ برای مجموعه اول با $\lambda_u = ۰, ۰/۲, ۰/۴, ۰/۶, ۰/۸$ الف: $n = ۱۰۰۰$ و ب: $n = ۱۰۰۰۰$

$B = ۱۰۰$ در نظر گرفته شده و از نماد خطوط پیوسته برای $E(\hat{\lambda}_{u_{n,k}}^{(۱)})$، خط‌چین برای $E(\hat{\lambda}_{u_{n,k}}^{(۲)})$ و نقطه‌چین برای $E(\hat{\lambda}_{u_{n,k}}^{(۳)})$ استفاده شده است. محدوده رسم نمودار، مقادیر kای نزدیک به n می‌باشد و با توجه به نمودارها نیز این مطلب قابل تأیید است که k تابعی از n بوده و با افزایش n همواره به آن نزدیک می‌شود.

هر سه برآوردگر اریب هستند و میزان اریبی آن‌ها به λ_u، k و n بستگی دارد. با افزایش λ_u از اریبی برآوردگرها کاسته می‌شود. در مقادیر kای کوچک، اریبی مثبت و به‌ازای مقادیر k نزدیک به n اریبی منفی است. به‌طور کلی رفتار سه برآوردگر $\hat{\lambda}_{u_{n,k}}^{(i)}$ بسیار نزدیک به‌هم بوده و با بزرگ شدن مقدار n، نمودارهای مربوط به آن‌ها تقریباً منطبق بر هم خواهد شد. با افزایش حجم نمونه، اریبی برآوردگرها تا حد قابل ملاحظه‌ای کاهش می‌یابد و می‌توان گفت اگر n به اندازه‌ی کافی بزرگ باشد، برآوردگرها نااریب خواهند بود. انحراف معیار برآوردگرهای مجموعه اول در شکل ۲ نمایش داده شده‌اند. نمودارهای ۲.الف به‌ازای $(n, \lambda_u) = (۵۰, ۰/۲)$، ۲.ب به‌ازای $(n, \lambda_u) = (۱۰۰۰, ۰/۲)$، ۲.ج به‌ازای $(n, \lambda_u) = (۵۰, ۰/۸)$ و ۲.د به‌ازای $(n, \lambda_u) = (۱۰۰۰, ۰/۸)$ رسم شده‌اند. انحراف معیار سه برآوردگر بسیار به‌هم نزدیک هستند و با افزایش n، میزان این انطباق بیشتر می‌شود. با افزایش k انحراف



شکل ۲: انحراف معیار ۱۰۰ بار شبیه‌سازی مونت‌کارلو از $\lambda_{u_{n,k}}^{(i)}$ در مجموعه اول به‌ازای مقادیر $\lambda_u =$ ۰/۲, ۰/۸

معیار برآوردگرها افزایش می‌یابد و میزان آن در مقادیر kی بزرگتر از $n-\sqrt{n}$، رشد چشمگیری دارد. انحراف معیار و اریبی برآوردگرها برای مجموعه‌های دوم و سوم نیز بررسی شده‌اند. طبق نتایج به‌دست آمده، مقادیر بزرگتر λ_u، انحراف معیار بیشتری به‌همراه دارد در حالی‌که در مجموعه اول تأثیر مقدار λ_u، ناچیز است. همچنین افزایش n در مجموعه‌های دوم و سوم، انحراف معیار سه برآوردگر را کاهش می‌دهد. در هر سه مجموعه به‌ازای مقادیر کوچک k، انحراف معیار برآوردگرها کاهش می‌یابد. میزان اریبی برآوردگرها در نمودارهای مربوط به ترکیب خطی محدب دو مفصل $M(u,v)$ و $\prod(u,v)$ و مفصل نرمال به‌طریق مشابهی تحلیل می‌شود. در نمودارهای مفصل نرمال میزان اریبی برآوردگرها بیشتر است. به‌ازای

$\rho = ۰$، اریبی مثبت و در $\rho = ۰/۸$، اریبی منفی است. چون اریبی و انحراف معیار سه برآوردگر در شبیه‌سازی‌های انجام شده بسیار نزدیک به‌هم هستند، از طریق این معیارها نمی‌توان بهترین برآوردگر را در حالت کلی تعیین کرد. در نتیجه بهترین برآوردگر بنا به داده‌های موجود در مسأله تعیین می‌شود. در بخش بعد روشی جدید برای انتخاب بهترین برآوردگر از بین برآوردگرهای λ_l و برآوردگرهای جدید λ_u برای داده‌های یک مثال کاربردی ارایه می‌شود.

۵ مثال کاربردی

برای مشخص کردن بهترین برآوردگر از بین سه برآوردگر ارائه شده، داده‌های مربوط به ۱۹۳۴ مورد از برگشت‌های روزانه نرخ ارز کشورهای آمریکا، تایلند و مالزی بر حسب یورو در دوره زمانی اول ژانویه ۲۰۰۴ تا پانزدهم جولای ۲۰۱۱ موجود در پایگاه بانک مرکزی اروپا بررسی می‌شود. کلایول و گاگن (۲۰۰۵)، بهترین مقدار k که به‌ازای آن می‌توان λ_l و λ_u، را به‌صورت تجربی برآورد کرد، $i_۰$ نامیده و آن‌را با توجه به رابطه (۳)، برای λ_l و λ_u به‌صورت جداگانه محاسبه کردند. براساس روش دوبریک و اسمیت (۲۰۰۵)، در محاسبه برآوردگرهای $\hat{\lambda}^{(i)}_{l_{n,k}}$، با افزایش حجم نمونه، مقدار $k_۰$ (تابعی از نمونه‌ی تصادفی است که بهترین مقدار را برای $\hat{\lambda}^{(i)}_{l_{n,k}}$ و $\hat{\lambda}^{(i)}_{u_{n,k}}$ تخمین می‌زند) تقریباً برابر با $\sqrt{n}$ می‌باشد. طبق مطالعات تجربی که بر روی داده‌های واقعی انجام شده است، نتایج زیر حاصل شده‌اند.

۱− برای برآوردگرهای مربوط به λ_u، مقدار $k_۰$ با افزایش حجم نمونه تقریباً برابر با $۰/۹۵n$ است.

۲− مقدار $k_۰$ به‌دست آمده برای λ_l و λ_u، نزدیک به $i_۰$ به‌دست آمده توسط کلایول و گاگن (۲۰۰۵) است.

بنابراین بهترین برآوردگر آن‌است که مقادیر آن در نقاط $k_۰$ و $i_۰$، نزدیک به‌هم باشد که به‌طریق زیر تعیین می‌شود:

ابتدا مقادیر $\hat{\lambda}^{(i)}_{l_{n,k}}(k_۰)$ و $\hat{\lambda}^{(i)}_{u_{n,k}}(k_۰)$ که در آن‌ها مقدار $k_۰$ برای اندازه وابستگی دمی پایین و بالا به‌ترتیب برابر با $\sqrt{n}$ و $۰/۹۵n$ است، محاسبه می‌شوند.

از بین برآوردگرهای مربوط به $\hat{\lambda}^{(i)}_{l_{n,k}}$، برآوردگری بهتر است که در آن عبارت $|\hat{\lambda}^{(i)}_{l_{n,k}}(k_\circ) - \hat{\lambda}^{(i)}_{l_{n,k}}(i_\circ)|$، مقداری کوچک‌تر داشته باشد. به‌عبارت دیگر به‌ازای آن برآوردگر، $\hat{\lambda}^{(i)}_{l_{n,k}}(k_\circ)$ و $\hat{\lambda}^{(i)}_{l_{n,k}}(i_\circ)$، به‌یکدیگر نزدیک‌تر باشند. بهترین برآوردگر مربوط به $\hat{\lambda}^{(i)}_{u_{n,k}}$ نیز به‌طریق مشابه انتخاب می‌شود.

طبق مطالب ذکر شده و با توجه به جدول ۲، در داده‌های مربوط به آمریکا-تایلند، آمریکا-مالزی و آمریکا-سنگاپور، از بین $\hat{\lambda}^{(i)}_{l_{n,k}}$ها به‌ترتیب $\hat{\lambda}^{(1)}_{l_{n,k}}$، $\hat{\lambda}^{(3)}_{l_{n,k}}$ و $\hat{\lambda}^{(3)}_{l_{n,k}}$ و از بین $\hat{\lambda}^{(i)}_{u_{n,k}}$ها به‌ترتیب $\hat{\lambda}^{(3)}_{u_{n,k}}$، $\hat{\lambda}^{(1)}_{u_{n,k}}$ و $\hat{\lambda}^{(3)}_{u_{n,k}}$، بهتر از بقیه عمل می‌کنند.

جدول ۱: اندازه وابستگی دمی پایین در داده‌های نرخ ارز

کشورها	برآوردگر	$\frac{i_\circ}{n}$	$i_\circ$	$\hat{\lambda}^{(i)}_{l_{n,k}}(\sqrt{n})$	$\hat{\lambda}^{(i)}_{l_{n,k}}(i_\circ)$	اختلاف برآوردها
آمریکا-تایلند	اول	۰/۰۱۷	۳۲/۹	۰/۵۸۷	۰/۵۹۱	۰/۰۰۴
	دوم	۰/۰۲۹	۵۶/۱	۰/۵۸۹	۰/۶۰۶	۰/۰۱۷
	سوم	۰/۰۳۱	۵۹/۹	۰/۵۷۲	۰/۵۹۳	۰/۰۲۱
آمریکا-مالزی	اول	۰/۰۴۴	۸۵/۱	۰/۶۶۶	۰/۶۳۵۱	۰/۰۳۰۹
	دوم	۰/۰۴۲	۸۱/۲	۰/۶۶۹۸	۰/۶۳۸۲	۰/۰۳۱۶
	سوم	۰/۰۲۵	۴۸/۳	۰/۶۵۸۹	۰/۶۵۹۵	۰/۰۰۰۶
آمریکا-سنگاپور	اول	۰/۰۳۸	۷۳/۵	۰/۶۹۳	۰/۶۲	۰/۰۷۳
	دوم	۰/۰۳۹	۷۵/۴	۰/۶۸۹	۰/۶۱۵	۰/۰۷۴
	سوم	۰/۰۱۸	۳۴/۸	۰/۶۷۷	۰/۷۰۹	۰/۰۳۲

جدول ۲: اندازه وابستگی دمی بالا در داده‌های نرخ ارز

کشورها	برآوردگر	$\frac{i_\circ}{n}$	$i_\circ$	$\hat{\lambda}^{(i)}_{u_{n,k}}(\circ/۹۵n)$	$\hat{\lambda}^{(i)}_{u_{n,k}}(i_\circ)$	اختلاف برآوردها
آمریکا-تایلند	اول	۰/۹۵۴	۱۸۴۵/۵	۰/۶۰۸	۰/۶۰۴	۰/۰۰۴
	دوم	۰/۹۵۳	۱۸۴۲/۵	۰/۶۰۷	۰/۶۰۵	۰/۰۰۲
	سوم	۰/۹۵۷	۱۸۵۱/۵	۰/۶۰۵	۰/۶۰۴	۰/۰۰۱
آمریکا-مالزی	اول	۰/۹۵۷	۱۸۵۰/۵	۰/۶۶۷	۰/۶۷۳	۰/۰۰۶
	دوم	۰/۹۷۲	۱۸۸۰/۵	۰/۶۵۹	۰/۶۳۵	۰/۰۲۴
	سوم	۰/۹۷۳	۱۸۸۱/۵	۰/۶۶۷	۰/۶۳۵	۰/۰۳۲
آمریکا-سنگاپور	اول	۰/۹۶۲	۱۸۶۰/۵	۰/۶۱۵	۰/۶۱۴	۰/۰۰۱۷
	دوم	۰/۹۷۲	۱۸۸۰/۵	۰/۶۱	۰/۶۱۷	۰/۰۰۷
	سوم	۰/۹۶۹	۱۸۷۳/۵	۰/۶۲۲	۰/۶۲۳	۰/۰۰۰۹

۶ بحث و نتیجه‌گیری

در این مقاله سه برآوردگر جدید برای اندازه‌های وابستگی دمی بالا ارائه و ثابت شد که این برآوردگرها سازگار و به‌طور مجانبی نااریب‌اند. در یک مطالعه شبیه‌سازی از سه مفصل خاص، خواص مجانبی برآوردگرها مورد بررسی قرار گرفت و ملاحظه شد که اریبی برآوردگرها بنا به مفصل مورد بررسی تغییر می‌کند. همچنین در هر سه خانواده مفصل، اریبی برآوردگرها بسیار نزدیک به‌هم بوده و با افزایش حجم نمونه، نمودارهای آن‌ها تقریباً منطبق بر هم می‌شود. انحراف معیار این سه برآوردگر نیز بسیار نزدیک به‌هم است. بنابراین با استفاده از شبیه‌سازی‌ها نمی‌توان بهترین برآوردگر را در حالت کلی تعیین کرد. پس تنها راه این است که بهترین برآوردگر بنا به داده‌های موجود در مسأله تعیین شود. برای این منظور در این مقاله روشی جدید ارائه شد که در آن ابتدا بهترین مقدار k $(k_\circ)$ که تابعی از n است، تعیین می‌شود. سپس $\hat{\lambda}^{(i)}_{l_{n,k}}(k_\circ)$ و $\hat{\lambda}^{(i)}_{u_{n,k}}(k_\circ)$، بهترین مقدار را برای λ_l و λ_u تخمین می‌زنند. از طرفی کلایول و گاگن (۲۰۰۵)، یک مقدار تجربی را به‌عنوان بهترین مقدار k معرفی کرده‌اند. این مقدار که $i_\circ$ نام‌گذاری شده است، با استفاده از داده‌های موجود در مسأله و تعاریف λ_l و λ_u تعیین می‌شود. بهترین برآوردگر از نظر ما برآوردگری است که مقادیر تقریباً یکسانی را در نقاط $k_\circ$ و $i_\circ$ اتخاذ کند.

تقدیر و تشکر

نویسندگان مقاله از پیشنهادات ارزنده هیئت داوران محترم مجله و همچنین از قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد، کمال تشکر را دارند.

مراجع

Coles, S. G., Heffernan, J. E. and Tawn, J. A. (1999), Dependence Measure for Extreme Value Analyses, *Extremes*, **2**, 339-365.

Caillault, C. and Guegan, D. (2005), Empirical Estimation of Tail Dependence Using Copulas: Application to Asian Markets, *Quantitative Finance*, **5**, 489-501.

Dobric, J. and Schmid, F. (2005), Nonparametric Estimation of the Lower Tail Dependence in Bivariate Copulas, *Journal of Applied Statistics*, **32**, 387-407.

Druet-Mari, D. and Kotz, S. (2001), *Correlation and Dependence*, Imperial College Press, London.

Deheuvels, P. (1979), La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d'indépendance, *Académie Royal de Belgique, Bulletin de la Classe dess Science*, **65**, 274-292.

Embrechts, P., McNeil, A. and Straumann, D. (2002), Correlation and Dependency in Risk Management: Properties and Pitfalls, in: M. A. H. Dempster(Ed), Risk Management: Value at Risk and Beyond, *Cambridge University Press*, 176-223.

Embrechts, P., Lindskog, F. and McNeil, A. (2003), Modeling Dependence with Copulas and Applications to Risk Management, in Rachev, S.(Ed.), Handbook of Heavy Tailed Distributions in Finance, *Elsevier*, 329-384.

Fermanian, J. D., Radulović, D. and Wegkamp, M. (2004), Weak Convergence of Empirical Copula Processes, *Bernoulli*, **10**, 847-860.

Gumbel, E. J. (1960), Bivariate Exponential Distributions, *Journal of the American Statistical Association*, **55**, 698-707.

Klein, L., Fischer, M. and Pleier, T. (2011), Weighted Power Mean Copulas: Theory and Application, *Discussion Papers, Department of Statistics and Econometrics University of Erlangen-Nurnberg, Germany.*

Joe, H. (1997), *Multivariate Models and Dependence Concepts*, Chapman and Hal, London.

Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer, New York.

Poulin, A., Huard, D., Favre, A. C. and Pugin, S. (2007), Importance of Tail Dependence in Bivariate Frequency Analysis, *Journal of Hydrologic Engineenng*, **12**, 394-403.

Sibuya, M. (1960), Bivariate Extreme Statistics, *Annals of the Institute of statistical Mathematics*, **11(2)**, 195-210.

Sklar, A. (1959), Fonctions de rèpartition à n Dimensions et Leurs Marges, *Publications of the Institute of Statistics, Université de Paris*, **8**, 229-231.