

تحلیلی از برآوردهای اندازه وابستگی دمی بالا

محمد امینی، هادی جباری نو قابی، 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 برآورده جدید به روش ناپارامتری برای اندازه وابستگی دمی بالا به دست آورده و نشان داده می‌شود که برآوردهای سازگار و به طور مجانبی ناریب هستند. سپس با شبیه‌سازی مونت‌کارلو از سه مفصل متفاوت، این سه برآورده با هم مقایسه شده و با به کارگیری داده‌های واقعی روشی جدید برای انتخاب بهترین برآورده ارائه می‌شود.

واژه‌های کلیدی: تابع مفصل، اندازه وابستگی دمی بالا، اندازه وابستگی دمی پایین، برآورده، ناریب، 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، α اسپیرمن، τ کندال، γ جینی و غیره هر کدام بنا به موقعیت و ویژگی‌های منحصر به فردی که دارند، مورد استفاده قرار می‌گیرند. امپرنس 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 موضوع بنای ریاضی (۲۰۱۰): ۶۲H۲۰

۱۲۰ تحلیلی از برآوردهای اندازه وابستگی دمی بالا

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.

سیبیویا (۱۹۶۰) نوعی از اندازه‌های وابستگی را معرفی کرد که دارای ویژگی‌های فوق بوده و اندازه وابستگی دمی بالا^۱ نام گرفته است. این اندازه به دو نوع اندازه وابستگی دمی بالا و پایین تفکیک می‌شود که هر یک در فاصله [۱۰, ۵۰] قرار دارند. نزدیک شدن آن‌ها به یک بیانگر وابستگی قوی بوده و محدوده صفر، محدوده استقلال متغیرهاست. دو بریک و اسمیت (۲۰۰۵) سه نوع برآوردهای ناپارامتری برای اندازه وابستگی دمی پایین معرفی کرده و برخی از ویژگی‌های این برآوردهای از قبیل سازگاری و نااریبی مجانبی را مورد بررسی قرار دادند. در این تحقیق سه نوع برآوردهای برای اندازه وابستگی دمی بالا به دست آورده و نشان داده می‌شود که این برآوردهای، سازگار و به طور مجانبی نااریب هستند. همچنین با شبیه‌سازی مونت‌کارلو رفتار مجانبی این سه برآوردهای در سه زمینه تحت بررسی (مفصل گامبل، ترکیب محدب مفصل‌های حاصل‌ضرب (u, v) و (u, u)) و کران بالای فرشه $M(u, v)$ و مفصل نرمال دوستگیر) مورد مقایسه قرار می‌گیرد. در بخش ۲ تعاریف مورد نیاز ارائه می‌شود. بخش ۳ شامل معرفی برآوردهای جدید مربوط به اندازه وابستگی دمی بالا و طریقه به دست آوردن آن‌هاست. خواص برآوردهای از جمله خواص سازگاری و نااریبی مجانبی آن‌ها در بخش ۴ ارائه می‌شود. در این بخش به دو روش نموداری و محاسباتی، خواص برآوردهای مربوط به اندازه وابستگی دمی بالا مورد بررسی قرار می‌گیرد. در بخش ۵ روشهای جدید برای انتخاب بهترین برآوردهای از بین سه برآوردهای 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 اندازه‌های وابستگی دمی

اندازه‌های وابستگی دمی برای اندازه‌گیری میزان وابستگی بین متغیرها در پیشامدهای نادر مانند بحران‌های اقتصادی به کار 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 گوشی بالای سمت راست مربع I^2 (اندازه وابستگی دمی بالا) و گوشی پایین سمت چپ آن (اندازه وابستگی دمی پایین) اندازه می‌گیرند.

تعريف 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

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

$$\lambda_l = \lim_{t \rightarrow 0^+} P[Y \leq G^{-1}(t) \mid X \leq F^{-1}(t)] \quad (2)$$

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 t ام 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 $H(x, y) = C(F(x), G(y))$ و G باشند، آنگاه مفصل C وجود دارد به قسمی که C یکتاست. اگر C که در آن اگر توابع توزیع F و G مطلقاً پیوسته باشند، مفصل C یکتاست. اگر تابع مفصل بین X و Y باشد، در صورت وجود حدود موجود در (۱) و (۲) داریم

$$\lambda_u = 2 - \lim_{t \rightarrow 1^-} \frac{1 - C(t, t)}{1 - t}, \quad \lambda_l = \lim_{t \rightarrow 0^+} \frac{C(t, t)}{t}. \quad (3)$$

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 برآوردهای λ_u

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

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

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{x \leq x(i), y \leq y(j) \text{ نمونه که}}{n} \quad (4)$$

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

برآوردهای اول: اولین برآوردهای برآورده از رابطه (۱) و با به کار گیری قضیه اسکلار به صورت

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

به دست می‌آید. بنابراین

$$\hat{\lambda}_{u_{n,i}}^{(1)} = 2 - \lim_{i \rightarrow n^-} \frac{1 - C_n\left(\frac{i}{n}, \frac{i}{n}\right)}{1 - \frac{i}{n}}. \quad (5)$$

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

$$\hat{\lambda}_{u_{n,k}}^{(1)} = 2 - \frac{1 - C_n\left(\frac{k}{n}, \frac{k}{n}\right)}{1 - \frac{k}{n}} \quad (6)$$

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

برآورده گر دوم: دومین برآورده گر پس از تعیین مدل رگرسیونی

$$C_n\left(\frac{i}{n}, \frac{i}{n}\right) = 1 - \left(1 - \frac{i}{n}\right)(2 - \lambda_u) + \varepsilon_i, \quad i = k, \dots, n$$

با جمله خطای ε_i و از روش کمترین توان های دوم به صورت

$$\hat{\lambda}_{u_{n,k}}^{(2)} = 2 - \frac{\sum_{i=k}^n \left(1 - \frac{i}{n}\right) [1 - C_n\left(\frac{i}{n}, \frac{i}{n}\right)]}{\sum_{i=k}^n \left(1 - \frac{i}{n}\right)^2}, \quad (V)$$

به دست می آید. این رابطه رگرسیونی با توجه به تعریف λ_u به صورت

$$\begin{aligned} \lambda_u &= 2 - \lim_{u \rightarrow 1} \frac{1 - C(u, u)}{1 - u} \\ (2 - \lambda_u)(1 - u) &= 1 - C(u, u) + o(1 - u) \end{aligned}$$

$$C(u, u) = 1 - (2 - \lambda_u)(1 - u) + \varepsilon_i$$

نتیجه می شود و برآورده گر مورد نظر با مشتق گیری از مجموع توان های دوم

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

و حل معادله

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

به صورت

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

به دست می آید. با توجه به تعریف λ_u در (۳) برای به دست آوردن برآورده گرها مقادیر i نزدیک به n ($i = k, \dots, n$) در نظر گرفته می شود.

برآوردهای سوم: برای محاسبه سومین برآوردهای مفصل C با ترکیب خطی محدب از مفصلهای M و W (کرانهای بالا و پایین فرشته)، به صورت

$$C(u, v) = \lambda_u M(u, v) + (1 - \lambda_u) W(u, v) \quad (\text{۸})$$

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 به ترتیب 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 $\hat{\lambda}_{u_{n,k}}^{(3)}$ ، با استفاده از رابطه

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

به دست می‌آید. با توجه به تعریف λ_u ، برآوردهای $\hat{\lambda}_{u_{n,k}}^{(3)}$ به ازای مقادیر k نزدیک به n و در فاصله $n - \sqrt{n} < k < n - \frac{n}{4}$ ، تعریف شدند، بنابراین $\max[2 \frac{i}{n} - 1, 0] = 2 \frac{i}{n} - 1$ و در نتیجه

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

که از آن برآوردهای

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

حاصل می‌شود.

۴ خواص برآوردهای

در این بخش، خواص مجانبی برآوردهای 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 $\lim_{n \rightarrow \infty} \frac{k}{n} = 1$ ، با بزرگ شدن n داریم

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

به عبارت دیگر هر سه برآوردگر مربوط به λ_u ، برآوردگرهایی سازگار و به طور مجانبی نالریب هستند.

برهان در ابتدا سازگاری برآوردگر نوع (۱) بررسی می شود.

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

که در آن k ، هم زمان با افزایش n به n نزدیک می شود. سرعت نزدیک شدن k به n کمتر از سرعت نزدیک شدن n به بی نهایت است و همواره رابطه $\sqrt{n} - k < n - k$ برقرار است. چون $\lim_{n \rightarrow \infty} k = n$ داریم

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

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

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

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

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

^۲ Functional limit theorem

برای برآوردگر نوع (۲) داریم

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

چون $k \leq i \leq n$ و $\dim_{n \rightarrow \infty} k = n$, مقدار i به n نزدیک می‌شود. به علاوه چون

$$\lim_{i \rightarrow n} \frac{\lambda - \frac{i}{n} + C_n(\frac{i}{n}, \frac{i}{n}) - \lambda_u(\lambda - \frac{i}{n})}{\lambda - \frac{i}{n}} = \lambda_u - \lambda_u = 0,$$

داریم

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

بنابراین با جایگذاری

$$C_n(\frac{i}{n}, \frac{i}{n}) = \lambda_u(\lambda - \frac{i}{n}) - \lambda + \frac{i}{n} + o(\lambda - \frac{i}{n}) \quad (10)$$

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

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

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

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

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

با توجه به این که تمامی برآوردهای معرفی شده سازگار هستند، $\hat{\lambda}_{u_{n,k}}^{(i)}$ در احتمال به λ_u همگرایست. از طرفی چون اندازه‌های وابستگی دمی در فاصله $[1, \infty]$ واقع می‌شوند، برآوردهای آنها نیز در این فاصله قرار دارند. بنابراین $|\hat{\lambda}_{u_{n,k}}^{(i)} - \lambda_u| \leq 1$ و در نتیجه $(\hat{\lambda}_{u_{n,k}}^{(i)} - \lambda_u)$ ، انتگرال پذیر یکنواخت است و طبق روابط همگرایی داریم

$$\lim_{n \rightarrow \infty} P(|\hat{\lambda}_{u_{n,k}}^{(i)} - \lambda_u| \geq \epsilon) = 0, \quad 0 \leq \lambda_u \leq 1 \text{ a.e.}, \quad i = 1, 2, 3$$

بنابراین λ_u ، یعنی این برآوردها به طور مجانبی ناریب هستند.

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

برای بررسی خواص برآوردهای $\hat{\lambda}_{u_{n,k}}^{(i)}$ $\{i = 1, 2, 3\}$ ، از طریق شبیه‌سازی مونت‌کارلو در نمونه‌های متناهی، سه مجموعه متفاوت خانواده مفصل گامبل (گامبل، ۱۹۶۰)، خانواده ترکیب محدب M و Π و خانواده مفصل نرمال با توابع مفصل زیر در نظر گرفته می‌شود.
خانواده مفصل گامبل:

$$C_{1\theta}(u, v) = \exp \{-[(-\log(u))^\theta + (-\log(v))^\theta]^{1/\theta}\}, \quad \theta \geq 1$$

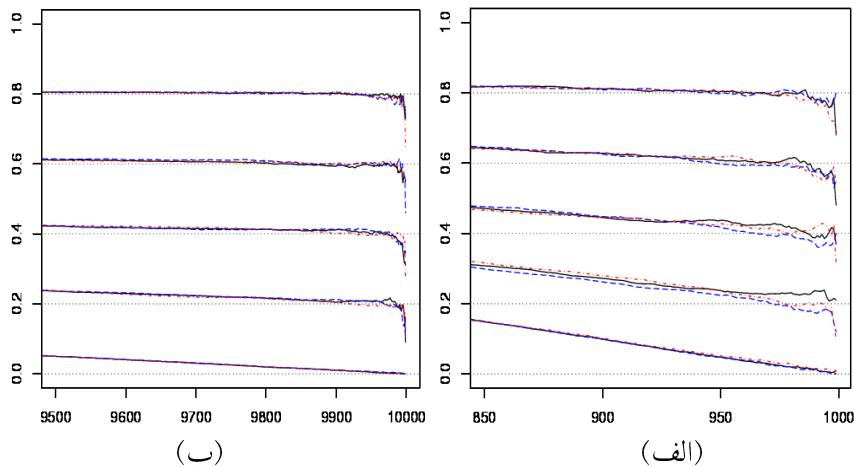
خانواده ترکیب محدب و M :

$$C_{2\theta}(u, v) = \theta M(u, v) + (1 - \theta) \prod(u, v), \quad 0 \leq \theta \leq 1$$

خانواده مفصل نرمال:

$$C_\rho(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left(-\frac{s^2 - 2\rho st + t^2}{2(1-\rho^2)} \right) ds dt.$$

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

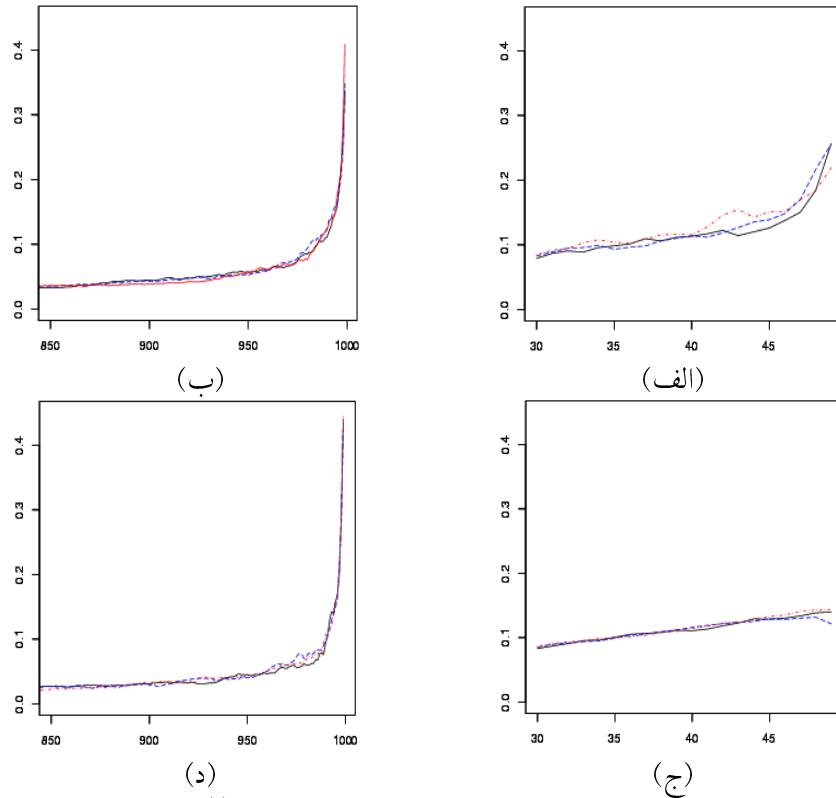


شکل ۱: میانگین‌های 100 بار شبیه‌سازی مونت کارلو از $\lambda_{u_{n,k}}^{(i)}$ برای مجموعه اول با $n = 10000$ و $k = 1, 2, 3$ برای $\lambda_u = 0, 0/2, 0/4, 0/6, 0/8$ الف:

$$n = 10000$$

$B = 100$ در نظر گرفته شده و از نماد خطوط پیوسته برای $E(\hat{\lambda}_{u_{n,k}}^{(1)})$, خط چین برای $E(\hat{\lambda}_{u_{n,k}}^{(2)})$ و نقطه چین برای $E(\hat{\lambda}_{u_{n,k}}^{(3)})$ استفاده شده است. محدوده رسم نمودار، مقادیر k 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 برآورده اریب هستند و میزان اریبی آن‌ها به λ_u , k و n بستگی دارد. با افزایش λ_u از اریبی برآوردهای کاسته می‌شود. در مقادیر k کوچک، اریبی مشبت و به‌ازای مقادیر k نزدیک به n اریبی منفی است. به‌طور کلی رفتار سه برآورده $\hat{\lambda}_{u_{n,k}}^{(i)}$ بسیار نزدیک بهم بوده و با بزرگ شدن مقدار n نمودارهای مربوط به آن‌ها تقریباً منطبق بر هم خواهد شد. با افزایش حجم نمونه، اریبی برآوردهای تا حد قابل ملاحظه‌ای کاهش می‌یابد و می‌توان گفت اگر n به اندازه‌ی کافی بزرگ باشد، برآوردهای ناریب خواهند بود. انحراف معیار برآوردهای مجموعه اول در شکل ۲ نمایش داده شده‌اند. نمودارهای ۲.الف به‌ازای $(n, \lambda_u) = (50, 0/2)$, ۲.ب به‌ازای $(n, \lambda_u) = (1000, 0/2)$, ۲.ج به‌ازای $(n, \lambda_u) = (50, 0/8)$ و ۲.د به‌ازای $(n, \lambda_u) = (1000, 0/8)$ رسم شده‌اند. انحراف معیار سه برآورده بسیار به‌هم نزدیک هستند و با افزایش n میزان این انطباق بیشتر می‌شود. با افزایش k انحراف



شکل ۲: انحراف معیار 10^0 بار شبیه‌سازی مونت کارلو از $\lambda_{u,n,k}^{(i)}$ در مجموعه اول به‌ازای مقادیر $\lambda_u = 0/2, 0/8$

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

۱۳۰ تحلیلی از برآوردهای اندازه وابستگی دمی بالا

$\rho = \rho$, اریبی مثبت و در $0/8 = \rho$, اریبی منفی است. چون اریبی و انحراف معیار سه برآوردهگر در شبیه‌سازی‌های انجام شده بسیار نزدیک به هم هستند، از طریق این معیارها نمی‌توان بهترین برآوردهگر را در حالت کلی تعیین کرد. در نتیجه بهترین برآوردهگر بنا به داده‌های موجود در مسئله تعیین می‌شود. در بخش بعد روشی جدید برای انتخاب بهترین برآوردهگر از بین برآوردهای λ_u و برآوردهای جدید λ_l برای داده‌های یک مثال کاربردی ارایه می‌شود.

۵ مثال کاربردی

برای مشخص کردن بهترین برآوردهگر از بین سه برآوردهگر 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_u و λ_l را به صورت تجربی برآورد کرد، با نامیده و آنرا با توجه به رابطه (۳)، برای λ_u و λ_l به صورت جداگانه محاسبه کردند. براساس روش دوبسیک و اسمیت (۲۰۰۵)، در محاسبه برآوردهای $\hat{\lambda}_{l_{n,k}}^{(i)}$ با افزایش حجم نمونه، مقدار k (تابعی از نمونه‌ی تصادفی) است که بهترین مقدار را برای $\hat{\lambda}_{l_{n,k}}^{(i)}$ و $\hat{\lambda}_{u_{n,k}}^{(i)}$ (تحمیل می‌زند) تقریباً برابر با \sqrt{n} می‌باشد. طبق مطالعات تجربی که بر روی داده‌های واقعی انجام شده است، نتایج زیر حاصل شده‌اند.

۱ - برای برآوردهای مربوط به λ_u ، مقدار k با افزایش حجم نمونه تقریباً برابر با $0/95n$ است.

۲ - مقدار k به دست آمده برای λ_u و λ_l نزدیک به $0/95n$ است.

بنابراین بهترین برآوردهگر آن است که مقادیر آن در نقاط k و i نزدیک به هم باشد که به طریق زیر تعیین می‌شود:

ابتدا مقادیر $(k_0 \hat{\lambda}_{l_{n,k}}^{(i)})$ و $(k_0 \hat{\lambda}_{u_{n,k}}^{(i)})$ که در آن‌ها مقدار k برای اندازه وابستگی دمی پایین و بالا به ترتیب برابر با \sqrt{n} و $0/95n$ است، محاسبه می‌شوند.

از بین برآوردهای مربوط به $\hat{\lambda}_{l_{n,k}}^{(i)}$, برآوردگری بهتر است که در آن عبارت $|\hat{\lambda}_{l_{n,k}}^{(i)}(k_0) - \hat{\lambda}_{l_{n,k}}^{(i)}(i_0)|$ برآورده، $\hat{\lambda}_{l_{n,k}}^{(i)}$ بهیکدیگر نزدیک تر باشند. بهترین برآوردهای مربوط به $\hat{\lambda}_{u_{n,k}}^{(i)}$ نیز به طریق مشابه انتخاب می‌شود.

طبق مطالعه ذکر شده و با توجه به جدول ۲، در داده‌های مربوط به آمریکا-تایلند، آمریکا-مالزی و آمریکا-سنگاپور، از بین $\hat{\lambda}_{l_{n,k}}^{(i)}$ ها به ترتیب $\hat{\lambda}_{l_{n,k}}^{(1)}$ و $\hat{\lambda}_{l_{n,k}}^{(2)}$ و از بین $\hat{\lambda}_{u_{n,k}}^{(i)}$ ها به ترتیب $\hat{\lambda}_{u_{n,k}}^{(1)}$ و $\hat{\lambda}_{u_{n,k}}^{(2)}$ ، بهتر از بقیه عمل می‌کنند.

جدول ۱: اندازه وابستگی دمی پایین در داده‌های نرخ ارز

کشورها	برآوردها	$\frac{1}{n}$	i_0	$\hat{\lambda}_{l_{n,k}}^{(i)}(\sqrt{n})$	اختلاف برآوردها	$\hat{\lambda}_{l_{n,k}}^{(i)}(i_0)$
آمریکا-تایلند	اول	۰/۰۱۷	۰/۰۸۹	۵۶/۱	۰/۰۲۹	۰/۶۰۶
	دوم	۰/۰۲۱	۰/۰۵۷۲	۵۹/۰	۰/۰۲۱	۰/۰۵۹۳
	سوم	۰/۰۲۰۹	۰/۰۶۶	۸۵/۱	۰/۰۴۴	۰/۶۲۵۱
آمریکا-مالزی	اول	۰/۰۲۱۶	۰/۰۶۹۸	۸۱/۲	۰/۰۲۲	۰/۶۸۲
	دوم	۰/۰۰۰۶	۰/۰۵۸۹	۴۸/۳	۰/۰۲۵	۰/۶۵۹۵
	سوم	۰/۰۷۳	۰/۰۶۲	۷۳/۵	۰/۰۳۸	۰/۶۲
آمریکا-سنگاپور	اول	۰/۰۷۴	۰/۰۶۸۹	۷۵/۴	۰/۰۳۹	۰/۶۱۵
	سوم	۰/۰۲۲	۰/۰۶۷۷	۲۴/۸	۰/۰۱۸	۰/۷۰۹

جدول ۲: اندازه وابستگی دمی بالا در داده‌های نرخ ارز

کشورها	برآوردها	$\frac{1}{n}$	i_0	$\hat{\lambda}_{u_{n,k}}^{(i)}(\sqrt{95n})$	اختلاف برآوردها	$\hat{\lambda}_{u_{n,k}}^{(i)}(i_0)$
آمریکا-تایلند	اول	۰/۰۵۴	۱۸۴۵/۵	۰/۶۰۸	۰/۰۰۴	۰/۶۰۴
	دوم	۰/۹۵۳	۱۸۴۲/۵	۰/۶۰۷	۰/۰۰۲	۰/۶۰۵
	سوم	۰/۹۵۷	۱۸۵۱/۵	۰/۶۰۵	۰/۰۰۱	۰/۶۰۴
آمریکا-مالزی	اول	۰/۹۵۷	۱۸۵۰/۵	۰/۶۶۷	۰/۰۰۶	۰/۱۷۳
	دوم	۰/۹۷۲	۱۸۸۰/۵	۰/۶۵۹	۰/۰۲۴	۰/۶۳۵
	سوم	۰/۹۷۳	۱۸۸۱/۵	۰/۶۶۷	۰/۰۳۲	۰/۶۳۵
آمریکا-سنگاپور	اول	۰/۹۶۲	۱۸۶۰/۵	۰/۶۱۵	۰/۰۰۱۷	۰/۶۱۴
	دوم	۰/۹۷۲	۱۸۸۰/۵	۰/۶۱	۰/۰۰۷	۰/۶۱۷
	سوم	۰/۹۶۹	۱۸۸۳/۵	۰/۶۲۲	۰/۰۰۰۹	۰/۶۲۳

۶ بحث و نتیجه‌گیری

در این مقاله سه برآوردهای جدید برای اندازه‌های وابستگی دمی بالا ارائه و ثابت شد که این برآوردهای سازگار و به طور مجانی ناریب‌اند. در یک مطالعه شبیه‌سازی از سه مفصل خاص، خواص مجانبی برآوردهای مورد بررسی قرار گرفت و ملاحظه شد که اریبی برآوردهای بنا به مفصل مورد بررسی تغییر می‌کند. همچنین در هر سه خانواده مفصل، اریبی برآوردهای بسیار نزدیک به هم بوده و با افزایش حجم نمونه، نمودارهای آنها تقریباً منطبق بر هم می‌شود. انحراف معیار این سه برآوردهای نیز بسیار نزدیک به هم است. بنابراین با استفاده از شبیه‌سازی‌ها نمی‌توان بهترین برآوردهای را در حالت کلی تعیین کرد. پس تنها راه این است که بهترین برآوردهای بنا به داده‌های موجود در مسئله تعیین شود. برای این منظور در این مقاله روشی جدید ارائه شد که در آن ابتدا بهترین مقدار k (k_0) که تابعی از n است، تعیین می‌شود. سپس (k_0) و $\hat{\lambda}_{l_{n,k}}^{(i)}$ بهترین مقدار را برای λ_l و λ_u تخمین می‌زنند. از طرفی کلایول و گاگن (۲۰۰۵)، یک مقدار تجربی را به عنوان بهترین مقدار k معرفی کرده‌اند. این مقدار که i نام‌گذاری شده است، با استفاده از داده‌های موجود در مسئله و تعاریف λ_l و λ_u تعیین می‌شود. بهترین برآوردهای از نظر ما برآوردهایی است که مقادیر تقریباً یکسانی را در نقاط k و i اتخاذ کند.

تقدیر و تشکر

نویسندهای این 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 des 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.

تحليلی از برآوردهای اندازه وابستگی دمی بالا ۱۳۴

- 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 Engineeringng*, **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 répartition à n Dimensions et Leurs Marges, *Publications of the Institute of Statistics, Université de Paris*, **8**, 229-231.