

# The 3<sup>rd</sup> Workshop on Copula and Various types of Dependencies 25-26 February 2015

Department of Statistics,  
Shahid Bahonar University of Kerman, Iran  
Ordered and Spatial Data Center of Excellence,  
Ferdowsi University of Mashhad, Iran

Construction methods  
Modeling using copula  
copula Theory  
copula and its application  
in spatial statistics  
Copula and its application  
in financial mathematics

Address: Department of statistics  
Faculty of Mathematics & Computer, Shahid Bahonar University of Kerman  
P.O.Box: 76169-14111, Kerman, Iran. Tel/Fax: +98 3433220057  
website: <http://statistics.fu.ac.ir> Email: [statisticsdept@mail.uk.ac.ir](mailto:statisticsdept@mail.uk.ac.ir)



باسمه تعالی



## مجموعه مقالات فارسی

سومین کارگاه آموزشی

# نظریه مفصل و انواع وابستگی‌ها

بخش آمار، دانشکده ریاضی و کامپیوتر

دانشگاه شهید باهنر کرمان

با همکاری

قطب علمی داده‌های ترتیبی و فضایی

دانشگاه فردوسی مشهد

۶ و ۷ اسفند ماه ۱۳۹۳

## پیشگفتار

به منظور تبادل نظر اعضای هیأت علمی و ارتقای سطح علمی دانشجویان تحصیلات تکمیلی دانشگاه‌ها، مراکز آموزش عالی کشور و کاربران نظریه مفصل، سومین کارگاه مفصل و انواع وابستگی‌ها روزهای ۶ و ۷ اسفندماه سال ۱۳۹۳ توسط بخش آمار، دانشکده ریاضی و کامپیوتر دانشگاه شهید باهنر کرمان با همکاری قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد و بخش‌های آمار دانشگاه‌های فردوسی مشهد، اصفهان، تربیت مدرس، یزد و انجمن آمار ایران در محل دانشکده ریاضی و کامپیوتر دانشگاه شهید باهنر کرمان، برگزار گردید. در این کارگاه دو روزه بیش از ۴۰ نفر از اعضای هیأت علمی، دانشجویان تحصیلات تکمیلی و کاربران مباحث مرتبط با نظریه مفصل شرکت نمودند. پس از فراخوان کارگاه، کمیته علمی اقدام به بررسی مقالات دریافتی نمود و در نهایت ارائه ۱۹ مقاله را مورد موافقت قرار داد.

خلاصه سخنرانی‌های عمومی این کارگاه در پایگاه اطلاعاتی قطب علمی داده‌های ترتیبی و فضایی دانشگاه فردوسی مشهد به نشانی <http://osdce.um.ac.ir> قابل دسترسی است. متن کامل تعداد ۶ مقاله‌ی فارسی و ۱۲ مقاله‌ی انگلیسی پذیرفته شده پس از تأیید کمیته علمی در مجموعه مقالات تدوین شده است که شامل دو بخش مقالات انگلیسی و فارسی می‌باشد.

**این مجموعه شامل خلاصه مقالات سخنرانی و پوسترهای فارسی و انگلیسی ارائه شده در سومین کارگاه آموزشی نظریه مفصل و انواع وابستگی‌ها است. بدیهی است که ارائه دهنده سخنرانی یا پوستر مسئول صحت مطالب، اصالت آنها و ویرایش علمی-ادبی مقاله خود بوده است و قطب علمی داده‌های ترتیبی و فضایی مسئولیتی در این خصوص ندارد.**

برگزاری کارگاه مذکور مرهون تلاش بی‌وقفه‌ی کمیته‌های علمی و اجرایی، پشتیبانی بخش آمار و مسئولین محترم دانشکده ریاضی و کامپیوتر دانشگاه شهید باهنر کرمان بخصوص جناب آقای دکتر فدایی ریاست محترم دانشکده بوده است. در این راستا بر خود لازم می‌دانیم به تاسی از حدیث شریف «من لم یشکر المخلوق لم یشکر الخالق» از زحمات و تلاش‌های صادقانه همه عزیزان در معاونت‌های محترم دانشگاه، دانشکده، حراست و انتظامات دانشگاه، حوزه ریاست و روابط عمومی و تمام همکاران محترم هیأت علمی و کارکنان عزیز و همچنین اعضای انجمن علمی بخش آمار که در امر برگزاری سومین کارگاه نظریه مفصل و انواع وابستگی‌ها مشارکت داشته‌اند، تشکر و قدردانی نمایم.

با آرزوی توفیق الهی

بخش آمار

دانشکده ریاضی و کامپیوتر

**دبیر کمیته علمی :**

دکتر ماه بانو تاتا، دانشگاه شهید باهنر کرمان

**اعضای کمیته علمی (به ترتیب حروف الفبا):**

- دکتر محمد امینی ، دانشگاه فردوسی مشهد
- دکتر علی دولتی، دانشگاه یزد
- دکتر محسن رضاپور، دانشگاه شهید باهنر کرمان
- دکتر محمدحسین علامت‌ساز، دانشگاه اصفهان
- دکتر غلامرضا محتشمی برزادران، دانشگاه فردوسی مشهد
- دکتر محسن محمدزاده، دانشگاه تربیت مدرس تهران

**دبیر کمیته اجرایی :**

دکتر رضا پورموسی، دانشگاه شهید باهنر کرمان

**اعضای کمیته اجرایی (به ترتیب حروف الفبا):**

- دکتر جعفر احمدی، دانشگاه فردوسی مشهد
- آقای رضا ایرانمنش، دانشگاه شهید باهنر کرمان
- دکتر محسن خسروی، دانشگاه شهید باهنر کرمان
- دکتر هادی جباری نوقایی ، دانشگاه فردوسی مشهد
- دکتر ایوب شیخی، دانشگاه شهید باهنر کرمان
- آقای هوشنگ طاهری‌زاده، دانشگاه شهید باهنر کرمان
- دکتر علیرضا عرب‌پور، دانشگاه شهید باهنر کرمان

### دبیر انجمن علمی بخش آمار:

آقای علیرضا ابراهیمی، دانشگاه شهید باهنر کرمان

### اعضای انجمن علمی بخش آمار (به ترتیب حروف الفبا):

- آقای محمدمین افشارمنش، دانشگاه شهید باهنر کرمان
- خانم حدیث چهردینی، دانشگاه شهید باهنر کرمان
- خانم زهرا خالقی، دانشگاه شهید باهنر کرمان
- خانم منصوره رازیان، دانشگاه شهید باهنر کرمان
- خانم سمانه صادقی، دانشگاه شهید باهنر کرمان
- آقای آرش فروهری، دانشگاه شهید باهنر کرمان

### محورها:

- روش‌های ساخت تابع مفصل
- توابع مفصل و مفاهیم وابستگی
- مدل‌سازی با استفاده از تابع مفصل
- کاربردهای توابع مفصل
- توابع مفصل و آمار فضایی

## فهرست

### بررسی ساختار وابستگی مفصل سیبویا

ع. احمدی، م. امینی و ه. جباری نوقایی ..... ۱

### تحلیل و بررسی نمودار کای و نقش آن در انتخاب مفصل مناسب

م. پوراحمدی و ا. حیدری ..... ۷

### برآورد پارامتر و انتخاب مدل ارشمیدسی بر اساس توزیع‌های متقارن تحت نرم $L_1$

ف. خوشنودی ..... ۱۱

### تخمین دبی پیک سیلاب با استفاده از تابع مفصل

س. زراعتی، م. ذونعمت کرمانی و م. رضاپور ..... ۱۵

### تشخیص وابستگی با استفاده از نمودارها

م. شاکر هدایت آباد، م. اصفهانی و م. امینی ..... ۱۹

### کاربردهایی از ترتیب تصادفی همبستگی

و. محتشمی، م. امینی و ج. احمدی ..... ۲۷



## بررسی ساختار وابستگی مفصل سیبویا

احمدی، ع. <sup>۱</sup>، امینی، م. <sup>۲</sup>، و جباری نوقابی، ه. <sup>۳</sup>

۱، ۲، ۳ گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد

author1@um.ac.ir

author2@um.ac.ir

author3@um.ac.ir

چکیده. در این مقاله خانواده‌ای از مفصل‌های تولید شده توسط تابع وابستگی سیبویا معرفی و سپس ساختار وابستگی آن بررسی می‌شود. در ادامه به بررسی ویژگی‌های این تابع وابستگی پرداخته می‌شود و علاوه بر این معرفی و بیان برآوردهای ناپارامتری آن از نتایج دیگر این مقاله می‌باشد.

### ۱. پیش‌گفتار

مفهوم تابع مفصل اولین بار در سال ۱۹۵۹ توسط اسکالر مطرح شد و پس از آن توسط محققان دیگر مورد بررسی قرار گرفت که مهم‌ترین تحقیقات انجام شده در رابطه با مفاهیم وابستگی و ترتیب‌های تصادفی وابسته بوده است. می‌دانیم که بعضی از اندازه‌ها مانند ضریب همبستگی پیرسن، توی کندال و روی اسپیرمن وابستگی کلی بین دو متغیر را اندازه می‌گیرند. هرچند اگر ما زیرمجموعه‌ای از داده‌ها را در نظر بگیریم داده‌ها ممکن است رفتارهای وابستگی متفاوتی داشته باشند. برای بررسی رفتار وابستگی مکانی بردار تصادفی  $(X, Y)$  اندازه‌های وابستگی مکانی زیادی پیشنهاد داده شده است. به عنوان مثال تابع وابستگی مکانی هلاند و وانگ (۱۹۸۷)، منحنی همبستگی بچرو و دکسوم (۱۹۹۳)، اندازه وابستگی مکانی بایرامو و همکاران (۲۰۰۳)، نلسن (۱۹۹۹) و سیبویا (۱۹۶۰). یک رده جدید و نامتقارن از مفصل، اولین بار توسط سیبویا (۱۹۶۰) معرفی و برخی از ویژگی‌های آن بررسی شده است. که مفصل‌های معروف مقدار فرین، مارشال اولکین و ارشمیدسی از جمله مفصل‌هایی هستند که در این رده قرار می‌گیرند. بطور کلی هر مفصلی می‌تواند در فرم سیبویا نوشته شود.

واژگان کلیدی. مفصل، تابع وابستگی سیبویا، برآوردهای ناپارامتری.  
\* سخنران

## ۲. نتایج اصلی

در این بخش تابع وابستگی سیبویا معرفی و برخی از ویژگی‌های آن که توسط سیبویا (۱۹۶۰) و کولو و همکاران (۲۰۰۷) مطرح شده، بیان می‌شود.  
در ادامه دو برآوردگر این تابع ارائه می‌شود و رفتار حدی آن‌ها مورد ارزیابی قرار می‌گیرد.

**تعریف ۱.۲.** فرض کنید  $X$  و  $Y$  متغیرهای تصادفی تعریف شده در فضای احتمال  $(\Omega, A, P)$  باشند و  $F(x, y) = p(X \leq x, Y \leq y)$  توزیع توام  $X$  و  $Y$  باشد و  $F_1(x) = P(X \leq x)$  و  $F_2(y) = P(Y \leq y)$  توزیع‌های حاشیه‌ای آن باشند آنگاه تابع وابستگی سیبویا را می‌توان به صورت زیر تعریف کرد.

$$\Lambda(F_1(x), F_2(y)) = \frac{F(x, y)}{F_1(x)F_2(y)} \quad \forall (x, y) \in R^2 \quad (1.2)$$

$$F_1(x), F_2(y) \geq 0$$

و برای  $F_1(x) = 0, F_2(y) = 0$  زمانی می‌تواند  $\Lambda$  تعریف شود که حدود زیر به ترتیب وجود داشته باشد

$$\lim_{F_1(x) \rightarrow 0} \frac{F(x, y)}{F_1(x)F_2(y)}$$

$$\lim_{F_2(y) \rightarrow 0} \frac{F(x, y)}{F_1(x)F_2(y)}$$

**تعریف ۲.۲.** برای هر  $x, y \in R$  و  $X, Y$  و  $PQD$  می‌باشند اگر و تنها اگر

$$F(x, y) \geq F_1(x)F_2(y) \quad \forall (x, y) \in R^2$$

و همچنین  $X$  و  $Y$  می‌باشند اگر و تنها اگر

$$F(x, y) \leq F_1(x)F_2(y) \quad \forall (x, y) \in R^2$$

## ۲-۱ ویژگی‌های تابع وابستگی سیبویا

از جمله ویژگی‌های  $\Lambda$  که توسط سیبویا (۱۹۶۰) و همچنین کولو و همکاران (۲۰۰۷) بیان شده است به شرح زیر می‌باشد.

الف- کرانداری: برای هر  $x, y \in R$

$$\max(0, \frac{F_1(x) + F_2(y) - 1}{F_1(x)F_2(y)}) \leq \Lambda(F_1(x), F_2(y)) \leq \min(\frac{1}{F_1(x)}, \frac{1}{F_2(y)})$$

ب-  $X$  و  $Y$  مستقل هستند اگر و تنها اگر

$$\Lambda(F_1(x), F_2(y)) = 1$$

ج-  $X$  و  $Y$   $PQD$  می‌باشند اگر و تنها اگر

$$\Lambda(F_1(x), F_2(y)) \geq 1$$

و همچنین  $X$  و  $Y$   $NQD$  می‌باشند اگر و تنها اگر

$$\Lambda(F_1(x), F_2(y)) \leq 1$$



د- فرض کنید  $\phi(\cdot)$  و  $\psi(\cdot)$  دو تابع حقیقی مقدار باشند به قسمی که  $\phi^{-1}(\cdot)$  و  $\psi^{-1}(\cdot)$  وجود داشته باشند در این صورت:

• اگر  $\phi(\cdot)$  و  $\psi(\cdot)$  هر دو صعودی باشند آنگاه برای هر  $(x, y) \in S$  که  $S$  تکیه‌گاه بردار تصافی  $(X, Y)$  می‌باشد.

$$\Lambda_{\phi(X)\psi(Y)}(x, y) = \Lambda_{X,Y}(\phi^{-1}(x), \psi^{-1}(y))$$

• اگر  $\phi(\cdot)$  نزولی و  $\psi(\cdot)$  صعودی باشد آنگاه برای هر  $(x, y) \in S$

$$\Lambda_{\phi(X)\psi(Y)}(x, y) = \frac{1}{\bar{F}_\lambda(\phi^{-1}(x))} - \frac{F_\lambda(\phi^{-1}(x))}{\bar{F}_\lambda(\phi^{-1}(x))} \Lambda_{X,Y}(\phi^{-1}(x), \psi^{-1}(y))$$

• اگر  $\phi(\cdot)$  صعودی و  $\psi(\cdot)$  نزولی باشند آنگاه برای هر  $(x, y) \in S$

$$\Lambda_{\phi(X)\psi(Y)}(x, y) = \frac{1}{\bar{F}_\tau(\psi^{-1}(y))} - \frac{F_\tau(\psi^{-1}(y))}{\bar{F}_\tau(\psi^{-1}(y))} \Lambda_{X,Y}(\phi^{-1}(x), \psi^{-1}(y))$$

• اگر  $\phi(\cdot)$  و  $\psi(\cdot)$  توابع نزولی باشند آنگاه برای هر  $(x, y) \in S$

$$\begin{aligned} \Lambda_{\phi(X)\psi(Y)}(x, y) &= \frac{1 - F_\lambda(\phi^{-1}(x)) - F_\tau(\psi^{-1}(y))}{\bar{F}_\lambda(\phi^{-1}(x))\bar{F}_\tau(\psi^{-1}(y))} \\ &+ \frac{F_\lambda(\phi^{-1}(x))F_\tau(\psi^{-1}(y))}{\bar{F}_\lambda(\phi^{-1}(x))\bar{F}_\tau(\psi^{-1}(y))} \Lambda_{X,Y}(\phi^{-1}(x), \psi^{-1}(y)) \end{aligned}$$

## ۲-۲ مفصل سیبویا

اگر  $(X, Y)$  یک بردار تصادفی با توزیع توام  $F(x, y)$  و توابع توزیع حاشیه‌ای  $F_\lambda(x)$  و  $F_\tau(y)$  باشند آنگاه بنا به قضیه اسکالر (۱۹۵۹) و رابطه (۱.۲)، مفصل سیبویا به صورت زیر تعیین می‌شود.

$$C(u, v) = uv\Lambda(u, v) \quad \forall (u, v) \in [0, 1]^2 \quad (2.2)$$

که تمام خاصیت‌های بیان شده در بخش ۱-۲، برای تابع  $\Lambda(u, v)$  نیز برقرار است. همچنین داریم:

$$\hat{C}(u, v) = u + v - 1 + (1 - u)(1 - v)\Lambda(1 - u, 1 - v)$$

علاوه بر این اندازه‌های وابستگی دمی  $\lambda_u$  و  $\lambda_l$  و ضرایب وابستگی کندال و اسپیرمن نیز بر حسب تابع  $\Lambda(u, v)$  بیان می‌شوند.

$$\lambda_l = \lim_{u \rightarrow 0^+} u\Lambda(u, u)$$

$$\lambda_u = \lim_{u \rightarrow 1^-} \frac{(1 - u) + (1 - u) - 1 + u^2\Lambda(u, u)}{1 - u}$$

$$\tau = 4 \int_0^1 \int_0^1 uv\Lambda(u, v) duv \Lambda(u, v) - 1$$

$$\rho_s = 12 \int_0^1 \int_0^1 uv\Lambda(u, v) duv - 3$$

برخی از اعضای رده مفصل‌های سیبویا، در جدول زیر ارائه می‌شود.

## ۲-۳ برآوردگرها

جدول ۱: برخی از مفصل‌های سیبویا

مفصل	محدوده $\theta$	$\Lambda_\theta(u, v)$
$FGM(\theta)$	$[-1, 1]$	$1 + \theta(1-u)(1-v)$
$Gumbel(\theta)$	$[0, 1]$	$\exp(-\theta \ln(u) \ln(v))$
$GCA(\theta)$	$[-1, 1]$	$\exp(\theta(1-u)(1-v))$
$AMH(\theta)$	$[-1, 1]$	$(1 - \theta(1-u)(1-v))^{-1}$

در این بخش دو برآوردگر برای تابع وابستگی  $\Lambda$  ارائه می‌شود و رفتار حدی آنها مورد ارزیابی قرار می‌گیرد. در ابتدا دو برآوردگر ناپارامتری را برای  $\Lambda$  با توجه به رابطه (۱۰۲) معرفی می‌کنیم.

- برای برآورد  $\Lambda$  در فرم (۱۰۲) کولو و همکاران (۲۰۰۷) برآوردگر تجربی  $\Lambda_n$  را پیشنهاد داده اند که به وسیله روش plug-in و با استفاده از برآوردگرهای تجربی توزیع‌های حاشیه ای و توام ساخته شده است.
- برای به دست آوردن یک برآوردگر بهتر برای نمونه های کوچک مورتنین و همکاران (۲۰۰۸) برآوردگر هموار  $\hat{\Lambda}$  را برای  $\Lambda$  در فرم (۱۰۲) پیشنهاد داده اند که برآوردگرهای هموار هسته‌ای را برای توزیع‌های توام و حاشیه‌ای در نظر می‌گیرد.

**تعریف ۳.۲.** فرض کنید  $\{(X_i, Y_i); i = 1, \dots, n\}$  یک نمونه تصادفی از  $(X, Y)$  باشد و  $k: R^2 \rightarrow R$  یک تابع هسته دو متغیره باشد بطوریکه  $\int \int k(u, v) dudv = 1$  و

$$k(x, y; h_1, h_2) = \frac{1}{h_1 h_2} k\left(\frac{x}{h_1}, \frac{y}{h_2}\right) \quad h_j(n) \geq 0; j = 1, 2$$

بطوریکه  $h_j(n) \rightarrow 0$  زمانیکه  $n \rightarrow \infty$  و

$$K(x, y) = \int_{-\infty}^x \int_{-\infty}^y k(u, v) dudv$$

و

$$K(x, y; h_1, h_2) = K\left(\frac{x}{h_1}, \frac{y}{h_2}\right)$$

آنگاه برآوردگر هموار توزیع توام  $F(x, y)$  بصورت زیر می‌باشد.

$$\hat{F}(x, y) = \frac{1}{n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_1}, \frac{y - Y_i}{h_2}\right)$$

**تعریف ۴.۲.** فرض کنید  $k_j(\cdot), j = 1, 2$  توابع متقارن، کراندار و حقیقی باشند بطوریکه  $\int k_j(u) du = 1$  و

$$k_j(u; h_j) = \frac{1}{h_j} k_j\left(\frac{u}{h_j}\right)$$

و

$$K_j(w) = \int_{-\infty}^w k_j(u) du, \quad K_j(w; h_j) = K_j\left(\frac{w}{h_j}\right)$$

آنگاه برآوردگر هموار توزیع‌های حاشیه‌ای  $F_1(x)$  و  $F_2(y)$  به ترتیب به صورت زیر می‌باشند.

$$\hat{F}_1(x) = \frac{1}{n} \sum_{i=1}^n K_1\left(\frac{x - X_i}{h_1}\right)$$

$$\hat{F}_2(y) = \frac{1}{n} \sum_{i=1}^n K_2\left(\frac{y - Y_i}{h_2}\right)$$

بنابراین برآوردگر هموار  $\hat{\Lambda}$  برای  $\Lambda$  در فرم (۱.۲) به صورت زیر می‌باشد.

$$\hat{\Lambda}(\hat{F}_1(x), \hat{F}_2(y)) = \frac{\hat{F}(x, y)}{\hat{F}_1(x)\hat{F}_2(y)} \quad \forall (x, y) \in R^2; \hat{F}_1(x), \hat{F}_2(y) \geq 0 \quad (3.2)$$

حال دو برآوردگر ناپارامتری را برای  $\Lambda$  با توجه به رابطه (۲.۲) بیان می‌کنیم.  
 • مورتین و همکاران (۲۰۰۸) یک برآوردگر تجربی برای  $\Lambda$  در فرم (۲.۲) به دست آورده‌اند. فرض کنید برآوردگرهای تجربی توزیع توام و حاشیه‌ای  $X$  و  $Y$  به ترتیب به صورت زیر باشند.

$$F_n(x, y) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x, Y_i \leq y)$$

$$F_{1n}(x, y) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x)$$

$$F_{2n}(x, y) = \frac{1}{n} \sum_{i=1}^n I(Y_i \leq y)$$

آنگاه بنابر فرمانیان و همکاران (۲۰۰۴) و با استفاده از روش *plug-in* داریم

$$C_n(u, v) = F_n(F_{1n}^{-1}(u), F_{2n}^{-1}(v)) \quad \forall (u, v) \in [0, 1]^2$$

که مفصل تجربی می‌باشد.

بنابراین برآوردگر تجربی  $\Lambda$  با توجه به رابطه (۲.۲) به صورت زیر می‌باشد.

$$\Lambda_n(u, v) = \frac{C_n(u, v)}{uv} \quad \forall (u, v) \in [0, 1]^2 \quad (4.2)$$

• برای به دست آوردن نتایج بهتر مورتین و همکاران (۲۰۰۸) استفاده از یک برآوردگر هموار هسته‌ای برای  $\Lambda$  در فرم (۲.۲) که به روش *plug-in* ساخته شده است را پیشنهاد می‌کنند.  
 فرض کنید  $\hat{F}_1$  و  $\hat{F}_2$  همان‌هایی باشند که قبلاً گفتیم آنگاه بنا به فرمانیان و همکاران (۲۰۰۴) برآوردگر هموار برای مفصل به صورت زیر می‌باشد.

$$\hat{C}(u, v) = \hat{F}(\hat{F}_1^{-1}(u), \hat{F}_2^{-1}(v)) \quad \forall (u, v) \in [0, 1]^2$$

که

$$\hat{F}_1^{-1}(u) = \inf\{x \in R; \hat{F}_1(x) \geq u\}$$

$$\hat{F}_\gamma^{-1}(v) = \inf\{y \in R; \hat{F}_\gamma(y) \geq v\}$$

بنابراین برآوردگر هموار  $\Lambda$  در فرم (۲.۲) به صورت زیر می باشد.

$$\hat{\Lambda}(u, v) = \frac{\hat{C}(u, v)}{uv} \quad \forall (u, v) \in [0, 1]^2 \quad (5.2)$$

در ادامه رفتار حدی این برآوردگرها را بیان می کنیم

شرایط نظم زیر را در نظر بگیرید.

۱-  $F$  مشتق مرتبه  $q$  کراندار دارد.

۲-  $F_1$  و  $F_\gamma$  لپ شیتس هستند.

۳-  $h \rightarrow 0$  زمانیکه  $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \sqrt{nh^q} = 0 \quad 4-$$

$$\int_R \int_R x^j y^l k(x, y) dx dy = 0 \quad 1 \leq j + l < q \quad 5-$$

$$\int_R \int_R |x|^j |y|^l |k(x, y)| dx dy < \infty \quad 1 \leq j + l < q \quad 6-$$

$$\int_R \int_R (|x| + |y|) dK(x, y) < \infty \quad 7-$$

قضیه ۵.۲. فرض کنید  $(X, Y)$  یک بردار تصادفی پیوسته باشد اگر شرایط ۱ تا ۷ برقرار باشد آنگاه

$$\hat{\Lambda}(\hat{F}_1(x), \hat{F}_\gamma(y)) \xrightarrow{P} \Lambda(F_1(x), F_\gamma(y)) \quad (x, y) \in R^2 a.s \quad n \rightarrow \infty$$

قضیه ۶.۲. فرض کنید  $(X, Y)$  یک بردار تصادفی پیوسته باشد و  $F$  توابع توزیع حاشیه ای پیوسته داشته باشد و تابع مفصل  $C(u, v)$  مشتقات جزئی پیوسته داشته باشد آنگاه

$$\Lambda_n(u, v) \xrightarrow{P} \Lambda(u, v) \quad \forall (u, v) \in [0, 1]^2$$

قضیه ۷.۲. فرض کنید  $(X, Y)$  یک بردار تصادفی پیوسته باشد. تحت شرایط ۱ تا ۷ و با فرض اینکه تابع مفصل  $C(u, v)$  دارای مشتقات جزئی پیوسته باشد آنگاه

$$\hat{\Lambda}(u, v) \xrightarrow{P} \Lambda(u, v) \quad \forall (u, v) \in [0, 1]^2$$

### مراجع

1. Bairamov, I., Kotz, S. and Kozubowski, T. J. (2003), *A new measure of linear local dependence*, Statistics, 37(3), 243-258
2. Bjerve, S. and Doksum, K. (1993), *Measures of association as functions of covariate values*, Ann. of Stat. 21(2) 890-902.
3. Fermanian, J. D., Radulovic, D. and Wegkamp, M. (2004), *Weak convergence of empirical copula processes*, Bernoulli 10(5), 847-860
4. Holland, P. W. and Wang, Y. J. (1987), *pendence Function for Continuous Bivariate Densities*, Communications in Statistics Theory and Methods, 16(3), 863-876.
5. Kolev, N. V., Gonçalves, M. and Dimitrov, B. (2007), *Probabilistic Properties of Sibuya's Dependence Function*, Department of Statistics, University of Sao Paulo
6. MORETTIN, P. A. and ABDEL LATIF, S. (2008), *Sibuya's Measure of Local Dependence*
7. Nelsen, R.B. (2006), *An Introduction to Copulas*, Springer
8. Sibuya, M. (1959), *Bivariate extreme statistics, I*, Ann. of Stat. 11 (2) (1959) 195-210.

## تحلیل و بررسی نمودار کای و نقش آن در انتخاب مفصل مناسب

پورا احمدی، م. ۱ و حیدری، ا. ۲\*

<sup>۱</sup> گروه آمار زیستی، دانشکده پیراپزشکی، کمیته تحقیقات دانشجویی، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران  
m.pourahmadi66@gmail.com

<sup>۲</sup> گروه آمار، دانشکده علوم ریاضی و آمار، دانشگاه شهید چمران اهواز  
ahmadheydari37@yahoo.com

چکیده. تابع مفصل یک ابزار قدرتمند برای بیان وابستگی کلی و موضعی بین متغیرهاست. هدف ما در این مقاله بررسی و تحلیل نموداری به نام کای می‌باشد که برخلاف نمودار همتراز، وابستگی دمی بین متغیرها را بهتر به تصویر می‌کشد. در واقع این نمودار، یک روش گرافیکی اختصاصی برای نشان دادن وابستگی دمی می‌باشد. در این مقاله نشان خواهیم داد نمودار کای در کنار آزمون‌های نیکویی برازش معیار خوبی برای انتخاب مفصل مناسب می‌باشد.

## تئوری نمودار کای (خی)

استفاده‌ای که در این جا از نمودارهای کای می‌شود فراتر از نشان دادن الگوی وابستگی بین متغیرهاست. هر مفصلی، وابستگی دمی خاص خود را دارد که به کمک آن‌ها می‌توان نوع مفصل به کار گرفته شده را تشخیص داد. قبل از معرفی این نمودار، تعریف تابع مفصل را به صورت زیر خواهیم داشت.  
قضیه ۱ (اسکلار، ۱۹۵۹): فرض کنید  $F$  تابع توزیع توأم  $n$ -بعدی با حاشیه‌های  $F_1, \dots, F_n$  باشد، در این صورت تابع مفصلی مانند  $C$  وجود دارد به طوری که برای هر  $x_1, \dots, x_n$  در  $\bar{R}$  داریم

$$F(x_1, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \quad (1)$$

نمودارهای کای، اولین بار توسط فیشر و سویترز (۱۹۸۵) بر اساس آماره کای اسکور برای نشان دادن استقلال دو متغیر به کار رفت. نمودار کای بر اساس  $\lambda$  و  $\chi$  شکل می‌گیرد که در زیر به معرفی آن‌ها می‌پردازیم. فرض کنید  $(x_1, y_1), \dots, (x_n, y_n)$  یک نمونه تصادفی برای یک جفت از متغیرهای تصادفی  $X$  و  $Y$  با توزیع توأم  $F_{XY}$  و توزیع‌های حاشیه‌ای  $F_X$  و  $F_Y$  باشد، داده‌های مشاهده شده  $(x_1, y_1), \dots, (x_n, y_n)$  را با فرمول‌های زیر به جفت‌های  $(\lambda_1, \chi_1), \dots, (\lambda_n, \chi_n)$  تبدیل می‌کنیم.

واژگان کلیدی. تابع مفصل، نمودار کای، وابستگی دمی، آزمون نیکویی برازش.  
\* سخنران

$$\chi_i = \frac{\hat{F}_{XY,i} - \hat{F}_{X,i}\hat{F}_{Y,i}}{\sqrt{\hat{F}_{X,i}(1 - \hat{F}_{X,i})\hat{F}_{Y,i}(1 - \hat{F}_{Y,i})}} \quad (۲)$$

$$\lambda_i = \mathcal{F} \operatorname{sgn}(\tilde{F}_{X,i}, \tilde{F}_{Y,i}) \cdot \max(\tilde{F}_{X,i}^{\vee}, \tilde{F}_{Y,i}^{\vee}) \quad (۳)$$

که در آن برای  $i = 1, \dots, n$ ،  $\tilde{F}_{X,i} = \hat{F}_{X,i} - 0.5$  و  $\tilde{F}_{Y,i} = \hat{F}_{Y,i} - 0.5$

فیشر و سویتزر ویژگی‌های نمودار کای را به صورت زیر خلاصه کردند:

- نمودار کای وابسته به رتبه‌ی داده‌ها می‌باشد.
- $\lambda_i$  و  $\chi_i$  در فاصله  $[-1, 1]$  مقیاس‌بندی شده است.
- اندازه  $\lambda_i$ ، فاصله‌ی یک نقطه‌ی داده  $(x_i, y_i)$  از مرکز مجموعه داده‌های دو متغیره می‌باشد.
- تحت فرض مستقل بودن بین متغیره‌های  $X$  و  $Y$  داریم:

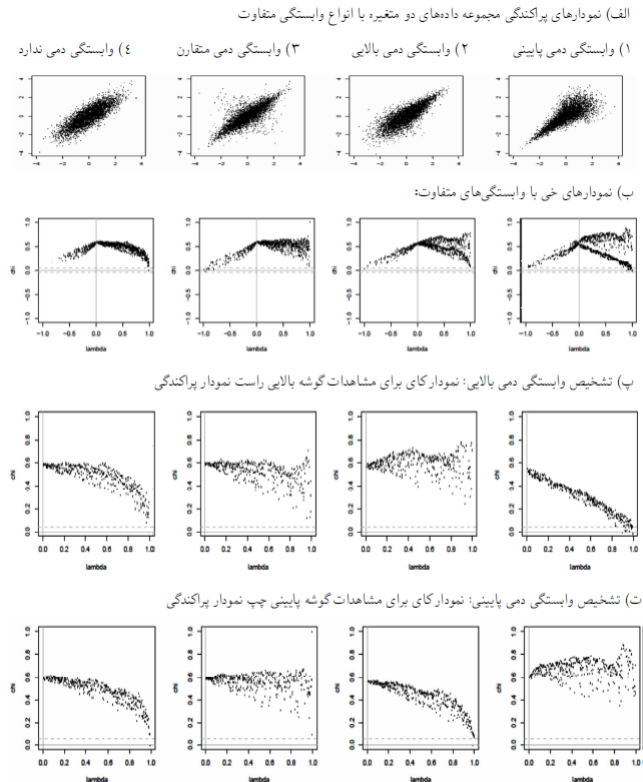
$$\lambda_i \sim U[-1, 1] \quad \text{و} \quad \chi_i \sim N\left(0, \frac{1}{n}\right)$$

- قرار گرفتن نقاط  $(\lambda_i, \chi_i)$  در باند بالای نمودار، حاکی از وابستگی مثبت بین متغیره‌ها و قرار گرفتن نقاط  $(\lambda_i, \chi_i)$  در باند پایین نمودار، وابستگی منفی بین متغیره‌ها را نشان می‌دهد.

#### انتخاب مفصل مناسب توسط نمودار کای

فیشر و سویتزر (۱۹۸۵) یک روش گرافیکی را برای ارزیابی وابستگی در مجموعه داده‌های دو متغیره مطرح کردند ولی وابستگی دمی داده‌ها را مورد بررسی قرار ندادند تا این‌که در سال ۲۰۰۴ آبرگر چگونگی حضور وابستگی دمی در مجموعه داده‌های دو متغیره را مورد بررسی قرار داد. با توجه به مفهوم وابستگی دمی برای ارزیابی حضور وابستگی دمی، ما علاقمند به مشاهدات با  $\lambda_i$ ‌های مثبت هستیم. اگر وابستگی دمی وجود نداشته باشد مقادیر  $\chi_i$ ، به خط صفر در سمت راست (برای  $\lambda$ ‌های نزدیک به ۱) نمودار کای بر می‌گردند. حضور وابستگی دمی، باعث انحراف مقادیر  $\chi$  از خط صفر برای مقادیر  $\lambda$ ‌های نزدیک به ۱ می‌باشد.

برای چک کردن وابستگی دمی، آبرگر پیشنهاد داد تا نمودار کای را فقط برای مشاهدات با  $\lambda_i$ ‌های مثبت در نظر بگیریم به طوری که مشاهدات در دو چارک جداگانه (در گوشه راست بالایی و گوشه چپ پایینی) نمودار پراکندگی دو متغیره قرار می‌گیرند که این جداسازی، قدرت ما را در تشخیص وابستگی دمی بیشتر می‌کند. برای آگاهی از اطلاعات بیشتر به اسچیرماچر (۲۰۰۸) مراجعه نمایید. شکل (۱) چگونگی وابستگی دمی در نمودار کای را نشان می‌دهد.



شکل ۱: ارزیابی وابستگی دمی و انواع آن (بالایی و پایینی) با استفاده از: الف) نمودارهای پراکندگی دو متغیره با وابستگی‌های متفاوت، ب) نمودار کای در حالت کلی، پ) نمودار کای برای تشخیص وابستگی دمی بالایی، ت) نمودار کای برای تشخیص وابستگی دمی پایینی

همان طور که ملاحظه می‌کنید مشاهدات در ستون اول از سمت چپ این نمودار، همگرا به خط افقی صفر در سمت راست هستند که نشانگر عدم وابستگی دمی بالایی و پایینی می‌باشد. در ستون ۲ بر خلاف ستون ۱، نمودار کای برای  $\lambda$  نزدیک به ۱، متمایل به خط صفر نیست که این نشان دهنده وابستگی دمی بالایی و پایینی است. در ستون ۳ و ۴، فرم گلوله ای شکل در سمت راست مشاهده می‌شود که وجود وابستگی دمی را نشان می‌دهد به طوری که برای ستون ۳، وابستگی دمی بالایی و برای ستون ۴، وابستگی دمی پایینی مشهود است.

### کاربرد: داده‌های اقتصادی

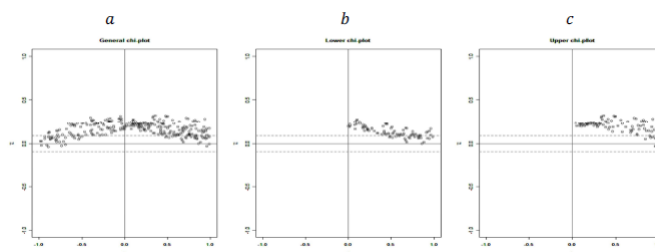
دو سری زمانی از مجموعه داده‌های فصلی اقتصاد کلان ایران (۲۰۰۷ - ۱۹۹۱) را در نظر می‌گیریم که به شرح زیر می‌باشد.

B: درآمد نفتی

C: تورم

خانواده‌های مفصل نرمال ( $N$ )،  $t$ -استیودنت ( $T$ )، کلاپتون ( $C$ )، گامیل ( $G$ )، فرانک ( $F$ )،  $BB1$ ،  $BBV$  و مستقل،  $\lambda$  مفصلی می‌باشند که ما برای این جفت داده در نظر گرفتیم.

نمودارهای کای برای این داده‌ها در شکل (۲) نشان داده شده است. از آنجایی که در دو نمودار  $b$  و  $c$ ، مقادیر  $\chi_i$  به خط صفر در سمت راست (برای  $\lambda$  های نزدیک به ۱) نمودار کای بر می‌گردند، وجود وابستگی دمی در این نمودارها مشاهده نمی‌شود، بنابراین تابع مفصل نرمال می‌تواند گزینه‌ی مناسبی باشد.



شکل ۲: ارزیابی وابستگی دمی با استفاده از:  $a$ ) نمودار کای در حالت کلی،  $b$ ) نمودار کای برای متوجه شدن وابستگی دمی پایینی،  $c$ ) نمودار کای برای تشخیص وابستگی دمی پایینی

هر چند که نمی‌توان مدل را مناسب صرفاً با روش گرافیکی انتخاب کرد ولی می‌تواند یک درک شهودی به ما القا کند و محدوده و بازه تابع‌های مفصل مورد آزمون را محدود کند که در کنار آزمون‌های نیکویی برازش، اطمینان در انتخاب را برای ما افزایش دهد. آزمون نیکویی برازش بر اساس تبدیل کندال را برای تابع مفصل‌های ذکر شده انجام می‌دهیم که در جدول (۱) خلاصه شده است.

جدول ۱:  $p$  - value های از آزمون بر اساس تبدیل کندال  $PS_n(K_n)$

$p$ - value	$N$	$T$	$C$	$G$	$F$	$BB\downarrow$	$BB\uparrow$
$PS_n(K_n)$	۰/۳۵	۰/۰۶	۰/۰۱	۰	۰	۰/۱۱	۰/۲

در جدول (۱)، مفصل نرمال، بزرگترین  $p$  - value را به خود اختصاص داده است که ادعای مطرح شده ما را بر اساس نمودار کای تأیید می‌کند.

### مراجع

1. K. Abberger, A simple graphical method to explore tail dependence in stockreturn pairs. Working paper, University of Konstanz, Germany, (2004).
2. N. I., Fisher, and P. Switzer, Chi-plots for assessing dependence. *Biometrika*, 72, 253-265, (1985).
3. N. I., Fisher, and P. Switzer, Graphical assessment of dependence: is a picture worth 100 test? *The American Statistician*, 55, 233-239, (2001).
4. Sklar, Fonctions de repartition an dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8, 229-231, (1959).



برآورد پارامتر و انتخاب مدل ارشمیدسی براساس توزیع‌های متقارن تحت نرم  $L_1$ خوشنودی، ف. <sup>۱</sup>\*

<sup>۱</sup> دانش‌آموخته کارشناسی ارشد آمار دانشگاه یزد، شاغل در معاونت فنی شرکت بیمه البرز  
aamaar90@gmail.com

چکیده. مسئله پیدا نمودن مدل احتمالی مناسب برای داده‌های چندمتغیره یکی از مسائل چالش برانگیز در آمار است. یکی از راهکارهای حل این مسئله، استفاده از مفصل است؛ به این ترتیب که اگر مفصل مناسبی برای داده‌ها پیدا کنیم، با داشتن توزیع‌های حاشیه‌ای می‌توان توزیع توأم مناسب برای داده‌ها را به دست آورد. در این مقاله به بیان روشی می‌پردازیم که در کلاس مفصل‌های ارشمیدسی، به طور هم‌زمان هم پارامترهای وابستگی مدل را برآورد می‌کند و هم آزمون نیکویی برازش انجام می‌دهد.

## ۱. مقدمه و مفاهیم مقدماتی

در این مقاله پس از بیان برخی قضایا و تعاریف مورد نیاز، براساس پژوهش‌های صورت گرفته توسط مک نیل و همکاران (۲۰۰۹) و کیو و همکاران (۲۰۱۰)، روشی بیان می‌شود که با استفاده از ویژگی توزیع‌های متقارن تحت نرم  $L_1$  به برآورد پارامتر مفصل و آزمون نیکویی برازش می‌پردازد؛ در پایان نیز نتایج نظری به دست آمده با مثالی از داده‌های واقعی مورد بررسی قرار خواهد گرفت.

قضیه ۱.۱. (قضیه اسکار برحسب توابع بقا) فرض کنید  $\bar{H}$  تابع بقای توأم  $X = (X_1, \dots, X_d)^T$  با توابع بقای حاشیه‌ای  $\bar{F}_j$ ،  $j = 1, \dots, d$  باشد. در این صورت مفصل بقای  $\hat{C}_\theta$ ، برای  $x \in \mathbb{R}^d$  در رابطه زیر صدق می‌کند:

$$\bar{H}(x) = \hat{C}_\theta(\bar{F}_1(x_1), \dots, \bar{F}_d(x_d)). \quad (1.1)$$

اگر  $\bar{F}_1, \dots, \bar{F}_d$  پیوسته باشند  $\hat{C}_\theta$  یکتاست در غیراین صورت  $\hat{C}_\theta$  به صورت یکتا بر  $\{u \in [0, 1]^d : u \in \text{range } \bar{F}_1 \times \dots \times \text{range } \bar{F}_d\}$  تعریف می‌گردد.

تعریف ۲.۱. (مفصل ارشمیدسی) گوییم مفصل  $C_\theta$  ارشمیدسی است اگر

$$C_\theta(u) = \psi_\theta(\psi_\theta^{-1}(u_1) + \dots + \psi_\theta^{-1}(u_d)), \quad u \in [0, 1]^d, \quad (2.1)$$

واژگان کلیدی. مفصل ارشمیدسی، نیکویی برازش، توزیع متقارن تحت نرم  $L_1$ ، توزیع شعاعی.

\* سخنران

که در آن  $\psi_\theta$  تابعی غیرصعودی و پیوسته با  $\psi_\theta(0) = 1$  و  $\lim_{x \rightarrow \infty} \psi_\theta(x) = 0$  که تابع مولد ارشمیدسی نامیده می‌شود و دارای معکوس  $\psi_\theta^{-1}$  است. جهت مطالعه ویژگی‌های مفصل‌های ارشمیدسی و نیز خانواده‌های شناخته شده و پرکاربرد آن می‌توان به نلسن (۲۰۰۶) مراجعه نمود.

**تعریف ۳.۱.** گوئیم بردار تصادفی  $X$  بر  $\mathbb{R}_+^d = [0, \infty)^d$  دارای یک توزیع متقارن تحت نرم  $L_1$  است، اگر

$$X \stackrel{d}{=} RS_d,$$

که در آن متغیر تصادفی نامنفی  $R$  مستقل از بردار تصادفی  $S_d$  است و  $S_d$  دارای توزیع یکنواخت بر

$$S_d = \{x \in \mathbb{R}_+^d; \|x\|_1 = 1\},$$

است. همچنین متغیر تصادفی  $R$  قسمت شعاعی  $X$  گفته می‌شود و توزیع آن نیز توزیع شعاعی نامیده می‌شود.

**قضیه ۴.۱.** فرض کنید  $U$  دارای مفصل بقای ارشمیدسی با مولد  $\psi_\theta$  باشد، در این صورت  $(\psi_\theta^{-1}(U_1), \dots, \psi_\theta^{-1}(U_d))$  دارای توزیع متقارن تحت نرم  $L_1$ ، با مفصل بقای  $\hat{C}_\theta$  و توزیع شعاعی  $F_R$  است و برای  $x \in [0, \infty)$  داریم:

$$F_R(x) = 1 - \sum_{k=0}^{d-2} \frac{(-1)^k x^k \psi_\theta^{(k)}(x)}{k!} - \frac{(-1)^{d-1} x^{d-1} \psi_\theta^{(d-1)}(x)}{(d-1)!},$$

که در آن  $a_+$  ماکسیمم  $a$  و  $0$  است.

**قضیه ۵.۱.** فرض کنید  $C_\theta$  مفصل ارشمیدسی  $d$ -بعدی با مولد  $\psi_\theta$  باشد و  $U$  نیز متغیر تصادفی با توزیع  $C_\theta$  باشد، در این صورت  $\sum_{i=1}^d \psi_\theta^{-1}(U_i)$  و  $(\frac{\psi_\theta^{-1}(U_1)}{\sum_{i=1}^d \psi_\theta^{-1}(U_i)}, \dots, \frac{\psi_\theta^{-1}(U_d)}{\sum_{i=1}^d \psi_\theta^{-1}(U_i)})$  مستقلند و برای

$j = 1, \dots, d$  متغیر تصادفی

$$V_\theta^j(U_1, \dots, U_d) = \left( 1 - \frac{\psi_\theta^{-1}(U_j)}{\sum_{i=1}^d \psi_\theta^{-1}(U_i)} \right)^{d-1},$$

دارای توزیع یکنواخت بر  $[0, 1]$  است.

در واقع قضیه‌ی ۵.۱ تجزیه‌ی بردار تصادفی  $U$  با مفصل ارشمیدسی  $C_\theta$  به دو بخش مستقل است که یک بخش دارای توزیع شعاعی  $F_R$  و بخش دیگر دارای توزیع یکنواخت است.

## ۲. برآورد و انتخاب مدل با معیار اطلاع شعاعی (RIC)

نمونه تصادفی مستقل و هم‌توزیع  $(X_{1i}, X_{2i}), i = 1, \dots, n$  از توزیعی توأم با توزیع‌های حاشیه‌ای  $F_1$  و  $F_2$  را در نظر بگیرید. فرض کنید مدلهایی که برای برازش بر داده‌ها در نظر گرفته شده‌اند، مدل‌های

با توابع مولد  $M_1, \dots, M_K$  با  $\psi_\theta^1, \dots, \psi_\theta^K$  باشند. فرض کنید

$$V_\theta(U_1, U_2) = 1 - \frac{\psi_\theta^{-1}(U_2)}{\sum_{i=1}^r \psi_\theta^{-1}(U_i)} = \frac{\psi_\theta^{-1}(U_1)}{\psi_\theta^{-1}(U_1) + \psi_\theta^{-1}(U_2)},$$

دارای توزیع یکنواخت بر  $[0, 1]$  است. همچنین با فرض آنکه  $V_\theta(t)$  و  $V_\theta^n(t)$  به ترتیب توزیع تجربی متغیر تصادفی  $V_\theta(U_1, U_2)$  و تابع توزیع آن باشند، برآوردگر  $\hat{\theta}_n$  از طریق مینیم کردن فاصله کرامر-فون میزس به صورت

$$d(\theta) = n \int_0^1 (V_\theta^n(t) - V_\theta(t))^2 dV_\theta(t) = n \int_0^1 (V_\theta^n(t) - t)^2 dt, \quad (1.2)$$

به دست می آید. براساس مطالب گفته شده، می توان مسئله انتخاب مدل ارشمیدسی چندمتغیره را به مسئله ای یک متغیره که دارای چهارگام زیر است، تبدیل کرد: (۱) با استفاده از مینیم کردن فاصله کرامر-فون میزس یعنی (۱.۲)، پارامتر مدل های مورد نظر را برآورد می کنیم و آن ها را با  $\hat{\theta}_n^1, \dots, \hat{\theta}_n^K$  نمایش می دهیم. (۲) برای  $k = 1, \dots, K$ ، توزیع شعاعی برآوردی  $F_{\hat{\theta}_n^k}^k$  با مولد  $\psi_{\hat{\theta}_n^k}^k$  برای بخش شعاعی مشاهده شده  $x_i = \psi_{\hat{\theta}_n^k}^{k-1}(u_{1i}) + \psi_{\hat{\theta}_n^k}^{k-1}(u_{2i})$  ( $i = 1, \dots, n$ ) را طبق (۳.۱) به دست می آوریم. (۳) برای  $k = 1, \dots, K$  توزیع شعاعی تجربی  $\hat{F}_{\hat{\theta}_n^k}^k$  با مولد  $\psi_{\hat{\theta}_n^k}^k$  را از رابطه ی

$$\hat{F}_{\hat{\theta}_n^k}^k(x) = \frac{1}{n} \sum_{i=1}^n I_{\{(u_{1i}, u_{2i}) | \psi_{\hat{\theta}_n^k}^{-1k}(u_{1i}) + \psi_{\hat{\theta}_n^k}^{-1k}(u_{2i}) \leq x\}},$$

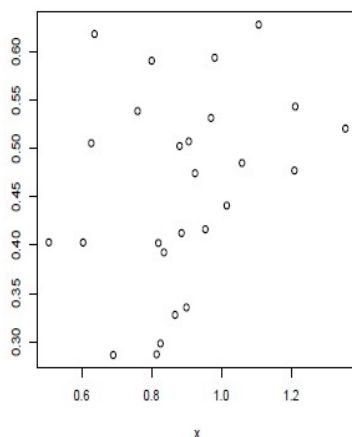
را به دست می آوریم. (۴) تعیین می کنیم که کدام توزیع شعاعی مشاهده شده  $\hat{F}_{\hat{\theta}_n^k}^k(x)$  نزدیکترین فاصله را به  $F_{\hat{\theta}_n^k}^k(x)$  دارد، یعنی در کدام مدل فاصله ی

$$d_{M_k}(\hat{\theta}_n^k) = \sum_{i=1}^n (\hat{F}_{\hat{\theta}_n^k}^k(x_i) - F_{\hat{\theta}_n^k}^k(x_i))^2.$$

مینیم می شود. به این معیار که جهت انتخاب و برآورد مدل ارشمیدسی کاربرد دارد، معیار اطلاع شعاعی گفته می شود.

### ۳. مثالی کاربردی از صنعت بیمه کشور

در این قسمت از داده های سری زمانی بیمه های اتومبیل کشور از سال ۱۳۶۶ تا ۱۳۹۱ به منظور یافتن ساختار توأم نسبت خسارت بیمه شخص ثالث و بیمه بدنه اتومبیل بهره جستیم؛ به عبارت دیگر بر اساس این داده ها و با استفاده از ۴ گام گفته شده در بخش ۲، در میان چهار خانواده مفصل ارشمیدسی علی-میخائیل-حق، گامبل، فرانک و جو به دنبال مناسب ترین مفصل هستیم. به همین منظور نسبت خسارت شخص ثالث با  $X$  و بدنه با  $Z$  نشان داده شده است. با مشاهده جدول ۱ می توان نتیجه گرفت که بهترین مفصل برای برآزش بر این داده ها، گامبل بوده که در این مدل  $\hat{\theta}_n = 1/19$ .



شکل ۱: نمودار پراکنش داده‌های  $X$  در برابر داده‌های  $Z$

جدول ۱: برآورد و نیکویی برازش مفصل برای متغیرهای تصادفی  $X$  و  $Z$  با RIC

خانواده‌های مفصل‌های ارشمیدسی	$\hat{\theta}_n$	$d(\hat{\theta}_n)$
گامبل	۱/۱۹	۰/۰۴۰
فرانک	۱/۱۸	۰/۰۴۱
جو	۱/۳۷	۰/۰۴۳
علی-میخائیل-حق	۰/۴۲	۰/۰۴۴

### مراجع

1. McNeil, A.J. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and L1-norm symmetric distribution. *Annals of Statistics*. 37 (5B), 3059-3097.
2. Nelsen, R. B. (2006). *An Introduction to Copulas*, Second Edition. Springer, New York.
3. Qu, X., Zhou, J. and Shen, X. (2010). Archimedean copula estimation and model selection via L1-norm symmetric distribution. *Insurance: Mathematics and Economics*. 46, 406-411.



## کاربردهایی از ترتیب تصادفی همبستگی

محتشمی، و. امینی، م. احمدی، ج.

گروه آمار، دانشکده علوم ریاضی و آمار، دانشگاه بیرجند

v.mohtashami@um.ac.ir

چکیده. در این مقاله، ارتباط ترتیب تصادفی همبستگی با سایر ترتیب‌های تصادفی بیان و شرایط لازم برای اینکه سود تنوع بخشی اکیدا مثبت باشد، بررسی شده است. همچنین با فرض برقراری ترتیب تصادفی همبستگی، برای تبدیل قطع زیان مجموع اجزای یک بردار تصادفی، کران بالای آن را می‌یابیم.

### ۱. مقدمه

ترتیب تصادفی همبستگی را اولین بار داهنه و گوورت (۱۹۹۶) برای حالت دو متغیره تعریف کردند. ترتیب تصادفی همبستگی در حالت دو متغیره ترتیبی جزئی بین توزیع‌های دو متغیره با حاشیه‌ای‌های یکسان می‌باشد و ابزاری مناسب برای استنتاج در مورد ریسک پذیری سبب‌ارایی با ریسک‌های دو به دو وابسته است.

لو و یی (۲۰۰۴)، یی و لو (۲۰۰۴)، یی و ونگ (۲۰۰۶) ارتباط بین ترتیب تصادفی همبستگی و سایر ترتیب‌های تصادفی را بررسی کردند. داهنه و همکاران (۲۰۰۹) کاربردهایی از ترتیب تصادفی همبستگی در مسائل بیمه مالی را بیان کردند.

در بخش اول ارتباط بین ترتیب تصادفی همبستگی و سایر ترتیب‌های تصادفی بررسی می‌شود. ارتباط سود تنوع بخشی و ترتیب تصادفی همبستگی در بخش دوم بیان می‌شود. در بخش سوم برای تبدیل قطع زیان مجموع اجزای بردار تصادفی کران بالا بدست می‌آید.

در این جا  $\underline{X} = (X_1, \dots, X_n)$  بردار تصادفی  $n$ -بعدی متعلق به رده فرشه است.

**تعریف ۱.۱.** تابع  $C : [0, 1]^n \rightarrow [0, 1]$  را یک مفصل  $n$ -متغیره گویند هرگاه:

الف) برای هر  $i = 1, \dots, n$ ،  $u_i \in (0, 1)$ ،

$$C(0, u_2, \dots, u_n) = \dots = C(u_1, \dots, u_{i-1}, 0, u_{i+1}, \dots, u_n)$$

$$= \dots = C(u_1, \dots, u_{n-1}, 0) = 0,$$

$$C(1, \dots, 1, u_i, 1, \dots, 1) = u_i \quad \forall 1 \leq i \leq n$$

2010 Mathematics Subject Classification. Primary 47A55; Secondary 39B52, 34K20, 39B82.

واژگان کلیدی. ترتیب تصادفی همبستگی، ترتیب تصادفی قطع زیان، تابع مفصل.

(ب)  $C$  تابعی  $n$ -صعودی است. یعنی برای هر  $a_i, b_i \in [0, 1]$  که  $a_i \leq b_i$ ،  

$$V_H(B) \geq 0,$$

که در آن

$$V_H(B) = \sum sgn(c)H(c),$$

مجموع بر روی تمام رئوس  $c$  در  $B$  بوده و تابع  $sgn(c)$  بصورت زیر تعریف می‌شود:

$$sgn(c) = \begin{cases} 1; & \text{اگر } c_k = a_k \text{ برای عدد زوج } k \\ -1; & \text{اگر } c_k = a_k \text{ برای عدد فرد } k. \end{cases}$$

قضیه زیر تابع توزیع  $n$ -متغیره را در چارچوب تابع مفصل بیان می‌کند.

**قضیه ۲.۱.** (قضیه اسکالر) فرض کنید  $H$  تابع توزیع  $n$ -متغیره با توابع توزیع حاشیه ای  $F_1, \dots, F_n$  باشد. در این صورت مفصل  $C$  موجود است به قسمی که

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)), \forall x_i \in \mathbb{R}, i = 1, \dots, n. \quad (1.1)$$

اگر  $F_i$  ها پیوسته باشند آنگاه  $C$  یکتاست در غیر این صورت با فرض اینکه دامنه توابع توزیع  $F_i$  را با  $RanF_i$  نشان دهیم  $C$  بطور منحصر بفرد بر  $RanF_1 \times \dots \times RanF_n$  مشخص می‌شود.

**نکته ۳.۱.**  $E(X-t)_+$  را تبدیل قطع زیان متغیر تصادفی  $X$  گویند.

**تعریف ۴.۱.** فرض کنید  $X$  و  $Y$  متغیرهای تصادفی با میانگین منتهای باشند. در این صورت:  
 الف) متغیر تصادفی  $X$  از  $Y$  در ترتیب تصادفی محدب ( $X \leq_{cx} Y$ ) کوچکتر است هرگاه

$$EX = EY, \quad \forall t \geq 0 \quad E(X-t)_+ \leq E(Y-t)_+.$$

ب) متغیر تصادفی  $X$  از  $Y$  در ترتیب تصادفی قطع زیان ( $X \leq_{sl} Y$ ) کوچکتر است هرگاه

$$\forall t \geq 0 \quad E(X-t)_+ \leq E(Y-t)_+.$$

**تعریف ۵.۱.** فرض کنید  $\underline{X} = (X_1, \dots, X_n)$  و  $\underline{Y} = (Y_1, \dots, Y_n)$  بردارهای تصادفی به ترتیب با توابع توزیع  $F$  و  $G$  باشند به قسمی که  $F, G \in R(F_1, \dots, F_n)$

الف)  $\underline{X}$  از  $\underline{Y}$  در ترتیب تصادفی همبستگی ( $\underline{X} \leq_{corr} \underline{Y}$ ) کوچکتر است هرگاه برای هر دو مجموعه جدا از هم  $A_1$  و  $A_2$  که  $A_1, A_2 \subseteq \{1, \dots, n\}$  و به ازای توابع غیرنزولی  $\alpha(\cdot)$  و  $\beta(\cdot)$

$$(2.1) \quad Cov(\alpha(X_i, i \in A_1), \beta(X_j, j \in A_2)) \leq Cov(\alpha(Y_i, i \in A_1), \beta(Y_j, j \in A_2)).$$

ب)  $\underline{X}$  را از  $\underline{Y}$  در ترتیب تصادفی زیر جمعی ( $\underline{X} \leq_{sm} \underline{Y}$ ) کوچکتر گویند هرگاه به ازای تمام توابع زیر جمعی  $g(\cdot)$ ،  $Eg(\underline{X}) \leq Eg(\underline{Y})$ ، به شرط اینکه امید ریاضی موجود باشد.

ج)  $\underline{X}$  را از  $\underline{Y}$  در ترتیب تصادفی چندگانه بالایی ( $\underline{X} \leq_{uo} \underline{Y}$ ) کوچکتر گویند هرگاه

$$\bar{F}_{\underline{X}}(x) \leq \bar{F}_{\underline{Y}}(x); \quad \forall x \in \mathbb{R}^n.$$

د)  $\underline{X}$  را از  $\underline{Y}$  در ترتیب تصادفی چندگانه پایینی ( $\underline{X} \leq_{lo} \underline{Y}$ ) کوچکتر گویند هرگاه

$$F_{\underline{X}}(x) \geq F_{\underline{Y}}(x); \quad \forall x \in \mathbb{R}^n.$$

**نکته ۶.۱.** در حالت  $n = 2$  ترتیب‌های تصادفی فوق با هم معادلند.

قضیه زیر ارتباط بین ترتیب تصادفی همبستگی ترتیب‌های تصادفی وابسته و ترتیب‌های تصادفی محذب و قطع زیان را بیان می‌کند.

قضیه ۷.۱. فرض کنید  $\underline{X}$  و  $\underline{Y}$  دو بردار تصادفی با توابع توزیع توأم متعلق به رده فرشه،  $C_F$  و  $C_G$  به ترتیب توابع مفصل متناظر با  $F$  و  $G$  باشند، اگر  $\underline{X} \leq_{corr} \underline{Y}$  آنگاه

- الف)  $\underline{X} \leq_{uo} \underline{Y}$  ،
- ب)  $\underline{Y} \leq_{lo} \underline{X}$  ،
- ج)  $C_F(u_1, \dots, u_n) > C_G(u_1, \dots, u_n)$  ،
- د)  $\underline{X} \leq_{sm} \underline{Y}$  ،
- ه)  $\sum_{i=1}^n X_i \leq_{sl} \sum_{i=1}^n Y_i$  ،
- و)  $\sum_{i=1}^n X_i \leq_{cx} \sum_{i=1}^n Y_i$  .

برای برهان قضیه فوق به لو و یی (۲۰۰۴)، یی و لو (۲۰۰۴)، یی و ونگ (۲۰۰۶) مراجعه شود.

تعریف ۸.۱. متغیرهای تصادفی  $X_1, \dots, X_n$  را توأما هم‌یکنوا گویند هرگاه یکی از شرایط زیر برقرار باشد

الف)

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \min[F_{X_1}(x_1), \dots, F_{X_n}(x_n)], \quad \forall x_1, \dots, x_n \geq 0,$$

ب) برای توابع حقیقی غیر نزولی  $u_i(\cdot), i = 1, \dots, n$ ، متغیر تصادفی  $Z$  وجود داشته باشد به قسمی که

$$(X_1, \dots, X_n) \stackrel{d}{=} (u_1(Z), \dots, u_n(Z)).$$

ج) برای متغیر تصادفی  $U$  که بطور یکنواخت بر بازه  $[0, 1]$  توزیع شده است

$$(X_1, \dots, X_n) \stackrel{d}{=} (F_{X_1}^{-1}(U), \dots, F_{X_n}^{-1}(U)).$$

نکته ۹.۱. در ادامه بردار تصادفی هم‌یکنوا را با  $(X_1^C, \dots, X_n^C)$  نشان می‌دهیم.

یادآوری می‌شود با فرض اینکه  $0 < p < 1$  و  $X$ ،  $g(X)$  متغیرهای تصادفی حقیقی مقدار باشند الف) اگر  $g(\cdot)$  تابعی غیر نزولی و از چپ پیوسته باشد آنگاه:

$$F_{g(X)}^{-1}(p) = g(F_X^{-1}(P)),$$

ب) اگر  $g(\cdot)$  تابعی غیر صعودی و از چپ پیوسته باشد آنگاه:

$$F_{g(X)}^{-1+}(p) = g(F_X^{-1}(1 - P)),$$

قضیه ۱۰.۱. بردار تصادفی  $(X_1, \dots, X_n)$  و بردار هم‌یکنوای متناظرش  $(X_1^C, \dots, X_n^C)$  را در نظر بگیرید. فرض کنید  $S^C = X_1^C + X_2^C + \dots + X_n^C$  باشد. آنگاه قاعده حق بیمه  $S^C$  از رابطه زیر بدست می‌آید

$$E[(S^C - d)_+] = \sum_{i=1}^n E[(X_i - d)_+], \quad (F_{S^C}^{-1+}(0) < d < F_{S^C}^{-1}(1)).$$

که در آن برای  $i = 1, \dots, n$  و  $d_i = F_{X_i}^{-1(\alpha_d)}(F_{S^C}(d))$  و  $\alpha_d \in [0, 1]$  از حل معادله  $F_{S^C}^{-1(\alpha_d)}(F_{S^C}(d)) = d$  بدست می‌آید.

نکته ۱۱.۱. اگر  $X_i$  متغیرهای تصادفی نامنفی و  $F_{X_i}$  توابعی اکیدا صعودی باشند آنگاه:

$$E[(S^C - d)_+] = \sum_{i=1}^n E[(X_i - F_{X_i}^{-1}(F_{S^C}(d)))_+], \quad (F_{S^C}^{-1+}(\circ) < d < F_{S^C}^{-1}(1)).$$

۲. سودهای تنوع بخشی اکیدا مثبت

فرض کنید  $(X_1, X_2)$  بردار تصادفی با توابع توزیع حاشیه‌ای  $F_1, F_2$ ،  $F_1$  و  $F_2$  یک متغیر تصادفی یکنواخت بر بازه  $(\circ, 1)$  باشد در این صورت:

$$(F_1^{-1}(U), F_2^{-1}(1-U)) \leq_{corr} (X_1, X_2) \leq_{corr} (F_1^{-1}(U), F_2^{-1}(U)), \quad (1.2)$$

نامساوی‌های شماره (۱.۲) نشان می‌دهند که بردار هم یکنواختی  $(F_1^{-1}(U), F_2^{-1}(U))$  در مفهوم ترتیب تصادفی همبستگی بزرگ‌ترین است. در حالیکه بردار یکنواخت معکوس  $(F_1^{-1}(U), F_2^{-1}(1-U))$  در مفهوم ترتیب تصادفی همبستگی کوچک‌ترین است.

تعریف ۱.۲. بردار تصادفی زیان  $\underline{X} = (X_1, \dots, X_n)$  و سرمایه متناظر  $\underline{c} = (c_1, \dots, c_n)$  را در نظر بگیرید. در این صورت سود تنوع بخشی  $D_{\underline{c}}[\underline{X}]$  به صورت زیر تعریف می‌شود

$$D_{\underline{c}}[\underline{X}] = \sum_{i=1}^n (X_i - c_i)_+ - \left( \sum_{i=1}^n (X_i - c_i) \right)_+. \quad (2.2)$$

قضیه ۲.۲. فرض کنید  $(X_1, X_2)$  یک بردار تصادفی با توزیع‌های حاشیه‌ای  $F_1(\cdot), F_2(\cdot)$  و  $U \sim U(\circ, 1)$  آنگاه

$$\circ \leq E[D_{\underline{c}}[F_1^{-1}(U), F_2^{-1}(U)]] \leq E[D_{\underline{c}}[\underline{X}]] \leq E[D_{\underline{c}}[F_1^{-1}(U), F_2^{-1}(1-U)]].$$

برای برهان قضیه فوق به داهنه و همکاران (۲۰۰۹) مراجعه شود. قضیه فوق بیان می‌کند که سود تنوع بخشی مورد انتظار برای هر بردار تصادفی از زیان‌ها که حاشیه‌ای‌های آن در دسترس است، دارای کمترین مقدار است، هرگاه مفصلی که بین توزیع‌های حاشیه‌ای ارتباط برقرار می‌کند، مفصل هم یکنوا باشد.

قضیه ۳.۲. فرض کنید  $(X_1, X_2)$  بردار تصادفی با توزیع حاشیه‌ای پیوسته  $F_i$  بر  $(F_i^{-1+}(\circ), F_i^{-1}(1))$  اکیدا صعودی باشند و

$$F^{-1+}(p) = \sup\{x \in \mathbb{R} | F_X(x) \leq p\}$$

$$F^{-1}(p) = \inf\{x \in \mathbb{R} | F_X(x) \geq p\}, \quad p \in [\circ, 1]$$

علاوه بر این قرار دهید  $\underline{c} = (c_1, c_2)$  به طوری که برای  $i = 1, 2$ ,

$$F_i^{-1+}(\circ) < c_i < F_i^{-1}(1),$$

توجه داریم که در این حالت هیچ  $p \in (\circ, 1)$  وجود ندارد به طوری که  $\underline{c} = (F_1^{-1}(p), F_2^{-1}(p))$  باشد. آنگاه

$$E[D_{\underline{c}}[\underline{X}]] > \circ.$$



## ۳. کران تبدیل قطع زیان مجموع اجزای یک بردار تصادفی

در این بخش در قالب دو مثال برای تبدیل قطع زیان مجموع اجزای یک بردار تصادفی کران ارائه می‌شود. برای جزئیات بیشتر به داهنه و همکاران (۲۰۰۲) مراجعه شود

مثال ۱.۳. بردار تصادفی  $(X_1, \dots, X_n)$  با توابع توزیع حاشیه‌ای نمایی را در نظر بگیرید. اگر

$$(X_1, \dots, X_n) \leq_{corr} (F_{X_1}^{-1}(U), \dots, F_{X_n}^{-1}(U))$$

برای تبدیل قطع زیان  $S = X_1 + \dots + X_n$  یک کران بالا بدست می‌آوریم.

قرار دهید  $S^C = F_{X_1}^{-1}(U) + \dots + F_{X_n}^{-1}(U)$  از آن جائیکه

$$(X_1, \dots, X_n) \leq_{corr} (F_{X_1}^{-1}(U), \dots, F_{X_n}^{-1}(U)) \Rightarrow S \leq_{sl} S^C, \quad (1.3)$$

کافیست تبدیل قطع زیان  $S^C$  را بیابیم. با توجه به اینکه به این که  $X_i$  ها دارای توزیع نمایی هستند

$$F_{X_i}(x) = 1 - \exp\left(-\frac{x}{\beta_i}\right), \quad \beta_i > 0, x_i \geq 0,$$

و

$$F_{X_i}^{-1}(p) = -\beta_i \log(1 - p),$$

بنابراین تبدیل قطع زیان متغیر تصادفی  $X_i$  به صورت زیر است:

$$E[(X_i - d)_+] = \beta_i \exp\left(-\frac{d}{\beta_i}\right), \quad 0 < d < +\infty,$$

و معکوس تابع توزیع  $S^C$  عبارتست از:

$$F_{S^C}^{-1}(p) = -\left(\sum_{i=1}^n \beta_i\right) \log(1 - p), \quad 0 < p < 1.$$

در نتیجه

$$E[(S^C - d)_+] = \beta \exp\left(-\frac{d}{\beta}\right), \quad \beta \sum_{i=1}^n \beta_i, \quad 0 < d < +\infty. \quad (2.3)$$

بنا به رابطه (۱.۳)، (۲.۳) کران تبدیل قطع زیان  $S$  است.

مثال ۲.۳. بردار تصادفی  $(\alpha_1 X_1, \dots, \alpha_n X_n)$  را در نظر بگیرید بطوریکه  $X_i$  ها از توزیع لگ

نرمال  $(\log X_i \sim N(\mu_i, \sigma_i^2))$  پیروی کند و  $\alpha_i \in \mathbb{R} - \{0\}$  علاوه بر این فرض کنید  $S =$

$\alpha_1 X_1 + \dots, \alpha_n X_n$  باشد برای تبدیل قطع زیان  $S$  یک کران بالا بدست می‌آوریم.

قرار دهید  $S^C = F_{\alpha_1 X_1}^{-1}(U) + \dots + F_{\alpha_n X_n}^{-1}(U)$  از آن جائیکه

$$(\alpha_1 X_1, \dots, \alpha_n X_n) \leq_{corr} (F_{\alpha_1 X_1}^{-1}(U), \dots, F_{\alpha_n X_n}^{-1}(U)) \Rightarrow S \leq_{sl} S^C. \quad (3.3)$$

بنابراین کافیست تبدیل قطع زیان  $S^C$  را بیابیم. با توجه به اینکه  $\Phi^{-1}(1 - p) = -\phi^{-1}(p)$  داریم

$$F_{\alpha_i X_i}^{-1}(p) = \alpha_i \exp\{\mu_i + \text{sgn}(\alpha_i) \sigma_i \Phi^{-1}(p)\}, \quad 0 < p < 1,$$

که در آن

$$\text{sgn}(\alpha_i) = \begin{cases} 1 & , \alpha_i > 0 \\ 0 & , \alpha_i < 0 \end{cases}$$

از طرفی تبدیل قطع زیان متغیر تصادفی لگ نرمال به صورت زیر است:

$$E[(X_i - d_i)_+] = \exp\{\mu_i + \frac{\sigma_i^2}{\gamma}\} \Phi(d_{i,1}) - d_i \Phi(d_{i,2}),$$

علاوه بر این

$$E[(d_i - X_i)_+] = -\exp\{\mu_i + \frac{\sigma_i^2}{\gamma}\} \Phi(-d_{i,1}) + d_i \Phi(-d_{i,2}), \quad (۴.۳)$$

که در آن  $d_i > 0$  و

$$d_{i,1} = \frac{\mu_i + \sigma_i^2 - \log(d_i)}{\sigma_i}, \quad d_{i,2} = d_{i,1} - \sigma_i, \quad (۵.۳)$$

از آن جاییکه

$$E[\alpha_i(X_i - d_i)_+] = -\alpha_i E[(d_i - X_i)_+],$$

اگر  $\alpha_i < 0$  بنا به روابط (۴.۳) و (۵.۳) داریم:

(۶.۳)

$$E[\alpha_i(X_i - d_i)_+] = \alpha_i \exp\{\mu_i + \frac{\sigma_i^2}{\gamma}\} \Phi(\text{sgn}(\alpha_i)d_{i,1}) - \alpha_i d_i \Phi(\text{sgn}(\alpha_i)d_{i,2}), \quad d_i > 0,$$

با توجه به اینکه توابع توزیع حاشیه‌ای اکیدا صعودی و پیوسته هستند. برای هر  $x < F_{SC}^{-1+}(0)$

$F_{SC}^{-1}(1)$ ،  $F_{SC}(x)$  بطور یکتا از رابطه  $F_{SC}^{-1}(F_{SC}(x)) = x$  یا بطور معادل از رابطه زیر بدست

می‌آید:

(۷.۳)

$$\sum_{i=1}^n \alpha_i \exp\{\mu_i + \text{sgn}(\alpha_i)\sigma_i \Phi^{-1}(F_{SC}(x))\} = x, \quad F_{SC}^{-1+}(0) < x < F_{SC}^{-1}(1),$$

با استناد به نکته ۱۱.۱ برای  $F_{SC}^{-1+}(0) < d < F_{SC}^{-1}(1)$  حق بیمه قطع زیان  $S^C$  بصورت زیر است:

$$E[(S^C - d)_+] = \sum_{i=1}^n E[(\alpha_i X_i - F_{\alpha_i X_i}^{-1}(F_{SC}(d)))_+], \quad (۸.۳)$$

بنا به روابط (۶.۳) و (۷.۳) و (۸.۳) برای  $F_{SC}^{-1+}(0) < d < F_{SC}^{-1}(1)$  داریم

(۹.۳)

$$E[(S^C - d)_+] = \sum_{i=1}^n \alpha_i \exp\{\mu_i + \frac{\sigma_i^2}{\gamma}\} \Phi(\text{sgn}(\alpha_i)\sigma_i - \Phi^{-1}(F_{SC}(d))) - d(1 - F_{SC}(d)),$$

در نتیجه بنا به (۳.۳) رابطه (۹.۳) کران بالا برای تبدیل قطع زیان  $S$  است.

#### ۴. نتیجه‌گیری

ترتیب تصادفی همبستگی بین دو بردار تصادفی، ترتیب تصادفی زیرجمعی و چندگانه بالایی را نتیجه می‌دهد و رابطه ترتیب را برای ترتیب تصادفی چندگانه پایینی و توابع مفصل متناظر عکس کرده و مجموع اجزای آن‌ها بوسیله ترتیب تصادفی محدب و قطع زیان مرتب می‌شود. علاوه بر این اگر ترتیب تصادفی همبستگی برقرار باشد با در نظر گرفتن برخی شرایط، سود تنوع بخشی اکیدا مثبت است. همچنین با

فرض برقراری ترتیب تصادفی همبستگی برای یک بردار تصادفی و بردار هم‌یکنوای متناظرش می‌توان برای مجموع اجزای آن بردار تصادفی کران بالا یافت.

### مراجع

1. Dhaene, J., Denuit, M., Goovaerts, M.J., Kass, R., Vyncke, D. (2002), *The concept of comonotonicity in actuarial science and finance: theory*, Insurance: Mathematics and Economics 31, 3-33.
2. Dhaene, J., Denuit, M., Vanduffel, S. (2009), *Correlation order, merging and diversification*, Insurance: Mathematics and Economics 45, 325-332.
3. Dhaene, J., Goovaerts, M.J. (1996), *Dependency of risks and stop-loss order*, ASTIN Bulletin, 26, 201-212.
4. Lu, T.Y., Zhang, Y. (2004), *Generalized correlation order and stop-loss order*, Insurance: Mathematics and Economics 35, 69-76.
5. Nelson, R. (2006), *An Introduction to Copulas*, Springer Science+Business Media, Inc.
6. Yi, Z., Weng, C. (2006), *On the Correlation order*, Statistics and Probability Letters, 76, 1410-1416.
7. Zhang, Y., Lu, T.Y. (2004), *A further study on correlation order*, Applied Mathematics of Chinese University, 19, 429-434.



## تشخیص وابستگی با استفاده از نمودارها

شاکر هدایت آباد، م.<sup>۱</sup>، اصفهانی، م.<sup>۲</sup>، و امینی، م.<sup>۳</sup>

<sup>۱،۲</sup> گروه آمار، دانشکده علوم ریاضی، دانشگاه فردوسی مشهد  
maryamshaker65@yahoo.com  
m-amini@um.ac.ir  
<sup>۲</sup> گروه آمار، دانشگاه ولایت  
esfahani64@yahoo.com

چکیده. در این مقاله پس از معرفی تابع مفصل و توزیع کندال، ارتباط آنها با یکدیگر را مورد بررسی قرار می‌دهیم. سپس نمودار کای و ویژگی‌های آن را مورد مطالعه قرار می‌دهیم. در ادامه نحوه رسم نمودار کندال را بیان کرده و ویژگی‌های آن را مورد بررسی قرار می‌دهیم. در پایان با ارائه یک مثال از داده‌های واقعی نحوه تشخیص وابستگی با استفاده از نمودارهای کای، کندال و پراکنندگی را بیان می‌کنیم.

### ۱. مقدمه

فیشر و اسویتزر (۱۹۸۵ و ۲۰۰۱) نمودار کای را معرفی کرده و برخی ویژگی‌های آن را بیان کردند. نمودار کای، به داده‌ها تنها از طریق مقادیر رتبه‌های آن‌ها بستگی دارد. جنست و بویز (۲۰۰۳) نمودار کندال را معرفی کردند. این نمودار یک ابزار بصری مهم برای برآورد وابستگی در نمونه است و مقادیر انحنای یعنی درجه وابستگی داده‌ها را نشان می‌دهد. نمودار کندال بر اساس مفهوم نمودار  $Q - Q$  می‌باشد با این تفاوت که با مفصل‌ها نیز در ارتباط است. یکی از ویژگی‌های مهم نمودار کندال تفسیر آسان تر آن نسبت به نمودار کای است و اینکه برای حالت‌های چند متغیره نیز قابل توسعه می‌باشد.

### ۲. تابع توزیع کندال

در این بخش پس از معرفی مفصل، تابع توزیع کندال و رابطه آن را با توی کندال بیان می‌کنیم.

تعریف ۱.۲. تابع  $I = [0, 1] : I^2 \rightarrow C$  یک مفصل است اگر در شرایط زیر صدق کند:

برای هر  $u, v \in I$ :

الف)  $C(u, 0) = C(0, v) = 0$ .

ب)  $C(u, 1) = u, C(1, v) = v$ .

ج) برای هر  $u_1, u_2, v_1, v_2 \in I$  طوری که  $u_1 \leq u_2$  و  $v_1 \leq v_2$  داریم:

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

واژگان کلیدی. مفصل، وابستگی، نمودار کای، نمودار کندال.  
\* سخنران

اهمیت مفصل‌ها در آمار در قضیه اسکالر (۱۹۵۹) بیان شده است.

**قضیه ۲.۲** (اسکلار). فرض کنید  $X$  و  $Y$  دو متغیر تصادفی با تابع توزیع توأم  $H$  و توزیع‌های حاشیه‌ای  $F$  و  $G$  باشند. مفصل  $C$  که روی  $(\text{Ran}F * \text{Ran}G)$  یکتا است، وجود دارد طوری که

$$H(x, y) = C(F(x), G(y)), \quad \forall x, y \quad (1.2)$$

علاوه بر این اگر  $C(u, v)$  یک مفصل و  $F(x)$  و  $G(y)$  توزیع‌های حاشیه‌ای باشند، آن گاه  $H(x, y)$  در رابطه (۱.۲)، یک تابع توزیع توأم با توابع حاشیه‌ای  $F$  و  $G$  است. برای جزئیات بیشتر به نلسن (۲۰۰۶) مراجعه شود.

**تعریف ۳.۲**. فرض کنید  $X$  و  $Y$  متغیرهای تصادفی با توزیع توأم  $H(x, y)$  باشند، آن گاه تابع توزیع متغیر تصادفی  $T = H(X, Y)$  را تابع توزیع کندال نامیده و با نماد  $K(t)$  نشان می‌دهند.  $K(t)$  تنها از طریق  $H(x, y)$  به مفصل متناظر  $C$  وابسته است و به توزیع‌های حاشیه‌ای  $F(x)$  و  $G(y)$  بستگی ندارد، زیرا از این که  $H(x, y) = C(F(x), G(y))$ ، برای هر  $t \in I$  داریم:

$$\begin{aligned} K(t) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{[H(x,y) \leq t]} dH(x, y) \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I_{[C(F(x), G(y)) \leq t]} dC(F(x), G(y)) \\ &= \int_0^1 \int_0^1 I_{[C(u,v) \leq t]} dC(u, v) = P(C(U, V) \leq t). \end{aligned}$$

علت نام‌گذاری تابع توزیع  $K$  به نام تابع توزیع کندال رابطه این توزیع با  $\tau$  کندال است که به صورت زیر بیان می‌شود:

$$\tau(X, Y) = 4E(H(X, Y)) - 1 = 4E(C(U, V)) - 1 = 3 - 4 \int_0^1 K(t) dt.$$

### ۳. تشخیص وابستگی با استفاده از نمودار

در این بخش پس از معرفی نمودارهای پراکندگی، کای و کندال نحوه تشخیص وابستگی با استفاده از این نمودارها را مورد بررسی قرار می‌دهیم.

**۱.۳. نمودار پراکندگی.** یک ابزار گرافیکی ساده برای ارزیابی استقلال متغیرهای  $X$  و  $Y$ ، رسم نمودار پراکندگی است. فرض کنید  $(X_1, Y_1)$  و... و  $(X_n, Y_n)$  یک نمونه تصادفی به حجم  $n \geq 2$  از تابع توزیع تجمعی دو متغیره پیوسته  $H$  با توابع حاشیه‌ای  $F$  و  $G$  باشد.

شکل ۱، مثالی بر اساس  $n = 100$  زوج شیه سازی شده از توزیع‌های نمایی مستقل  $F(x) = 1 - \exp(-2x)$  و  $G(y) = 1 - \exp(-10y)$  را بیان می‌کند و شکل ۲، زوج‌های  $(U_i, V_i) = (F(X_i), G(Y_i))$ ،  $1 \leq i \leq n$  را نمایش می‌دهد. همان طور که ملاحظه می‌شود داده‌ها روی حاشیه‌ها به صورت تصادفی پراکنده شده‌اند که استقلال را نشان می‌دهد.

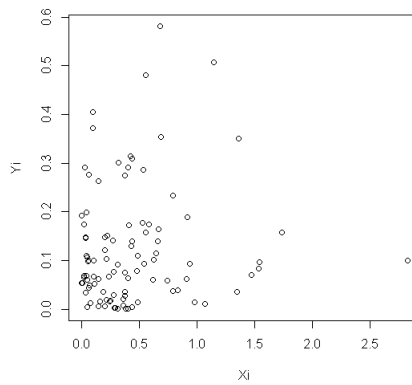
اگر چه توزیع‌های حاشیه‌ای اساساً مجهول هستند، یک روش ساده برای کنترل اثر آن‌ها، رسم زوج‌های  $(\hat{F}_n(X_i), \hat{G}_n(Y_i))$ ،  $1 \leq i \leq n$  می‌باشد که در آن  $\hat{F}_n$  و  $\hat{G}_n$  به ترتیب توابع توزیع تجربی  $X_i$  و  $Y_i$  هستند. یعنی:

$$\hat{F}_n(t) = \frac{1}{n} \#\{i, X_i \leq t\}$$

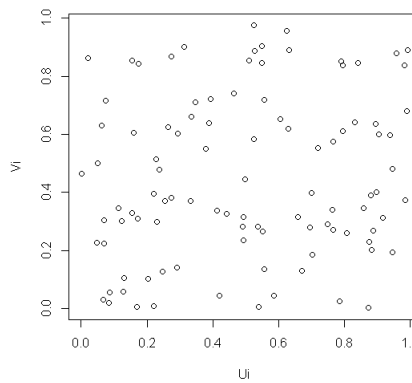
$$\hat{G}_n(t) = \frac{1}{n} \#\{i, Y_i \leq t\}$$

که در آن  $\#$  نشان‌دهنده تعداد است. رسم زوج‌های  $(\hat{F}_n(X_i), \hat{G}_n(Y_i))$  معادل با رسم زوج‌های  $(\frac{R_i}{n}, \frac{S_i}{n})$  می‌باشد که در آن  $R_i$  رتبه‌ی  $X_i$  ها و  $S_i$  رتبه‌ی  $Y_i$  ها است. چون رتبه‌ها اغلب تحت تبدیلات یکنوا پایا هستند، داده‌های تبدیل یافته ممکن است به عنوان مشاهداتی از مفصل منحصر به فرد زیر نیز ملاحظه شوند:

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)).$$



شکل ۱: نمودار پراکندگی متغیرهای تصادفی نمایی



شکل ۲: نمودار پراکندگی با حاشیه‌های یکنواخت روی  $[0, 1]$

۲.۳. نمودار کای. در این قسمت به نحوه رسم نمودار کای می پردازیم و سپس با ذکر یک مثال نحوه تشخیص وابستگی با استفاده از این نمودار را بیان می کنیم.  
فرض کنید  $(X_1, Y_1)$  و ... و  $(X_n, Y_n)$  یک نمونه تصادفی از  $(X, Y)$  با تابع توزیع توأم  $H$  باشند. برای هر زوج  $(X_i, Y_i)$  با  $1 \leq i \leq n$ ، فرض کنید:

$$H_i = \frac{1}{n-1} \# [j \neq i : X_j \leq X_i, Y_j \leq Y_i],$$

$$F_i = \frac{1}{n-1} \# [j \neq i : X_j \leq X_i],$$

$$G_i = \frac{1}{n-1} \# [j \neq i : Y_j \leq Y_i].$$

آن گاه نمودار کای، نمودار پراکندگی زوج  $(\lambda_i, \chi_i)$  برای  $\lambda_i \in [-\frac{1}{n-1}, \frac{1}{n-1}]$  می باشد طوری که

$$\chi_i = \frac{H_i - F_i G_i}{\sqrt{F_i(1-F_i)G_i(1-G_i)}} \quad (1.3)$$

و

$$\lambda_i = \text{sign}(\tilde{F}_i \tilde{G}_i) \max(\tilde{F}_i^2, \tilde{G}_i^2). \quad (2.3)$$

که در آن  $\tilde{F}_i = F_i - \frac{1}{n-1}$  و  $\tilde{G}_i = G_i - \frac{1}{n-1}$  و  $\lambda_i \in [-1, 1]$  اندازه‌ی فاصله‌ی بردار  $(X_i, Y_i)$  از مرکز مجموعه داده‌ها است. طرف راست رابطه‌ی (۱.۳)، ضریب همبستگی  $n-1$  بردار  $(X_{ij}, Y_{ij})$  را نشان می‌دهد که  $X_{ij}$  و  $Y_{ij}$  به صورت زیر تعریف می‌شوند:

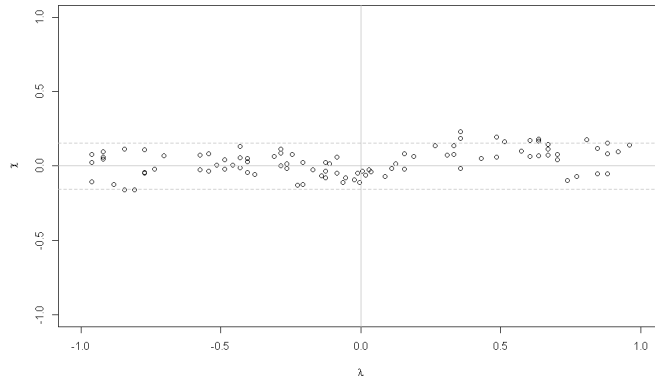
$$X_{ij} = \begin{cases} 1; & X_j \leq X_i \\ 0; & X_j > X_i \end{cases}, \quad Y_{ij} = \begin{cases} 1; & Y_j \leq Y_i \\ 0; & Y_j > Y_i \end{cases}$$

نمودار کای، نمودار پراکندگی  $(\lambda_i, \chi_i)$  با حدود کنترل در  $[-\frac{c_p}{\sqrt{n}}, \frac{c_p}{\sqrt{n}}]$  می باشد که در آن  $c_p$  توسط روش مونت کارلو شبیه سازی شده و  $c_p$  تقریباً  $100 \times p$  درصد زوج‌های  $(\lambda_i, \chi_i)$  بین خطوط واقع شده را مشخص می‌کند.

اگر  $Y_i$  ها صعودی (نزولی) و  $Y$  بطور مثبت (بطور منفی) با  $X$  وابسته باشد، آن گاه  $\lambda_i > 0$  (  $\lambda_i < 0$  ). همچنین  $\lambda_i > 0$  یعنی هر دو  $X_i$  و  $Y_i$  نسبت به میانه ی توزیع مربوطه بزرگ یا کوچک می‌شوند.  $\lambda_i < 0$  یعنی  $X_i$  و  $Y_i$  روی طرف مقابل نسبت به میانه مربوطه آن‌ها هستند.  
نمودار کای در سری زمانی و شرایطی که یکی از متغیرها غیر تصادفی است، به کار برده می‌شود. همان طور که توسط فیشر و اسویتزر (۱۹۸۵) توضیح داده شده، صورت کسر  $\chi_i$  مربوط به آزمون‌های استاندارد ناپارامتری استقلال بر اساس ضریب همبستگی روی اسپیرمن  $(\rho_n)$  و توی کندال  $(\tau_n)$  بیان می‌شود. که می توان آن را به صورت زیر نوشت:

$$\sum_{i=1}^n (H_i - F_i G_i) = \frac{n}{12} (3\tau_n - \frac{n+1}{n-1} \rho_n).$$

شکل ۳ نمودار کای توسط  $100$  زوج تصادفی نمایی مستقل مشابه شکل ۱ را نشان می‌دهد که در آن  $c_p \approx 1/78$  استفاده شده که متناظر با  $p = 0/95$  است و تقریباً همه‌ی نقاط بین دو خط قرار گرفته‌اند. به طور کلی قرار گرفتن مشاهدات بین حدود افقی استقلال را نشان می‌دهد.



شکل ۳: نمودار کای متغیرهای نمایی

۳.۳. نمودار کندال. نمودار کندال چندین ویژگی توصیفی نمودار کای، شامل تکیه روی رتبه‌ها و نسبت آن‌ها با آزمون‌های ناپارامتری استقلال را حفظ می‌کند. در این قسمت نحوه رسم نمودار کندال را بیان کرده و سپس با استفاده از یک مجموعه داده واقعی به نحوه تشخیص وابستگی از روی این نمودار می‌پردازیم. یک راه مشترک ارزیابی نرمال بودن به روش تصویری، رسم نمودار  $Q - Q$  می‌باشد که زوج‌های  $(Z_{i:n}, X_{(i)})$  را رسم می‌کند که در آن  $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$  نشان دهنده آماره‌های ترتیبی نمونه و  $Z_{i:n}$ ، چندک مرتبه  $i$  ام توزیع نرمال مرتبط با نمونه تصادفی نرمال استاندارد به حجم  $n$  می‌باشد طوری که  $Z_{i:n} = E(Z_{(i)})$  که در آن  $Z_{(1)} \leq Z_{(2)} \leq \dots \leq Z_{(n)}$  آماره‌های ترتیبی نمونه تصادفی  $Z_1$  و  $\dots$  و  $Z_n$  از توزیع نرمال استاندارد است.

روش ساخت این نمودار برای یک نمونه تصادفی  $(X_1, Y_1)$  و  $\dots$  و  $(X_n, Y_n)$  به صورت زیر است: ۱- برای هر  $1 \leq i \leq n$ ،  $H_i$  را به صورت زیر تعریف می‌کنیم:

$$H_i = \frac{1}{n-1} \# [j \neq i : X_j \leq X_i, Y_j \leq Y_i].$$

۲- مقادیر  $H_i$  را به صورت  $H_{(1)} \leq H_{(2)} \leq \dots \leq H_{(n)}$  مرتب می‌کنیم.

۳- نمودار زوج‌های  $(W_{i:n}, H_{(i)})$  را رسم کرده که در آن  $W_{i:n}$  بیانگر امید ریاضی  $i$  امین آماره مرتب در یک نمونه تصادفی به حجم  $n$  از توزیع  $K_0(t) = t - t \ln t$  است. در نتیجه داریم:

$$W_{i:n} = n \binom{n-1}{i-1} \int_0^1 t [K_0(t)]^{i-1} \times [1 - K_0(t)]^{n-i} dK_0(t). \quad (3.3)$$

شکل ۴ نمودار کندال یک نمونه تصادفی به اندازه  $n = 100$  از متغیرهای تصادفی نمایی مستقل همانند شکل‌های ۱ و ۳ می‌باشد که به دلیل نزدیک بودن نقاط به خط راست استقلال را نشان می‌دهد. جنست و رایوست (۱۹۹۳) بیان کردند که تحت شرایط نظم متعادل، تابع توزیع تجربی  $K_n$  از شبه مشاهدات  $H_1, \dots, H_n$  به‌طور مجانبی گاوسی، برآوردگر  $\sqrt{n}$ -سازگار برای  $H(X, Y) \leq t$  است که همگرایی  $\hat{H}_n$  به  $H$  را نتیجه می‌دهد به طوری که  $H_i = \hat{H}_n(X_i, Y_i)$  تابع توزیع تجربی بر اساس  $(X_j, Y_j)$ ،  $i \neq j$  می‌باشد و ثابت کردند که  $K_n(t)$  در احتمال به  $K(t)$  همگرا است. بنابراین



برای  $0 \leq p \leq 1$  در  $K_n^{-1}(p)$  احتمال به  $K^{-1}(p)$  همگرا است. چون  $K_n$  و عکس آن کراندار هستند، این همگرایی به طور طبیعی نتایج زیر را می‌دهد:

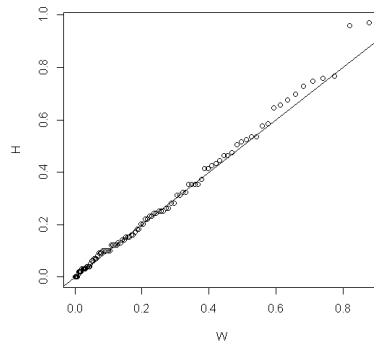
۱- برای عدد صحیح دلخواه  $n \geq 1$  و  $0 \leq p \leq 1$ ، فرض کنید  $[np]$  کوچک ترین عدد صحیح بزرگ تر یا مساوی  $np$  باشد، آن گاه

$$H_{([np])} = K_n^{-1}(p) \xrightarrow{n \rightarrow \infty} K^{-1}(p),$$

بنابراین تحت فرضیه صفر استقلال داریم:

$$\lim_{n \rightarrow \infty} E(H_{([np])}) = \lim_{n \rightarrow \infty} W_{[np]:n} = K_0^{-1}(p).$$

- ۲- برای نمونه به حجم کافی بزرگ  $n$ ، زوج‌های  $(W_{i:n}, H_{(i)})$  تمایل به تمرکز در امتداد منحنی  $w \mapsto (K_0^{-1}(p), K^{-1}(p))$  دارند. به عبارت دیگر، نقاط در نمودار کندال مانند نمودار  $w \mapsto K^{-1}\{K_0(w)\}$  است.
- ۳- تحت فرضیه صفر استقلال، وقتی که  $K = K_0$ ، گراف تمایل به خطی شدن دارد.
- ۴- وقتی که متغیرهای  $X$  و  $Y$  هم یکنوا با  $-1$  باشند،  $\tau(X, Y) = -1$  تمام نقاط روی نمودار در محور افقی ( $p \equiv 0$ ) قرار می‌گیرند، زیرا در این حالت برای همه مقادیر ممکن  $0 \leq p \leq 1$ ،  $K^{-1}(p) = 0$  است.
- ۵- وقتی که متغیرهای  $X$  و  $Y$  هم یکنوا با  $1$  باشند،  $\tau(X, Y) = 1$  چون روی بازه  $[0, 1]$ ،  $K^{-1}(p) = p$  است، آن گاه تمام نقاط روی منحنی  $K_0(p)$  قرار می‌گیرند.



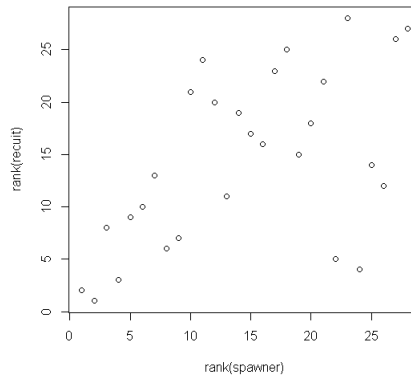
شکل ۴: نمودار کندال متغیرهای نمایی

#### ۴. تحلیل داده های واقعی

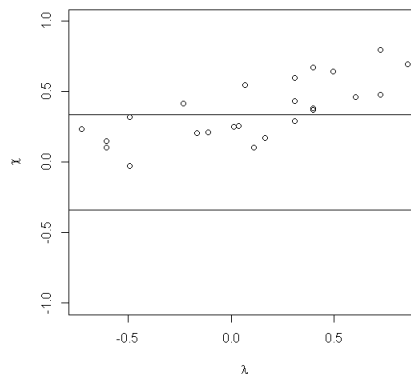
در این بخش برای یک مجموعه داده شامل ۲۸ اندازه‌ی سهام تخم ریزی سالانه ماهی آزاد و تولید مربوط به اندازه ماهی جدید در رودخانه اسکینا (کلمبیا پریتیش، کانادا) که این مجموعه داده توسط کالنبرگ و لدوینا (۱۹۹۹) آرایه شده است، نمودارهای پراکندگی، کای و کندال را رسم کرده و از نظر وابستگی با یکدیگر مقایسه می‌کنیم.

شکل‌های ۵ و ۶ و ۷ نمودار پراکندگی، نمودار کای و نمودار کندال را برای این داده ها نشان می‌دهد. تمام نمودارها بیانگر ارتباط مثبت کلی در داده ها است (به دلیل این که  $\tau_n \approx 0.41$  و

۵۵/۰)  $\rho_n \approx$  نمودار کندال وجود یک خوشه در گوشه سمت چپ پایین از نمودار پراکندگی رتبه‌ای را مشخص می‌کند. این پدیده نسبت به برابری چندین  $H_i$ ، تقریباً در امتداد خط  $y = 1 - x$  در نمودار پراکندگی رتبه‌ای قرار گرفته است. نمودار کای ارتباط مثبت بین متغیرها را نشان می‌دهد چون داده‌ها بالای حدود کنترل قرار دارند. فیشر و اسویتزر (۲۰۰۱) توجهی به این ویژگی از داده‌ها بر اساس نمودار کای نکردند. همچنین نمودار پراکندگی نشان می‌دهد، ارتباط خطی قابل توجهی در داده‌ها وجود دارد.

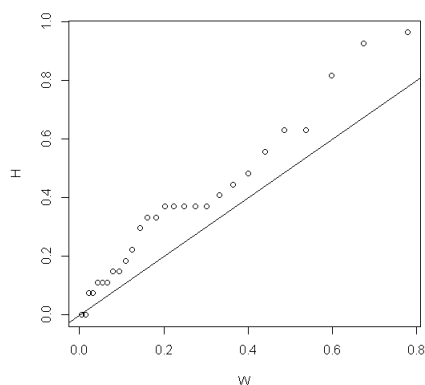


شکل ۵: نمودار پراکندگی اندازه‌ی سهام تخم‌ریزی سالانه ماهی آزاد و تولید مربوط به اندازه ماهی جدید



شکل ۶: نمودار کای اندازه‌ی سهام تخم‌ریزی سالانه ماهی آزاد و تولید مربوط به اندازه ماهی جدید

از این مثال می‌توان نتیجه گرفت که نمودار کندال نسبت به نمودار کای آسان‌تر تفسیر می‌شود.



شکل ۷: نمودار کندال اندازه ی سهام تخم ریزی سالانه ماهی آزاد و تولید مربوط به اندازه ماهی جدید

### مراجع

1. Fisher, N. I. and Switzer, P. (1985). *Chi-Plots for Assessing Dependence*. Biometrika, 72, 253-265.
2. Fisher, N. I. and Switzer, P. (2001). *Graphical Assessment of Dependence: Is a Picture Worth 100 Tests?*. The American Statistician, 55, 233-239.
3. Genest, C. and Rivest, L.-P. (1993). *Statistical inference procedures for bivariate Archimedean copulas.*. Journal of the American Statistical Association, 88, 1034-1043.
4. Genest, C. and Boies, J. C. (2003). *Statistical Models and Methods for Lifetime Data Detecting dependence with kendall plots*. Amer. statist. Vol. 57, No. 4, pp. 275-284.
5. Kallenberg, W. C. M. and Ledwina, T.(1999). *Data-Driven Rank Tests for Independence*. Journal of American Statistical Association, 94, 285-301.
6. Sklar, A.(1959). *Functions de repartition a n dimensions et leurs marges*. Publ. Inst. Statist. Univ. Paris 8, 229-231.



## تخمین دبی پیک سیلاب با استفاده از تابع مفصل

زراعتی، س. <sup>۱</sup> \*، ذونعمت کرمانی، م. <sup>۲</sup>، و رضاپور، م. <sup>۳</sup>

<sup>۱،۲</sup> بخش مهندسی آب، دانشکده کشاورزی، دانشگاه شهید باهنر کرمان  
zeraati@agr.uk.ac.ir

zounemat@uk.ac.ir

<sup>۳</sup> بخش آمار، دانشکده علوم ریاضی و آمار، دانشگاه شهید باهنر کرمان  
mohsenrzp@uk.ac.ir

چکیده. سیل به عنوان یکی از مهمترین بلایای طبیعی در کشور مطرح است. در حقیقت سیل افزایش ارتفاع آب در رودخانه می باشد که باعث کاهش ظرفیت طبیعی رود و اشغال بخشی از دشت های حاشیه رودخانه می شود. علت وقوع این سیلاب ها باران های شدید، دخالت های بی رویه و فعالیت های بشر در نقاط مختلف، توسعه بی رویه کشاورزی در بسیاری از مناطق و نابودی جنگل ها و مراتع می باشد. از آنجا که بسیاری از شهرها و زیرساخت ها از قبیل بزرگ راه ها، سیستم های زهکشی در نزدیکی یا محل تلاقی جریان های رودخانه ها واقع شده اند با شناخت عوامل بروز سیلاب و بررسی تعدادی از سیل ها با دوره بازگشت مشخص میتوان با درجه ای از تقریب احتمال وقوع سیلاب را مشخص کرد. در چند دهه اخیر این مساله مد نظر بسیاری از محققین و طراحان قرار گرفته است. از آنجا که جریان در یک شاخه از رودخانه می تواند تحت تاثیر جریان شاخه های دیگر رودخانه باشد، دانستن ارتباط بین جریان ها در بسیاری از طرح های هیدرولوژیکی حائز اهمیت است. از این رو برای تخمین دبی پیک در محل تلاقی جریان های رودخانه ها نیازمند روشی کارآمد هستیم. در این تحقیق از روش آماری مفصل برای تخمین دبی پیک سیلاب استفاده شد که نتایج نشان داد تخمین دبی پیک در منطقه مورد مطالعه از حوضه جازموریان با استفاده از تابع مفصل نسبت به سایر روش ها بهتر است.

### ۱. پیش‌گفتار

از روش‌های معمول که در حال حاضر برای تخمین دبی پیک سیل به کار گرفته می شود برنامه فراوانی سیل ملی بر پایه آنالیز منطقه‌ای است که می توان برای تخمین سیل در محل تلاقی جریان‌های شاخه‌های رودخانه به کار رود (یو اس جی اس، ۲۰۰۲). روش‌های بر مبنای بارش-رواناب از روش‌های تجربی هستند که در عمل با توجه به در دسترس بودن مجموعه ی گسترده‌ای از داده‌های بارش-رواناب مورد استفاده قرار می گیرند. روش‌های بر مبنای دبی شامل رگرسیون چند متغیره و تجزیه و تحلیل تناوب‌های سیل منطقه می باشد (هاسکینگ و والیس، ۱۹۹۷) که میزان دبی سیلاب از طریق رگرسیون جریان و

2010 Mathematics Subject Classification. Primary 47A55; Secondary 39B52, 34K20, 39B82.

واژگان کلیدی. دبی پیک، سیلاب، تابع مفصل.

\* سخنران

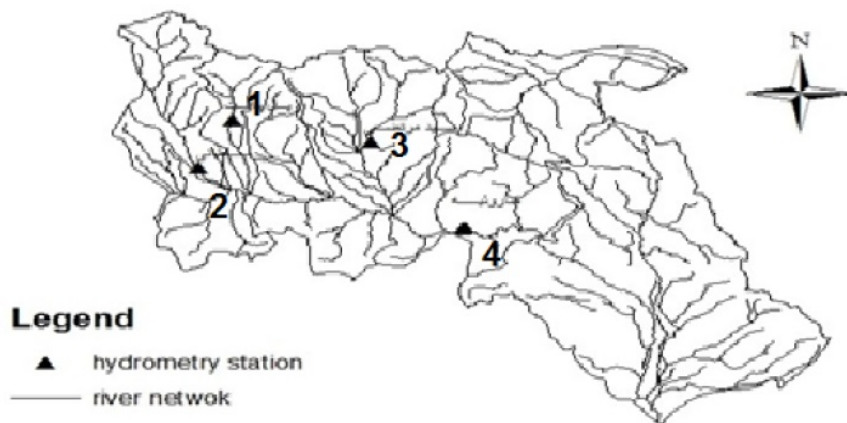
ویژگی‌های حوضه آبخیز برآورد می‌شود. برخی روش‌های دیگر ممکن است به اندازه کافی برای مدل کردن تغییرات انعطاف پذیر و قابل اجرا نباشند، چون اکثر رخدادهای پیچیده هیدرولوژیکی مثل سیلاب‌ها، طوفان‌ها رخدادهای چند متغیره‌ای هستند که توسط تعدادی متغیرهای تصادفی توصیف می‌شوند، هرچند توزیع‌های چند متغیره تعمیم یافته از توزیع‌های یک متغیره بسیاری در مقالات مورد بررسی قرار گرفته است اما مشکل اساسی این توزیع‌ها این است که خانواده‌های توزیع‌های چند متغیره و حاشیه‌ای یکسان است. در بیشتر موارد استفاده از توزیع نرمال چند متغیره برای مدل کردن دبی‌های پیک به این علت که توزیع‌های حاشیه‌ای مربوط به داده‌ها نامتقارن هستند مناسب نمی‌باشد. به علاوه در توزیع‌های حاشیه‌ای پیچیده‌تر مثل توزیع‌های ترکیبی که در حال حاضر به طور گسترده برای مدل کردن پدیده‌های ناهمگن استفاده می‌شوند (تتیرینگتن و همکاران، ۱۹۸۵؛ وست ۱۹۹۲) امکان استفاده از توزیع‌های چند متغیره استاندارد وجود ندارد. ساختار چند متغیره‌ای که محدودیت‌ها و مشکلات ذکر شده را نداشته باشد بر اساس مفهوم تابع مفصل قابل بیان است. این توابع اولین بار در هیدرولوژی به وسیله (دی مایکل و سالوادوری، ۲۰۰۳) برای ایجاد یک مدل دو متغیره توصیف کننده شدت و مدت بارش استفاده شدند. پس از آن مفهوم تابع مفصل به سرعت در زمینه‌های مختلف هیدرولوژی به کار گرفته شد. در این مقاله به کارگیری تابع مفصل جهت تخمین دبی پیک در محل تلاقی جریان‌های رودخانه برای زیر ساخت‌ها بسیار ضروری است چون وقتی که این زیر ساخت‌ها در نزدیکی محل تلاقی جریان قرار می‌گیرند ممکن است از طریق یک یا چند جریان بالا به زیر آب روند که تابع مفصل با ساخت ساختار وابسته بین شاخه‌های جریان بالادست راهی برای ساختن توزیع‌های چند متغیره با انتخاب توزیع‌های حاشیه‌ای مختلف فراهم می‌کند که در نتیجه تخمین دبی پیک پایین دست را امکان پذیر می‌سازد.

## ۲. مواد و روش‌ها

در روش‌های چند متغیره، تجربی و رگرسیونی محدودیت انتخاب توزیع‌های چند متغیره و حاشیه‌ای یکسان وجود دارد اما در واقعیت جریان‌ها از توزیع‌های یکسان تبعیت نمی‌کنند. در این تحقیق از روش تابع مفصل استفاده شد تا با توجه به توزیع‌های حاشیه‌ای متفاوت هر یک از جریان‌های بالادست بهترین ساختار توام بدست آید و با توجه به آن جریان‌های پیک پایین دست محاسبه شود. از این رو ابتدا داده‌های پیک سالانه و روزانه برای ۱۷ سال را برای چهار ایستگاه هیدرومتری جدول (۱) واقع در حوضه مورد مطالعه شکل (۱) اخذ گردید. سپس برای ساخت توزیع توام، توزیع‌های حاشیه‌ای جریان شاخه‌های فرعی معرفی شد که نتایج نشان داد توزیع‌های برازش داده شده تقریب مناسبی برای داده‌های جریان منطقه فراهم می‌کنند. با معرفی تابع مفصل، توزیع‌های چند متغیره با حاشیه‌های انتخابی ساخته شد و بر طبق ساختار توزیع توام که با توجه به مفصل گامبل از خانواده مفصل‌های ارشمیدسی بدست آمد، دبی پیک پایین دست در ایستگاه ۴ از طریق دو ساختار اختصاص داده شده با تولید اعداد تصادفی از طریق تابع مفصل و دو ساختار دیگر با توجه به مقادیر اولیه جریان محاسبه شد. در نهایت ساختار جریان‌های پیک پایین دست ارزیابی شدند.

جدول ۱: موقعیت ایستگاه‌های هیدرومتری

نام ایستگاه	شماره ایستگاه	طول جغرافیایی	عرض جغرافیایی
پل بافت	۱	۵۶°۳۸'	۲۹°۱۵'
سلطانی	۲	۵۶°۳۲'	۲۹°۰۵'
سید مرتضی	۳	۵۶°۵۸'	۲۹°۱۲'
کنارویه	۴	۵۸°۱۸'	۲۸°۵۲'



شکل ۱: منطقه مورد مطالعه و موقعیت ایستگاه ها

## ۳. دست‌آورد های پژوهش

برای ارزیابی دبی پیک چهار ساختار در نظر گرفته شد که ساختار اول و دوم به ترتیب از جمع جریان های بالادست و یک رابطه رگرسیونی بر اساس مفصل برازش داده شده بدست آمدند و ساختار سوم و چهارم به ترتیب از جمع جریان های بالادست و یک رابطه رگرسیونی بین جریان ها حاصل از مقادیر جریان اولیه بدست آمدند. به این منظور برای ساختارهای اول و دوم یک مجموعه شبیه سازی شده برای محاسبه پیک جریان و تخمین توزیع احتمالاتی با توجه به مفصل گامبل بدست آمد و با توجه به توزیع برازش داده شده بر هر یک از ساختارها جریان پیک پایین دست محاسبه شد. ساختار سوم و چهارم از مقادیر اولیه بدست آمدند. نتایج ارزیابی در جدول ۲ آورده شده است. مهمترین نکته در این ارزیابی تخمین احتمال توام شاخه های فرعی می باشد که تابع مفصل این مسئله را ساده کرده است و با تعداد کافی شبیه سازی و تولید اعداد تصادفی نتایج مشابه به راه حل تحلیلی بدست آمد. نتایج شبیه سازی تابع توزیع توام ساختار دوم نتایج بهتری را نسبت به ساختارهای دیگر نشان می دهد. می توان نتیجه گرفت که روش احتمال توام یک راه آسان برای تخمین دبی پیک در محل تلاقی جریان ها با دقت قابل قبولی را فراهم می کند که نیاز به ثبت داده های دبی در رودخانه اصلی نمی باشد. همچنین روش مفصل برای تابع توزیع چند متغیره ابزاری قوی برای کاهش فرضیات در اکثر توابع توزیع چند متغیره با توابع توزیع یکسان می باشد که از محاسبات پیچیده اجتناب می کند و کاربرد آن نیز بسیار ساده می باشد.

جدول ۲: جدول ارزیابی دبی پیک شبیه سازی شده ایستگاه x۴

ساختار	RMSE	$R^2$	NASH
اول	۱۶۶.۷۲	۰.۸۸۶	۰/۲۵۱۱
دوم	۱۴۹.۱۶	۰.۸۸۱	۰/۰۰۱۵
سوم	۱۹۴.۳۲	۰.۹۳۷	۰/۶۹۹۶
چهارم	۱۹۵.۸۳	۰.۹۷	۰/۷۲۶۷

## مراجع

1. De Michele, C. and Salvadori, G.(2003), *A Generalized Pareto intensity-duration model of storm rainfall exploiting 2-Copulas*, J. Geophys. Res.,108(D2), doi:10.1029/2002JD002534.
2. Hosking, J. R. M., and Wallis, J. R. (1997), *Regional frequency analysis, 1st Ed.*, Cambridge University Press, New York.
3. Titterton, D. M., A. F. M. Smith, and U. E. Makov (1985) , *Statistical Analysis of Finite Mixture Distributions*, John Wiley, Hoboken, N. J.
4. U.S. Geological Survey (2002), *The National Flood Frequency Program, Version 3: A Computer Program for Estimating Magnitude and Frequency of Floods for Ungaged Sites*, U.S. Geological Survey Water-Resources Investigations Report 02-4168
5. West, M. (1992) , *Modelling with mixtures Bayesian Statistics 4*. New York: Oxford University press, 503-24