

## تحلیل وابستگی دو متغیره با استفاده از اندازه‌های واگرایی جفری و هلینجر براساس برآورد تابع چگالی مفصل به روش تبدیل پروبیت بهبودیافته

مرتضی محمدی، مهدی عمادی، محمد امینی

گروه آمار، دانشگاه فردوسی مشهد

تاریخ دریافت: ۱۳۹۸/۱۰/۲۴ تاریخ پذیرش و انتشار: ۱۳۹۹/۰۴/۰۶

**چکیده:** اندازه‌های واگرایی می‌توانند به عنوان معیارهای وابستگی در نظر گرفته می‌شوند و برحسب تابع چگالی مفصل بازنویسی شوند. در این مقاله، معیارهای وابستگی جفری و هلینجر با استفاده از روش تبدیل پروبیت بهبودیافته برآورد می‌شوند و سازگاری مجانبی آن‌ها اثبات می‌گردد. علاوه بر این، یک مطالعه شبیه‌سازی برای سنجش دقت برآوردها انجام شده است. نتایج شبیه‌سازی نشان می‌دهند که برای حجم نمونه کم یا شدت وابستگی ضعیف، معیار وابستگی هلینجر عملکردی بهتری نسبت به معیارهای وابستگی کولبک-لیبلر و جفری دارند. در انتها، کاربردی از روش‌های مورد بررسی در هیدرولوژی ارائه شده است. **واژه‌های کلیدی:** واگرایی، وابستگی، چگالی مفصل، تبدیل پروبیت.

## ۱ مقدمه

یکی از مهمترین نکات مورد توجه هنگام کار با داده‌ها، بررسی وابستگی بین متغیرها است. برای بررسی وابستگی بین متغیرها از معیارهای مختلفی استفاده می‌شود. از جمله پرکاربردترین این معیارها می‌توان به ضرایب همبستگی خطی پیرسون،  $\rho$ -اسپیرمن ( $\rho_s$ ) و  $\tau$ -کندال<sup>۱</sup> ( $\tau$ ) اشاره کرد. توان آزمون استقلال براساس این معیارها در برخی موارد حتی در حجم نمونه بی‌نهایت به سمت یک میل نمی‌کند. برای حل این

آدرس الکترونیکی نویسنده مسئول مقاله: مهدی عمادی، emadi@um.ac.ir  
کد موضوع بندی ریاضی (۲۰۱۰): 62G10, 62G07, 62H15

<sup>1</sup>Kendall's Tau

مشکل، معیار وابستگی و آزمون استقلال براساس فاصله کرامر-فون میزس<sup>۱</sup> ( $CvM$ ) بین تابع توزیع توام و حاصل ضرب توابع توزیع حاشیه‌ای توسط بلوم و همکاران (۱۹۶۱) پیشنهاد گردید.

آزمون استقلال براساس فاصله کرامر-فون میزس بین تابع مفصل تجربی و حاصل ضرب توابع حاشیه‌ای توسط جنست و رمیلارد (۲۰۰۴) پیشنهاد شد. بلالیا و همکاران (۲۰۱۷) نشان دادند که آزمون‌های استقلال براساس تابع مفصل تجربی، در حالتی که وابستگی ضعیف باشد یا وابستگی در دم‌ها وجود داشته باشد، عملکرد ضعیفی دارند. همچنین آن‌ها نشان دادند که وقتی داده‌ها از مفصل تی ( $T$ ) با درجه آزادی ۲ و  $\tau$ -کندال برابر با صفر تولید شوند، توان آزمون استقلال براساس مفصل تجربی و مفصل برآورد شده به روش برنشتاین بسیار ضعیف است و در نهایت آن‌ها برای رفع این مشکل، استفاده از تابع چگالی مفصل را به جای تابع توزیع مفصل پیشنهاد کردند.

از طرف دیگر، برای حل مشکلات مربوط به ضرایب همبستگی کلاسیک، اندازه و اگرایی کولبک-لیبلر<sup>۲</sup> ( $KL$ ) بین تابع چگالی توأم متغیرها و حاصل ضرب توابع چگالی حاشیه‌ای آن‌ها به عنوان معیاری برای اندازه‌گیری وابستگی بین متغیرها توسط جو (۱۹۸۹) معرفی شد. میچز و زاگرافوس (۲۰۰۶) از اندازه‌های فی-واگرا به عنوان معیار وابستگی و آزمون استقلال برای توزیع‌های نرمال و تی چند متغیره استفاده کردند. آن‌ها اندازه‌های فی-واگرا را با استفاده از برآورد ماکسیمم درست‌نمایی پارامترها، برآورد کردند. محتشمی برزادران و امینی (۲۰۱۰) اندازه‌های و اگرایی مختلف را برحسب تابع مفصل بدست آوردند و خواص آن‌ها را نیز مورد بررسی قرار دادند. خسروی و همکاران (۱۳۹۱) فاصله هلینجر را به عنوان اندازه فی-واگرای بهینه برای بررسی میزان وابستگی در خانواده مفصل فارلی-گامبل-مورگنسترن تعمیم یافته پیشنهاد دادند. ما و سان (۲۰۱۱) برای اولین بار رابطه بین اطلاع متقابل و تابع مفصل را با عنوان "آنتروپی مفصل" معرفی کردند. بلومنتریت و اسمیت (۲۰۱۲) آنتروپی مفصل را که برابر با منفی اطلاع متقابل است، به عنوان معیاری برای بیان وابستگی در نظر گرفتند. این معیار وابستگی برحسب تابع چگالی مفصل است و آن‌ها برای برآورد این معیار از روش‌های ناپارامتری مختلفی برای برآورد تابع چگالی مفصل استفاده کردند.

با توجه به محدود بودن تکیه‌گاه تابع چگالی مفصل در بازه  $[0, 1]$ ، روش‌های استاندارد برآورد تابع چگالی دو متغیره برحسب تابع هسته دقت مناسبی به‌ویژه در نزدیکی مرزها برای برآورد تابع چگالی مفصل ندارند. چارپنتیر و همکاران (۲۰۰۶) نشان دادند که برآوردگرهای حاصل از روش‌های استاندارد، مقدار قابل توجهی از چگالی احتمال را خارج از مربع واحد قرار می‌دهد و در نتیجه ممکن است برآورد تابع

<sup>1</sup>Cramer-von Mises

<sup>2</sup>Kullback-Leibler

چگالی مفصل به روش استاندارد یک تابع چگالی در بازه  $[0, 1]^2$  نباشد. برای حل این مشکلات که به مشکلات مرزی معروف است، برخی روش‌های ناپارامتری را برحسب تابع هسته از جمله انعکاس آینه<sup>۱</sup>، تبدیل پروبیت<sup>۲</sup> و هسته بتا<sup>۳</sup> توسط چارپنتیر و همکاران (۲۰۰۶) پیشنهاد شد.

ناگلر (۲۰۱۴) نشان داد که در بین روش‌های ناپارامتری، روش تبدیل پروبیت عملکرد بهتری دارد و مشکلات مرزی برآورد تابع چگالی مفصل را رفع خواهد کرد، اما مشکل ضرب‌های نامتناهی منجر به عملکرد ضعیف آن در نزدیکی مرزها خواهد شد. با توجه به این مشکلات، ون و وو (۲۰۱۵) روش اصلاح‌شده تبدیل پروبیت را برای برآورد تابع چگالی مفصل معرفی کرده‌اند. جیننس و همکاران (۲۰۱۷) یک روش تبدیل پروبیت اصلاح شده دیگر با عنوان تبدیل پروبیت درستنمایی موضعی<sup>۴</sup> ( $LLPT$ ) معرفی کردند. بعلاوه روش موجک (جنست و همکاران، ۲۰۰۹)، روش برنشتاین (بوزمارنی و همکاران، ۲۰۱۳) و روش اسپلاین (کایرمن و چلاس، ۲۰۱۳) برای برآورد تابع چگالی مفصل بررسی شده است. ناگلر (۲۰۱۸) یک مقایسه جامع بین روش‌های مختلف برآورد تابع چگالی مفصل انجام داده است و کدهای لازم را در بسته  $kdecompula$  در نرم افزار  $R$  قرار داده است. او نشان داد که روش  $LLPT$  از درجه دوم با برآورد پارامتر هموارسازی به روش نزدیک‌ترین همسایگی در سطوح مختلف وابستگی دارای عملکرد بهتری در مقایسه با سایر روش‌های مورد بررسی است.

در این مقاله، برای برآورد معیارهای وابستگی جفری و هلینجر که برحسب تابع چگالی مفصل هستند از روش  $LLPT$  استفاده خواهد شد. نویسندگان مقاله اندازه‌های وابستگی جفری و هلینجر بازنویسی شده برحسب تابع چگالی مفصل را با استفاده از روش  $LLPT$  برآورد کرده‌اند. این بازنویسی موجب افزایش دقت و سهولت برآورد می‌شود، زیرا در برآورد این اندازه‌ها به جای برآورد توابع چگالی توام و حاشیه‌ای‌ها کفایت تنها تابع چگالی مفصل برآورد شود. برای سهولت کار فقط حالت دو متغیره پیوسته بررسی خواهد شد. بخش ۲ مروری کوتاه بر مفاهیم مورد نیاز از تابع مفصل است. بخش ۳ به بررسی روش تبدیل پروبیت اصلاح شده برای برآورد تابع چگالی مفصل اختصاص یافته است. بخش ۴ به معرفی روش برآورد اندازه‌های وابستگی مفصل مبنا براساس اندازه‌های واگرایی جفری و هلینجر پرداخته و سازگاری آن‌ها مطالعه شده است. در بخش ۵ نتایج عددی برای مقایسه دقت برآوردگرهای مورد بررسی ارائه شده است. تحلیل داده‌های هیدرولوژی براساس اندازه‌های وابستگی مفصل مبنا از نتایج دیگر این مقاله است.

<sup>1</sup>Mirror-reflection

<sup>2</sup>Probit transformation

<sup>3</sup>Beta kernel

<sup>4</sup>Local Likelihood Probit Transformation

## ۲ تابع مفصل

تابع مفصل ارتباط بین توزیع چند متغیره و توزیع‌های حاشیه‌ای یک متغیره را بیان می‌کند (اسکلار، ۱۹۵۹). اگر  $X$  و  $Y$  دو متغیر تصادفی به ترتیب با توابع توزیع  $F_X$  و  $F_Y$  و تابع توزیع توأم  $F$  باشند، آنگاه تابع مفصل  $C$  وجود دارد به طوری که

$$F(x, y) = C(F_X(x), F_Y(y)), \quad x, y \in \mathbb{R}. \quad (۱)$$

اگر  $F_X$  و  $F_Y$  پیوسته باشند، آنگاه  $C$  منحصر به فرد است. در صورتی که تابع مفصل  $C$  مطلقاً پیوسته باشد، ارتباط بین تابع چگالی توأم  $(f)$  و تابع چگالی مفصل  $(c)$  به صورت

$$f(x, y) = c(F_X(x), F_Y(y))f_X(x)f_Y(y), \quad x, y \in \mathbb{R}, \quad (۲)$$

بدست می‌آید، که در آن  $f_X$  و  $f_Y$  به ترتیب توابع چگالی حاشیه‌ای  $X$  و  $Y$  هستند. در جدول ۱، فضای پارامتر و ضریب همبستگی  $\tau$ -کندال ( $\tau$ ) برای برخی از پرکاربردترین اعضاء خانواده مفصل‌های ارشمیدسی و بیضوی ارائه شده است، که در آن  $D_1(\theta) = \frac{1}{\theta} \int_0^\theta \frac{t}{e^t-1} dt$  تابع دبای،  $\Phi_2$  توزیع نرمال استاندارد دو متغیره با پارامتر همبستگی  $\rho$  و  $\Phi^{-1}$  معکوس توزیع نرمال استاندارد و  $t_{\nu, \rho}$  توزیع تی استاندارد دو متغیره با پارامتر همبستگی  $\rho$  و  $t_\nu^{-1}$  معکوس توزیع تی استاندارد با درجه آزادی  $\nu$  است. توابع مفصل کلیتون، گامبل-هوگارد و فرانک از اعضاء خانواده ارشمیدسی و توابع مفصل گاوسی و تی از اعضاء خانواده بیضوی هستند. ضریب  $\tau$ -کندال بین دو متغیر تصادفی با تابع مفصل  $C$  توسط رابطه  $1 - \tau = \int_{[0,1]^2} C(u, v) dC(u, v)$  بدست می‌آید (جو، ۲۰۱۴).

جدول ۱. برخی از مهم‌ترین توابع مفصل ارشمیدسی و بیضوی

مفصل	$C(u, v)$	فضای پارامتر	$\tau$ -کندال
کلیتون	$(u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}$	$\theta \in (-1, +\infty) - \{0\}$	$\frac{\theta}{\theta+2}$
گامبل-هوگارد	$\exp\{-[(-\ln u)^\theta + (-\ln v)^\theta]^{1/\theta}\}$	$\theta \in [1, +\infty)$	$\frac{\theta-1}{\theta}$
فرانک <sup>۱</sup>	$\frac{-1}{\theta} \log\left\{1 + \frac{(e^{-u\theta}-1)(e^{-v\theta}-1)}{e^{-\theta}-1}\right\}$	$\theta \in (-\infty, +\infty) - \{0\}$	$1 + \frac{2}{\theta}(D_1(\theta) - 1)$
گاوسی <sup>۲</sup>	$\Phi_2(\Phi^{-1}(u), \Phi^{-1}(v); \rho)$	$\rho \in [-1, +1]$	$\frac{2}{\pi} \arcsin(\rho)$
تی <sup>۳</sup>	$t_{\nu, \rho}(t_\nu^{-1}(u), t_\nu^{-1}(v); \rho)$	$\rho \in [-1, +1], \nu > 1$	$\frac{2}{\pi} \arcsin(\theta)$

### ۳ برآورد تابع چگالی مفصل

فرض کنید  $(X_i, Y_i)_{i=1, \dots, n}$  یک نمونه تصادفی دلخواه باشد. در حالت کلی با استفاده از روش هسته می‌توان یک برآورد ساده برای تابع چگالی مفصل را به صورت

$$\hat{c}_n(u, v) = \frac{1}{n} \sum_{i=1}^n K_{b_n}(u - \tilde{U}_i) K_{b_n}(v - \tilde{V}_i), \quad (u, v) \in [0, 1]^2, \quad (3)$$

در نظر گرفت که در آن  $K(\cdot), K_{b_n}(\cdot) = \frac{K(\cdot)}{b_n}$  تابع هسته و  $b_n$  پهنای باند است. همچنین در (۳)، مقادیر  $\tilde{U}_i = n\hat{F}_X(x_i)/(n+1)$  و  $\tilde{V}_i = n\hat{F}_Y(y_i)/(n+1)$  به عنوان شبه مشاهدات در نظر گرفته می‌شوند و  $\hat{F}_X$  و  $\hat{F}_Y$  به ترتیب توابع توزیع تجربی برای متغیرهای تصادفی  $X$  و  $Y$  هستند. برآوردگر (۳) دارای مشکلات مرزی است و روش تبدیل پروبیت رویکردی است که این مشکلات را با تغییر تکیه‌گاه محدود تابع چگالی مفصل به  $\mathcal{R}^2$  برطرف خواهد کرد.

فرض کنید بردار  $(U_i, V_i)_{i=1, \dots, n}$  نمونه‌ای تصادفی از مفصل دو متغیره  $C$  باشند. بردار حاصل از تبدیل پروبیت روی بردار شبه مشاهدات به صورت  $(\hat{S}_i, \hat{T}_i) = (\Phi^{-1}(\tilde{U}_i), \Phi^{-1}(\tilde{V}_i))$ ، یک بردار تصادفی با تابع مفصل  $C$  و حاشیه‌ای‌های نرمال استاندارد است که  $\Phi$  تابع چنک توزیع نرمال استاندارد است. با توجه به (۲)، برآورد تابع چگالی مفصل به روش تبدیل پروبیت را می‌توان به صورت

$$\hat{c}_n^{(PT)}(u, v) = \frac{\hat{f}_n(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \quad (u, v) \in (0, 1)^2, \quad (4)$$

در نظر گرفت، که در آن  $\phi$  تابع چگالی نرمال استاندارد است. یک روش استاندارد برای برآورد تابع چگالی دو متغیره  $\hat{f}_n$  در (۴) بر مبنای تابع هسته را می‌توان به صورت

$$\hat{f}_n(s, t) = \frac{1}{n|\mathbf{H}_{ST}|^{\frac{1}{2}}} \sum_{i=1}^n \mathbf{K}[\mathbf{H}_{ST}^{-\frac{1}{2}} \begin{pmatrix} s - \hat{S}_i \\ t - \hat{T}_i \end{pmatrix}], \quad (5)$$

نوشت، که در آن  $\mathbf{K} : \mathbb{R}^2 \rightarrow \mathbb{R}$  یک تابع هسته دو متغیره و  $\mathbf{H}_{ST}$  ماتریس هموارسازی است. برآوردگر (۴) دارای مشکل ضرب‌های نامتناهی در نزدیکی مرزها است. برای حل این مشکل از روش  $\mathcal{LLPT}$  استفاده می‌شود، که در آن به جای برآوردگر استاندارد هسته در (۵)، یک چندجمله‌ای روی لگاریتم

تابع چگالی داده‌های تبدیل یافته به صورت موضعی برازش داده می‌شود. فرض کنید بردار دلخواه  $(s, t)$  در  $\mathbb{R}^2$  و  $(s', t')$  نزدیک به  $(s, t)$  باشد. لوادر (۱۹۹۶) برآورد لگاریتم درجه دوم درست‌نمایی موضعی را براساس شبه نمونه تبدیل شده  $(\hat{S}_i, \hat{T}_i)_{i=1, \dots, n}$ ، به صورت

$$\begin{aligned} \log f(s', t') &= a_{\gamma,0}(s, t) + a_{\gamma,1}(s, t)(s' - s) + a_{\gamma,2}(s, t)(t' - t) \\ &\quad + a_{\gamma,3}(s, t)(s' - s)^2 + a_{\gamma,4}(s, t)(t' - t)^2 + a_{\gamma,5}(s, t)(s' - s)(t' - t) \\ &\equiv P_{a_\gamma}(s' - s, t' - t), \end{aligned}$$

تعریف کرد، که در آن بردار  $a_\gamma(s, t) \equiv (a_{\gamma,0}(s, t), \dots, a_{\gamma,5}(s, t))$  از حل مساله ماکسیمم درست‌نمایی وزنی به صورت

$$\begin{aligned} \hat{a}_\gamma(s, t) &= \arg \max_{a_\gamma} \left\{ \sum_{i=1}^n \mathbf{K} \left( \mathbf{H}_{ST}^{-\frac{1}{\gamma}} \begin{pmatrix} s - \hat{S}_i \\ t - \hat{T}_i \end{pmatrix} \right) P_{a_\gamma}(\hat{S}_i - s, \hat{T}_i - t) \right. \\ &\quad \left. - n \int_{\mathbb{R}^2} \mathbf{K} \left( \mathbf{H}_{ST}^{-\frac{1}{\gamma}} \begin{pmatrix} s - s' \\ t - t' \end{pmatrix} \right) \exp(P_{a_\gamma}(s' - s, t' - t)) ds' dt' \right\}. \end{aligned}$$

حاصل می‌شود. بنابراین برآورد تابع  $f(s, t)$  برابر با  $\exp\{\hat{a}_\gamma(s, t)\}$  خواهد بود. در نتیجه برآورد تابع چگالی مفصل به روش  $\mathcal{LLPT}$  از درجه دوم برابر

$$\hat{c}_n^{(\mathcal{LLPT})}(u, v) = \frac{\tilde{f}^p(\Phi^{-1}(u), \Phi^{-1}(v))}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))}, \quad (u, v) \in [0, 1]^2. \quad (6)$$

است. دقت برآورد تابع چگالی مفصل با استفاده از (۶)، وابسته به دقت برآورد بهینه پارامتر هموارسازی است که از روش نزدیکترین همسایگی و اعتبارسنجی متقابل کمترین توانهای دوم برای برآورد پارامتر هموارسازی است. برای بررسی سازگاری معیارهای وابستگی پیشنهادی، مقادیر ارببی و واریانس برآوردگر

(۶) مورد نیاز است که به صورت (جینس و همکاران، ۲۰۱۷)

$$\begin{aligned} Bias(\hat{c}_n^{(LLPT)}(u, v)) &= -\frac{b_n^*}{\lambda} \frac{c(u, v)}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} \left\{ \frac{\partial^* g}{\partial x^*} + \frac{\partial^* g}{\partial y^*} + \mathfrak{r} \frac{\partial^* g}{\partial x^* \partial y^*} \right. \\ &\quad \left. + \mathfrak{r} \left( \frac{\partial^* g}{\partial x^*} \frac{\partial g}{\partial x} + \frac{\partial^* g}{\partial y^*} \frac{\partial g}{\partial y} + \frac{\partial^* g}{\partial x^* \partial y^*} \frac{\partial g}{\partial y} + \frac{\partial^* g}{\partial x \partial y^*} \frac{\partial g}{\partial x} \right) \right\} (x, y) + o(b_n^*), \quad (۷) \end{aligned}$$

$$Var(\hat{c}_n^{(LLPT)}(u, v)) = \frac{\mathfrak{u}}{\lambda \pi n b_n^*} \frac{c(u, v)}{\phi(\Phi^{-1}(u))\phi(\Phi^{-1}(v))} + o\left(\frac{1}{n b_n^*}\right). \quad (۸)$$

است، که در آن  $x = \Phi^{-1}(u)$ ،  $g(x, y) = \log c(\Phi(x), \Phi(y)) + \log \phi(x) + \log \phi(y)$  و  $y = \Phi^{-1}(v)$ .

#### ۴ اندازه‌های وابستگی مفصل مبنا

معیارهای متفاوتی برای اندازه‌های واگرایی کولب-لیبلر، جفری و هلینجر برای محاسبه فاصله بین دو تابع چگالی احتمال پیوسته  $f_1$  و  $f_2$  به ترتیب به صورت

$$KLD(f_1|f_2) = \int_{\mathbb{R}} f_1(x) \log \frac{f_1(x)}{f_2(x)} dx, \quad (۹)$$

$$JD(f_1|f_2) = \int_{\mathbb{R}} (f_1(x) - f_2(x)) (\log f_1(x) - \log f_2(x)) dx, \quad (۱۰)$$

$$HD(f_1|f_2) = \int_{\mathbb{R}} (\sqrt{f_1(x)} - \sqrt{f_2(x)})^2 dx. \quad (۱۱)$$

تعریف شده‌اند (کولب و لیبلر، ۱۹۵۱). اندازه‌های واگرایی بین تابع چگالی توام دو متغیر تصادفی پیوسته و حاصلضرب توابع چگالی حاشیه‌ای آن‌ها را می‌توان به عنوان معیارهای وابستگی در نظر گرفت. فرض کنید  $(X, Y)$  یک بردار تصادفی با توزیع‌های حاشیه‌ای  $F_X$  و  $F_Y$  و تابع مفصل متناظر  $C$  باشند.

با توجه به (۲) و (۹)، اندازه وابستگی کولیک-لیبار مفصل مینا توسط **ما و سان (۲۰۱۱)** به صورت

$$KLD(c) \equiv KLD(f|f_X f_Y) = \int_{[0,1]^2} c(u, v) \log(c(u, v)) \, dudv \quad (۱۲)$$

معرفی شد. به طور مشابه، اندازه‌های وابستگی جفری و هلینجر مفصل مینا با استفاده از رابطه (۲) توسط **محتمشی برزادران و امینی (۲۰۱۰)** به صورت

$$JD(c) \equiv JD(f|f_X f_Y) = \int_{[0,1]^2} (c(u, v) - ۱) \log(c(u, v)) \, dudv, \quad (۱۳)$$

$$HD(c) \equiv HD(f|f_X f_Y) = \int_{[0,1]^2} (\sqrt{c(u, v)} - ۱)^2 \, dudv. \quad (۱۴)$$

معرفی شده است.

گزاره ۱. فرض کنید  $X$  و  $Y$  دو متغیر تصادفی با تابع توزیع مفصل  $C$ ، تابع چگالی مفصل  $c$ ، تابع چگالی توام  $f$  و توابع چگالی حاشیه‌ای به ترتیب  $f_X$  و  $f_Y$  باشند. در اینصورت، اندازه‌های وابستگی معرفی شده در (۱۳) و (۱۴) تحت تبدیلات نزولی (صعودی) پایا هستند.

برهان: اگر  $\alpha$  و  $\beta$  دو تبدیل نزولی دلخواه روی متغیرهای تصادفی  $X$  و  $Y$  باشند، با استفاده از قضیه ۲۰۴۰۳ در **نلسن (۲۰۰۷)**، برای تابع مفصل  $C$  روابط  $C_{\alpha(X), Y}(u, v) = v - C_{X, Y}(1 - u, v)$  و  $C_{\alpha(X), \beta(Y)}(u, v) = u + v - ۱ - C_{X, Y}(1 - u, 1 - v)$  برای تابع چگالی مفصل  $c_{\alpha(X), \beta(Y)}(u, v) = c_{X, Y}(1 - u, 1 - v)$  و  $c_{\alpha(X), Y}(u, v) = c_{X, Y}(1 - u, v)$  برقرار است. بنابراین

$$\begin{aligned} JD_{\alpha(X), Y}(c) &= \int_{[0,1]^2} (c_{\alpha(X), Y}(u, v) - ۱) \log(c_{\alpha(X), Y}(u, v)) \, dudv \\ &= \int_{[0,1]^2} (c(z, v) - ۱) \log(c(z, v)) \, dzdv \\ &= JD_{X, Y}(c), \end{aligned}$$



که در آن تساوی دوم با تغییر متغیر  $z = 1 - u$  برقرار است. با استلالی مشابه،

$$\begin{aligned} JD_{\alpha(X),\beta(Y)}(c) &= \int_{[0,1]^2} (c_{\alpha(X),\beta(Y)}(u,v) - 1) \log(c_{\alpha(X),\beta(Y)}(u,v)) dudv \\ &= \int_{[0,1]^2} (c(z,w) - 1) \log(c(z,w)) dzdw \\ &= JD_{X,Y}(c), \end{aligned}$$

که در آن تساوی دوم با تغییر متغیر  $z = 1 - u$  و  $w = 1 - v$  برقرار است. به طور مشابه می‌توان این نتایج را برای اندازه وابستگی هلینجر بر حسب تابع چگالی مفصل بدست آورد. اگر  $\alpha$  و  $\beta$  دو تبدیل صعودی دلخواه روی متغیرهای تصادفی  $X$  و  $Y$  باشند، آنگاه از این‌که  $C_{\alpha(X),\beta(Y)}(u,v) = C_{X,Y}(u,v)$  و  $C_{\alpha(X),\beta(Y)}(u,v) = C_{X,Y}(u,v)$ ، با استلالی مشابه برهان کامل می‌شود.

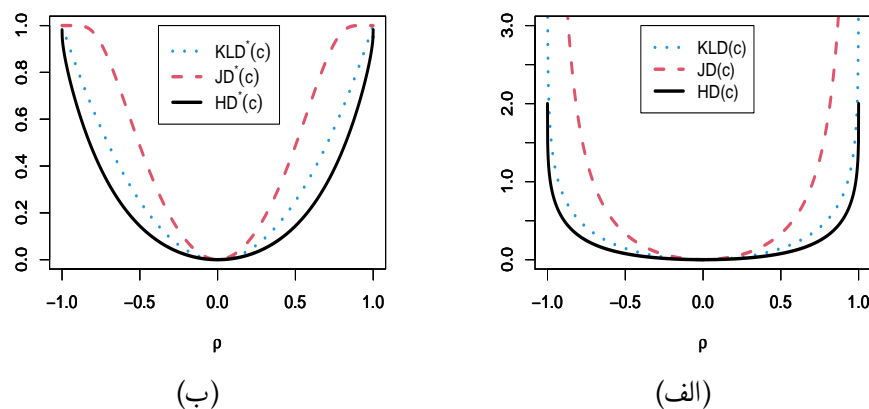
مثال ۱. (میچز و زاگرافوس ، ۲۰۰۶) اگر بردار تصادفی  $(X, Y)$  از مفصل گاوسی دومتغیره با پارامتر وابستگی  $\rho$  باشند، آنگاه

$$KLD(c) = -\frac{1}{2} \log(1 - \rho^2), \quad (15)$$

$$JD(c) = \frac{1}{1 - \rho^2} - 1, \quad (16)$$

$$HD(c) = 2 \left( 1 - 2 \frac{(1 - \rho^2)^{\frac{1}{4}}}{(4 - \rho^2)^{\frac{1}{4}}} \right). \quad (17)$$

این اندازه‌ها به ازای  $\rho = 0$  برابر صفر و برای مقادیر مثبت  $\rho$ ، نسبت به  $\rho$  صعودی هستند. اندازه‌های وابستگی (۱۵)-(۱۷) در بازه  $[0, \infty)$  قرار دارند. با استفاده از تبدیل اکیدا صعودی  $T(x) = 1 - e^{-2x}$  که در آن  $T : [0, \infty) \rightarrow [0, 1]$ ، اندازه‌های وابستگی استاندارد محاسبه می‌شوند. برای نمایش معیار وابستگی استاندارد کولبک-لیبلر، جفری و هلینجر بر حسب تابع مفصل به ترتیب از نمادهای  $KLD^*(c)$ ،  $JD^*(c)$  و  $HD^*(c)$  استفاده می‌شود. رفتار معیارهای وابستگی نسبت به مقادیر مختلف پارامتر  $\rho$  برای مفصل گاوسی در حالت استاندارد و غیراستاندارد در شکل ۱ رسم شده است. همانطور که ملاحظه می‌شود با افزایش شدت وابستگی (افزایش قدر مطلق پارامتر  $\rho$ )، مقادیر اندازه‌های وابستگی (۱۵)-(۱۷) افزایش می‌یابد و کمترین مقدار هر سه اندازه وابستگی برابر صفر است و به ازای  $\rho = 0$



شکل ۱. اندازه‌های وابستگی برای مفصل گاوسی با پارامتر  $\rho$  در حالت الف-غیراستاندارد و ب-استاندارد

حاصل می‌شود. اندازه وابستگی هلینجر تنها انداز وابستگی در بین اندازه‌های وابستگی در شکل ۱ است که وابستگی خطی ( $\rho = \pm 1$ ) را اندازه‌گیری می‌کند. همچنین ترتیب معیارهای وابستگی پس از اعمال تبدیل استاندارد  $T(x)$  تغییر نکرده است و ماکسیم مقدار آن‌ها برابر عدد یک شده است. براساس مطالعه میچیز و زاگرافوس (۲۰۰۶)، انتظار می‌رود اندازه وابستگی هلینجر عملکرد بهتری نسبت به اندازه‌های وابستگی جفری و کولیک-لیبلر داشته باشد، زیرا مقایسه کمتری نسبت به آن‌ها اختیار می‌کند و همچنین مقدار ماکسیم آن در حالت غیراستاندارد متناهی است.

#### ۴.۱ برآورد ناپارامتری اندازه‌های وابستگی

برای نمونه تصادفی  $(U_i, V_i)_{i=1, \dots, n}$ ، برآورد گشتاوری اندازه وابستگی کولیک-لیبلر را با جایگذاری برآورد تابع چگالی مفصل به روش  $\mathcal{LLPT}$  در (۱۲) می‌توان به صورت

$$\widehat{KLD}(c) = \frac{1}{n} \sum_{i=1}^n \log(\hat{c}^{(\mathcal{LLPT})}(\tilde{U}_i, \tilde{V}_i)). \quad (18)$$

ارائه کرد. به طور مشابه، برآوردگرهای گشتاوری اندازه‌های وابستگی جفری و هلینجر با جایگذاری برآورد تابع چگالی مفصل به روش  $\mathcal{LLPT}$  در (۱۳) و (۱۴) به صورت

$$\widehat{JD}(c) = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{1}{\hat{c}(\mathcal{LLPT})(\tilde{U}_i, \tilde{V}_i)}\right) \log(\hat{c}(\mathcal{LLPT})(\tilde{U}_i, \tilde{V}_i)), \quad (19)$$

$$\widehat{HD}(c) = \frac{1}{n} \sum_{i=1}^n \left(1 - \frac{1}{\sqrt{\hat{c}(\mathcal{LLPT})(\tilde{U}_i, \tilde{V}_i)}}\right)^2. \quad (20)$$

معرفی می‌شوند، که در آن  $(\tilde{U}_i, \tilde{V}_i)_{i=1, \dots, n}$  بردار شبه مشاهدات است. خاصیت سازگاری برآورد ناپارامتری آنتروپی توسط احمد و لین (۱۹۷۶) اثبات شده است. اکنون سازگاری مرتبه دوم برآوردگرهای (۱۹) و (۲۰) در گزاره ۲ اثبات می‌گردد.

گزاره ۲. فرض کنید  $(U_i, V_i)_{i=1, \dots, n}$  یک نمونه تصادفی با تابع توزیع مفصل  $C$  و تابع چگالی مفصل  $c$  باشد. اگر  $b_n \rightarrow 0$ ,  $nb_n \rightarrow \infty$  هنگامی که  $n \rightarrow \infty$  و همچنین شرایط  $E(c(U, V)) < \infty$  و  $E(c(U, V)^2) < \infty$  برقرار باشند، آنگاه

الف) هنگامی که  $n \rightarrow \infty$ ,  $E(\widehat{JD}(c) - JD(c))^2 \rightarrow 0$ .

ب) هنگامی که  $n \rightarrow \infty$ ,  $E(\widehat{HD}(c) - HD(c))^2 \rightarrow 0$ .

برهان: الف) فرض کنید  $\hat{c}_n^{(\mathcal{LLPT})} \equiv \hat{c}$  و  $I = [0, 1]$

$$JD(c) = \int_I \left(1 - \frac{1}{c(u, v)}\right) \log(c(u, v)) dC(u, v), \quad (21)$$

$$\widehat{JD}(c) = \int_I \left(1 - \frac{1}{\hat{c}(u, v)}\right) \log(\hat{c}(u, v)) dC_n(u, v), \quad (22)$$

$$\theta(c) = \int_I \left(1 - \frac{1}{c(u, v)}\right) \log(c(u, v)) dC_n(u, v). \quad (23)$$

با استفاده از نامساوی مینکوفسکی می‌توان نوشت،

$$\begin{aligned} E(\widehat{JD}(c) - JD(c))^2 &= E[(\widehat{JD}(c) - \theta(c)) + (\theta(c) - JD(c))]^2 \\ &\leq [E^{1/2}(\widehat{JD}(c) - \theta(c))^2 + E^{1/2}(\theta(c) - JD(c))^2]^2 \\ &= (P_1^{1/2} + P_2^{1/2})^2. \end{aligned}$$

بنابراین برای سازگاری قسمت الف کافیت به ازای  $i = 1, 2$  نشان داده شود،  $P_i \rightarrow 0$  با جایگذاری (۲۲) و (۲۳) در  $P_1$  و با استفاده از بسط تیلور تابع  $g(\hat{c}(u, v)) = (1 - \frac{1}{\hat{c}(u, v)}) \log(\hat{c}(u, v))$  برای مقادیر ثابت  $U = u$  و  $V = v$ ، در نقطه  $c(u, v)$  می‌توان نوشت،

$$\begin{aligned} P_1 &= E \left[ \int_{I^r} \left( \left(1 - \frac{1}{\hat{c}(u, v)}\right) \log(\hat{c}(u, v)) - \left(1 - \frac{1}{c(u, v)}\right) \log(c(u, v)) \right) dC_n(u, v) \right]^2 \\ &= E \left[ \int_{I^r} \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^2(u, v)} \right) (\hat{c}(u, v) - c(u, v)) dC_n(u, v) + o(b_n^r) \right]^2 \\ &= E \left[ \frac{1}{n} \sum_{i=1}^n \left( \frac{c(U_i, V_i) - 1 + \log(c(U_i, V_i))}{c^2(U_i, V_i)} \right) (\hat{c}(U_i, V_i) - c(U_i, V_i)) + o(b_n^r) \right]^2 \\ &= \frac{1}{n^2} E \left[ \sum_{i=1}^n \left( \frac{c(U_i, V_i) - 1 + \log(c(U_i, V_i))}{c^2(U_i, V_i)} \right)^2 (\hat{c}(U_i, V_i) - c(U_i, V_i))^2 \right] \\ &\quad + \frac{1}{n^2} E \left[ \sum_{\substack{i, j=1 \\ i \neq j}}^n \left( \frac{c(U_i, V_i) - 1 + \log(c(U_i, V_i))}{c^2(U_i, V_i)} \right) \left( \frac{c(U_j, V_j) - 1 + \log(c(U_j, V_j))}{c^2(U_j, V_j)} \right) \right. \\ &\quad \left. \times (\hat{c}(U_i, V_i) - c(U_i, V_i)) (\hat{c}(U_j, V_j) - c(U_j, V_j)) \right] + o(b_n^r) \\ &= A_1 + A_2 + o(b_n^r). \end{aligned}$$

اکنون هر یک از عبارات  $A_1$  و  $A_2$  به صورت جداگانه محاسبه می‌شود. با توجه به هم‌توزیعی  $\hat{c}(U_i, V_i)$

برای  $i = 1, \dots, n$  می‌توان نوشت،

$$\begin{aligned} A_1 &= \frac{1}{n} E \left[ \left( \frac{c(U, V) - 1 + \log(c(U, V))}{c^\nu(U, V)} \right)^\nu (\hat{c}(U, V) - c(U, V))^\nu \right] \\ &= \frac{1}{n} E_{(U, V)} \{ E_{\hat{c}} \left[ \left( \frac{c(U, V) - 1 + \log(c(U, V))}{c^\nu(U, V)} \right)^\nu (\hat{c}(U, V) - c(U, V))^\nu | U, V \right] \} \\ &= \frac{1}{n} \int_{I^\nu} E_{\hat{c}} \left[ \left( \frac{c(U, V) - 1 + \log(c(U, V))}{c^\nu(U, V)} \right)^\nu \right. \\ &\quad \left. \times (\hat{c}(U, V) - c(U, V))^\nu | U = u, V = v \right] c(u, v) du dv. \end{aligned}$$

با توجه به (۷) و (۸) برای هر  $(u, v) \in I^\nu$  داریم،

$$\begin{aligned} &E_{\hat{c}} \left[ \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^\nu(u, v)} \right)^\nu (\hat{c}(u, v) - c(u, v))^\nu \right] \\ &= \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^\nu(u, v)} \right)^\nu E_{\hat{c}} (\hat{c}(u, v) - c(u, v))^\nu \\ &= \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^\nu(u, v)} \right)^\nu (Var_{\hat{c}}(\hat{c}(u, v)) + Bias_{\hat{c}}^\nu(\hat{c}(u, v))) \\ &= \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^\nu(u, v)} \right)^\nu \left( O\left(\frac{1}{nb_n^\nu}\right) + O(b_n^\nu) \right) \rightarrow 0, \quad n \rightarrow \infty. \end{aligned}$$

علاوه بر این،

$$\begin{aligned} &E_{\hat{c}} \left[ \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^\nu(u, v)} \right)^\nu (\hat{c}(u, v) - c(u, v))^\nu \right] c(u, v) \\ &\leq \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^\nu(u, v)} \right)^\nu (Var_{\hat{c}}(\hat{c}(u, v)) + E_{\hat{c}}^\nu(\hat{c}(u, v)) + c^\nu(u, v)) c(u, v) \\ &\leq \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^\nu(u, v)} \right)^\nu \left( O\left(\frac{1}{nb_n^\nu}\right) + \nu c^\nu(u, v) + O(b_n^\nu) + c^\nu(u, v) \right) c(u, v) \\ &\rightarrow \nu c^\nu(u, v) \left( \frac{c(u, v) - 1 + \log(c(u, v))}{c^\nu(u, v)} \right)^\nu, \quad n \rightarrow \infty. \end{aligned}$$

بنابراین تابع زیر انتگرال  $A_1$  همگرا به صفر و از بالا کراندار است. در این صورت هنگامی که  $n \rightarrow \infty$

با استفاده از قضیه همگرایی تسلطی<sup>۱</sup>،  $A_1 \rightarrow 0$  را می‌توان نتیجه گرفت. با توجه به نتایج بدست آمده برای  $A_1$  و استفاده از نامساوی کوشی-شوارتز برای  $A_2$  می‌توان نوشت،

$$A_2 \leq \frac{1}{n^2} n(n-1)(\sqrt{A_1})^2 \rightarrow 0, \quad n \rightarrow \infty.$$

بنابراین هنگامی که  $n \rightarrow \infty$ ،  $A_2 \rightarrow 0$  و در نتیجه  $P_1 \rightarrow 0$  برقرار است. از طرفی با توجه به (۲۱) و (۲۳) برای  $P_2$  می‌توان نوشت،

$$\begin{aligned} P_2 &= E\left(\frac{1}{n} \sum_{i=1}^n \left(1 - \frac{1}{c(U_i, V_i)}\right) \log(c(U_i, V_i))\right. \\ &\quad \left. - \int_{\Gamma} \left(1 - \frac{1}{c(u, v)}\right) \log c(u, v) dC(u, v)\right)^2 \\ &= \frac{1}{n} \text{Var}\left(\left(1 - \frac{1}{c(U, V)}\right) \log c(U, V)\right) = O\left(\frac{1}{n}\right). \end{aligned}$$

در نتیجه هنگامی که  $n \rightarrow \infty$ ،  $P_2 \rightarrow 0$  و اثبات کامل می‌شود.

(ب) اثبات این قسمت به روشی مشابه اثبات قسمت الف قابل ارائه است.

## ۵ مطالعه شبیه‌سازی

در این بخش به کمک شبیه‌سازی مونت-کارلو و استفاده از بسته‌های *copula* و *kdecopula* در نرم‌افزار *R*، دقت برآوردگرهای معیارهای وابستگی ارائه شده در بخش ۴.۱ بر اساس معیارهای اریبی و میانگین توان دوم خطا (*MSE*) مورد ارزیابی قرار می‌گیرند. دقت برآوردگرهای کولبک-لیبلر، جفری و هلینجر برای توابع مفصل کلیتون، گامبل-هوگارد، فرانک، گاوسی و تی (با درجات آزادی ۳ و ۱۰) با شدت وابستگی‌های متفاوت ( $\tau = 0.2, 0.4, 0.6, 0.8$ ) و در حجم نمونه‌های مختلف ( $n = 20, 50, 100$ ) با یکدیگر مقایسه می‌شوند. توابع مفصل در نظر گرفته شده، سطوح مختلف وابستگی را شامل می‌شوند.

نتایج ارزیابی دقت برآوردگرهای  $\widehat{JD}$  و  $\widehat{HD}$ ،  $\widehat{KLD}$  بر مبنای ۱۰۰۰ مرتبه تکرار شبیه‌سازی برای مفصل‌های کلیتون، گامبل-هوگارد و فرانک در جدول ۲ و برای مفصل‌های گاوسی و تی با درجات آزادی ۳ و ۱۰ در جدول ۳ ارائه شده است. نتایج این جداول نشان می‌دهد که هر سه برآوردگر سازگار هستند،

<sup>1</sup>Dominated Convergence Theorem

جدول ۰۲. اریبی و  $MSE$  اندازه‌های وابستگی برای مفصل‌های کلیتون، گامبل-هوگارد و فرانک

مفصل	$n$	برآوردگر	اریبی				MSE			
			$\tau = 0.8$	$\tau = 0.6$	$\tau = 0.4$	$\tau = 0.2$	$\tau = 0.8$	$\tau = 0.6$	$\tau = 0.4$	$\tau = 0.2$
کلیتون	۲۰	$\widehat{KLD}$	۰/۱۴۹۵	۰/۱۳۴۸	۰/۱۱۱۲	۰/۰۹۴۷	۰/۳۲۹	۰/۳۸۷	۰/۴۵۸	۰/۴۸۱
		$\widehat{HD}$	۰/۰۵۴۵	۰/۰۳۵۱	۰/۰۴۲۹	۰/۰۸۲۱	۰/۰۷۱	۰/۰۸۴	۰/۱۳۴	۰/۱۶۴
		$\widehat{JD}$	-۰/۰۸۲۳	-۰/۰۳۲۵۵	-۱/۰۵۶۸۱	-۵/۰۵۱۶	۰/۱۷۹	۰/۲۱۲	۱/۰۹۶۲۳	۲/۰۴۶۵۶
کلیتون	۵۰	$\widehat{KLD}$	۷/۰۰۹۹	۰/۰۸۳۱	۰/۰۶۵۵	۰/۰۲۸۱	۰/۱۵۹	۰/۱۹۶	۰/۲۴۹	۰/۲۸۴
		$\widehat{HD}$	۰/۰۴۰۸	۰/۰۳۲۷	۰/۰۳۸۱	۰/۰۱۱۸	۰/۰۳۴	۰/۰۵۵	۰/۱۰۶	۰/۱۶۱
		$\widehat{JD}$	-۰/۰۵۹۰	-۰/۰۲۶۴۰	-۱/۰۳۴۵۷	-۳/۱۶۷۰	۰/۰۸۵۰	۰/۱۰۸۱	۱/۰۸۶۵۳	۲/۱۶۵۹
کلیتون	۱۰۰	$\widehat{KLD}$	۰/۰۵۵۹	۰/۰۴۵۷	۰/۰۳۰۴	-۰/۰۰۲۰	۰/۰۵۵	۰/۰۷۹	۰/۱۱۱	۰/۱۴۴
		$\widehat{HD}$	۰/۰۳۱۹	۰/۰۲۱۰	۰/۰۲۰۳	۰/۰۰۹۱	۰/۰۱۹	۰/۰۴۲	۰/۰۰۹۸	۰/۱۵۶
		$\widehat{JD}$	-۰/۰۴۷۶	-۰/۰۲۲۴۱	-۱/۰۵۸۴	-۱/۰۸۱۶	۰/۰۴۱	۰/۰۹۰۹	۱/۰۵۳۹	۲/۰۶۱۶
گامبل-هوگارد	۲۰	$\widehat{KLD}$	۰/۱۵۴۸	۰/۱۴۴۵	۰/۱۲۶۶	۰/۱۱۷۶	۰/۳۲۷	۰/۴۱۱	۰/۴۷۹	۰/۵۵۲
		$\widehat{HD}$	۰/۰۵۹۵	۰/۰۵۹۳	۰/۰۶۸۲	۰/۱۰۰۶	۰/۰۷۷	۰/۱۲۰	۰/۱۲۹	۰/۴۱۶
		$\widehat{JD}$	-۰/۰۸۱۵	-۰/۱۴۶۴	-۰/۰۸۳۵۳	-۳/۰۴۵۴	۰/۲۱۷	۰/۳۵۲	۰/۷۷۸۲	۱/۰۴۶۳۶
گامبل-هوگارد	۵۰	$\widehat{KLD}$	۰/۱۰۲۳	۰/۰۹۲۱	۰/۰۸۴۴	۰/۰۷۶۲	۰/۱۶۰	۰/۲۱۴	۰/۲۵۷	۰/۳۰۹
		$\widehat{HD}$	۰/۰۴۴۶	۰/۰۵۶۳	۰/۰۴۲۶	۰/۰۶۳۰	۰/۰۳۶	۰/۰۷۸	۰/۱۲۲	۰/۲۶۱
		$\widehat{JD}$	-۰/۰۵۱۴	-۰/۱۴۹۱	-۰/۰۸۰۱۹	-۲/۰۴۹۸۱	۰/۰۸۵	۰/۰۴۶۱	۰/۷۰۵۴	۱/۰۹۴۹
گامبل-هوگارد	۱۰۰	$\widehat{KLD}$	۰/۰۵۷۷	۰/۰۵۰۹	۰/۰۴۳۳	۰/۰۳۰۴	۰/۰۵۶	۰/۰۸۱	۰/۱۰۹	۰/۱۳۵
		$\widehat{HD}$	۰/۰۳۵۱	۰/۰۲۷۰	۰/۰۲۳۱	۰/۰۵۹۲	۰/۰۲۲	۰/۰۸۰	۰/۱۱۳	۰/۱۶۲
		$\widehat{JD}$	-۰/۰۴۰۶	-۰/۱۲۴۲	-۰/۰۶۶۷۹	-۱/۰۴۶۴	۰/۰۴۳	۰/۰۳۲۶	۰/۵۳۱۸	۰/۹۱۵۶
فرانک	۲۰	$\widehat{KLD}$	۰/۱۵۴۶	۰/۱۴۶۰	۰/۱۳۵۱	۰/۱۱۵۴	۰/۳۲۰	۰/۳۸۴	۰/۳۹۹	۰/۴۲۲
		$\widehat{HD}$	۰/۰۶۰۱	۰/۰۶۵۷	۰/۰۸۶۲	۰/۱۰۷۱	۰/۰۷۷	۰/۱۲۹	۰/۱۹۷	۰/۳۳۱
		$\widehat{JD}$	-۰/۰۷۱۳	-۰/۱۰۰۰۶	-۰/۰۷۳۱۶	-۶/۱۵۶۳	۰/۲۰۵	۰/۳۲۶	۰/۵۹۸۵	۱/۱۵۶۴
فرانک	۵۰	$\widehat{KLD}$	۰/۱۰۴۰	۰/۱۰۱۴	۰/۰۹۰۲	۰/۰۸۷۷	۰/۱۵۹	۰/۱۹۱	۰/۲۰۱	۰/۲۹۵
		$\widehat{HD}$	۰/۰۴۳۲	۰/۰۵۳۱	۰/۰۷۵۸	۰/۰۸۰۱	۰/۰۳۵	۰/۰۶۸	۰/۰۹۷	۰/۱۲۵
		$\widehat{JD}$	۰/۰۵۱۸	-۰/۰۹۵۷	-۰/۰۷۰۳۰	-۴/۰۹۶۱	۰/۰۷۹	۰/۰۳۱۰	۰/۵۳۹۷	۰/۹۱۵۶
فرانک	۱۰۰	$\widehat{KLD}$	۰/۰۵۹۴	۰/۰۵۸۴	۰/۰۴۹۶	۰/۰۲۴۲	۰/۰۵۴	۰/۰۷۳	۰/۰۸۴	۰/۰۹۱
		$\widehat{HD}$	۰/۰۳۳۲	۰/۰۴۴۲	۰/۰۶۴۵	۰/۰۷۱۳	۰/۰۱۹	۰/۰۶۹	۰/۰۷۸	۰/۱۰۴
		$\widehat{JD}$	-۰/۰۴۰۲	-۰/۰۸۸۱	-۰/۰۵۷۱۶	-۲/۰۴۸۱	۰/۰۳۸	۰/۰۲۱۴	۰/۴۰۳۰	۰/۸۴۱۶

زیرا برای تمامی مفصل‌های در نظر گرفته شده، با افزایش حجم نمونه مقادیر اریبی و  $MSE$  برای هر سه برآوردگر کاهش و میزان دقت افزایش می‌یابد. با افزایش شدت وابستگی ( $\tau$ )، مقدار میانگین توان دوم خطا افزایش پیدا می‌کند. به طور کلی، در حجم نمونه کم و متوسط ( $n \leq 50$ ) یا در شدت وابستگی ضعیف ( $\tau < 0.5$ )، عملکرد برآوردگر معرفی شده برای اندازه وابستگی هلینجر به مراتب بهتر از دو برآوردگرها دیگر است. همچنین در تمامی حالات در نظر گرفته شده، عملکرد برآوردگر هلینجر بهتر از برآوردگر جفری است. در حجم نمونه‌های ۲۰ و ۵۰، اندازه‌های وابستگی هلینجر به ازای تمامی سطوح شدت وابستگی

جدول ۰۳. اربیی و MSE اندازه‌های وابستگی برای مفصل‌های گاوسی و تی با درجات آزادی ۳ و ۱۰

مفصل	n	برآوردگر	MSE				اربیی			
			$\tau = 0.8$	$\tau = 0.6$	$\tau = 0.4$	$\tau = 0.2$	$\tau = 0.8$	$\tau = 0.6$	$\tau = 0.4$	$\tau = 0.2$
۲۰		$\widehat{KLD}$	۰/۰۵۰۴	۰/۰۴۱۸	۰/۰۳۶۵	۰/۰۳۲۱	۰/۱۲۱۲	۰/۱۳۰۵	۰/۱۳۸۳	۰/۱۵۱۳
		$\widehat{HD}$	۰/۰۴۴۷	۰/۰۲۱۲	۰/۰۰۸۵	۰/۰۰۶۷	۰/۱۱۲۳	۰/۰۴۴۳	۰/۰۴۱۷	۰/۰۵۴۰
		$\widehat{JD}$	۳/۱۲۶۴	۱/۷۴۱۹	۰/۰۶۷۷	۰/۰۲۰۲	-۲/۶۱۵۴	-۱/۳۰۰۳	-۰/۱۹۳۷	-۰/۰۶۲۳
۵۰	گاوسی	$\widehat{KLD}$	۰/۰۲۵۳	۰/۰۲۱۹	۰/۰۱۸۵	۰/۰۱۵۲	۰/۰۷۹۸	۰/۰۸۳۶	۰/۰۹۴۳	۰/۱۰۱۸
		$\widehat{HD}$	۰/۰۱۶۹	۰/۰۰۸۱	۰/۰۰۵۳	۰/۰۰۳۱	۰/۰۰۶۶	۰/۰۰۳۵۵	۰/۰۰۳۸۴	۰/۰۰۴۰۳
		$\widehat{JD}$	۳/۰۵۶۹	۱/۷۰۰۹	۰/۰۵۹۳	۰/۰۰۷۴	-۱/۷۸۴۱	-۱/۲۸۶۴	-۰/۱۷۳۹	-۰/۰۵۰۵
۱۰۰		$\widehat{KLD}$	۰/۰۱۴۲	۰/۰۰۹۲	۰/۰۰۷۲	۰/۰۰۵۲	۰/۰۰۴۱	۰/۰۰۴۶۶	۰/۰۰۵۳۲	۰/۰۰۵۸۴
		$\widehat{HD}$	۰/۰۱۵۲	۰/۰۰۷۷	۰/۰۰۴۷	۰/۰۰۱۷	۰/۰۰۶۱۴	۰/۰۰۱۷۵	۰/۰۰۳۵۵	۰/۰۰۲۱۰
		$\widehat{JD}$	۲/۷۴۶۹	۱/۵۳۲۶	۰/۰۴۸۷	۰/۰۰۳۳	-۱/۱۵۶۸	-۱/۲۲۱۶	-۰/۱۵۹۹	-۰/۰۰۴۰۱
۲۰		$\widehat{KLD}$	۰/۰۶۲۴	۰/۰۵۹۷	۰/۰۴۵۳	۰/۰۳۸۱	۰/۱۲۹۱	۰/۱۴۰۰	۰/۱۴۷۷	۰/۱۵۸۲
		$\widehat{HD}$	۰/۰۲۵۱	۰/۰۲۱۴	۰/۰۱۶۸	۰/۰۱۰۰	۰/۱۱۶۰	۰/۱۰۸۳	۰/۰۸۳۶	۰/۰۶۶۱
		$\widehat{JD}$	۱/۶۱۵۳	۰/۳۵۴۷	۰/۰۵۵۴	۰/۰۲۲۲	-۱/۹۰۳۹	-۰/۵۰۷۷	-۰/۱۱۹۱	-۰/۰۸۷۳
۵۰	تی (ν = ۳)	$\widehat{KLD}$	۰/۰۳۲۸	۰/۰۲۹۰	۰/۰۲۳۳	۰/۰۱۷۷	۰/۰۸۵۸	۰/۰۹۳۳	۰/۰۹۷۹	۰/۱۰۲۳
		$\widehat{HD}$	۰/۰۲۳۶	۰/۰۲۱۲	۰/۰۱۲۸	۰/۰۰۴۴	۰/۰۷۸۳	۰/۰۹۱۸	۰/۰۷۸۸	۰/۰۴۹۱
		$\widehat{JD}$	۱/۴۱۶۲	۰/۲۷۴۸	۰/۰۳۸۲	۰/۰۱۰۶	-۱/۳۹۱۴	-۰/۴۳۶۵	-۰/۰۹۲۲	-۰/۰۵۱۳
۱۰۰		$\widehat{KLD}$	۰/۰۱۵۵	۰/۰۱۲۸	۰/۰۰۸۹	۰/۰۰۶۰	۰/۰۲۷۲	۰/۰۰۴۵۶	۰/۰۰۵۰۸	۰/۰۰۵۳۹
		$\widehat{HD}$	۰/۰۲۰۲	۰/۰۱۹۸	۰/۰۱۰۶	۰/۰۰۲۶	۰/۰۰۶۹۲	۰/۰۰۶۰۱	۰/۰۰۴۲۴	۰/۰۰۲۵۹
		$\widehat{JD}$	۱/۲۶۴۵	۰/۱۸۰۱	۰/۰۲۶۸	۰/۰۰۵۶	-۱/۰۱۶۶	-۰/۲۰۰۸	-۰/۰۵۷۱	-۰/۰۰۳۳۲
۲۰		$\widehat{KLD}$	۰/۰۵۷۱	۰/۰۴۸۶	۰/۰۳۹۲	۰/۰۳۵۲	۰/۱۳۳۷	۰/۱۳۸۳	۹/۰۱۵۰	۰/۱۵۷۷
		$\widehat{HD}$	۰/۰۳۹۹	۰/۰۱۱۸	۰/۰۱۳۶	۰/۰۰۷۳	۰/۱۲۴۱	۰/۰۹۱۴	۰/۰۶۲۶	۰/۰۵۳۵
		$\widehat{JD}$	۳/۰۱۵۳	۰/۶۴۰۸	۰/۰۴۷۸	۰/۰۱۹۱	-۳/۲۳۳۸	-۰/۷۶۳۴	-۰/۱۳۵۵	-۰/۰۰۷۵۷
۵۰	تی (ν = ۱۰)	$\widehat{KLD}$	۰/۰۲۹۱	۰/۰۲۴۹	۰/۰۲۰۶	۰/۰۱۶۵	۰/۰۸۴۷	۰/۰۹۲۴	۰/۱۰۰۵	۰/۱۰۷۵
		$\widehat{HD}$	۰/۰۲۱۶	۰/۰۱۱۰	۰/۰۰۸۵	۰/۰۰۳۹	۰/۰۸۱۷	۰/۰۸۷۰	۰/۰۵۷۸	۰/۰۴۶۲
		$\widehat{JD}$	۲/۱۱۵۶	۰/۵۸۱۷	۰/۰۳۴۷	۰/۰۰۹۱	-۲/۱۵۴۶	-۰/۷۲۷۳	-۰/۱۱۳۳	-۰/۰۰۵۲۷
۱۰۰		$\widehat{KLD}$	۰/۰۱۳۲	۰/۰۱۰۶	۰/۰۰۸۱	۰/۰۰۵۸	۰/۰۴۵۲	۰/۰۵۰۲	۰/۰۵۶۸	۰/۰۶۰۳
		$\widehat{HD}$	۰/۰۴۲۹	۰/۰۲۱۶	۰/۰۰۸۶	۰/۰۰۲۲	۰/۰۷۱۸	۰/۰۸۰۷	۰/۰۴۰۷	۰/۰۲۶۰
		$\widehat{JD}$	۲/۲۱۶۴	۰/۴۳۷۲	۰/۰۲۴۵	۰/۰۰۴۶	-۱/۸۱۶۳	-۰/۵۹۷۲	-۰/۰۸۸۶	-۰/۰۰۳۱۰

برای توابع مفصل در نظر گرفته شده، دقت بیشتری در مقایسه با اندازه وابستگی کولیک-لیبلر و جفری دارد. با افزایش حجم نمونه ( $n \geq 100$ )، دقت برآورد معیار وابستگی جفری به شدت کاهش پیدا می‌کند. در حجم نمونه ۱۰۰ و شدت وابستگی ضعیف ( $\tau < 0.5$ )، اندازه وابستگی هلینجر عملکردهتری نسبت به دو اندازه وابستگی جفری و هلینجر دارد. در حجم نمونه ۱۰۰ با افزایش شدت وابستگی، عملکرد اندازه وابستگی کولیک-لیبلر بهتر از دو معیار وابستگی جفری و هلینجر می‌شود. همچنین در شدت وابستگی خیلی ضعیف ( $\tau = 0.2$ )، اندازه وابستگی جفری عملکرد بهتری نسبت به کولیک-لیبلر دارد.

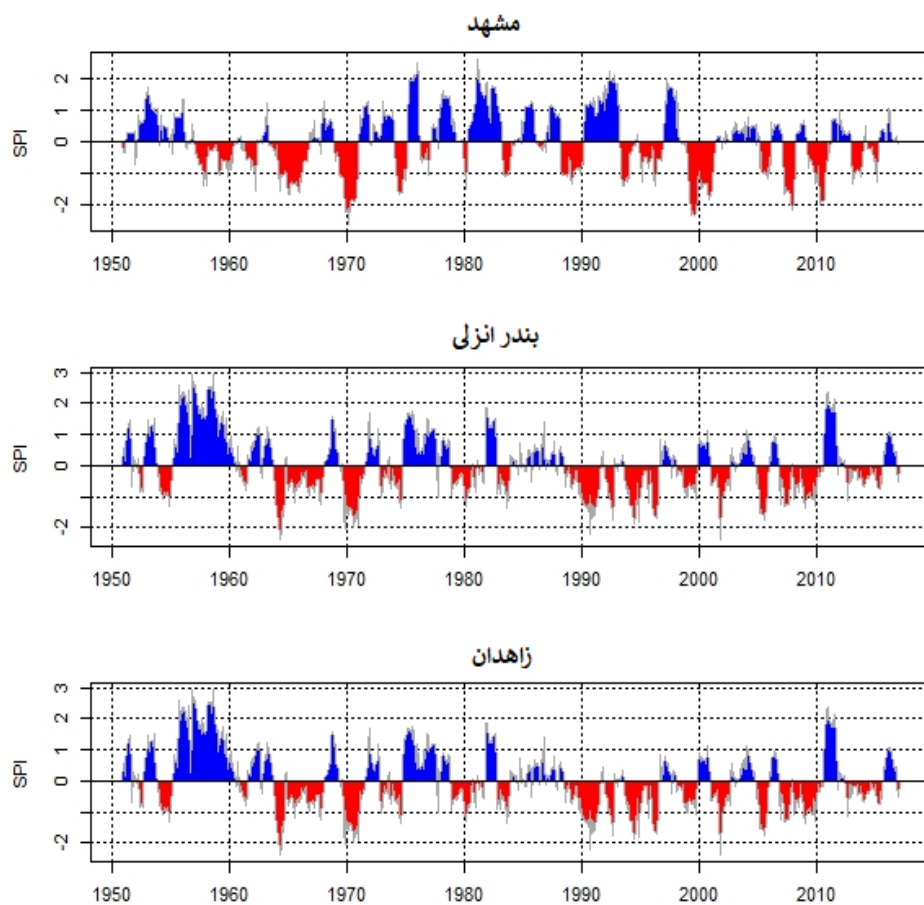


## ۶ کاربرد در هیدرولوژی

تحلیل آماری خشکسالی یکی از نیازهای مهم هوا و اقلیم شناسی، هیدرولوژی و سیستم‌های برنامه‌ریزی و مدیریت منابع آب است. از مشخصه‌های مهم خشکسالی می‌توان به شدت، مدت و طول دوره خشکسالی اشاره کرد که به یکدیگر وابسته هستند و تحلیل آن‌ها با استفاده از روش‌های آماری، باعث درک و شناخت بهتر خشکسالی می‌شود. از جمله منابع مهم در این زمینه می‌توان به شیائو (۲۰۰۶)، جنست و فاوور (۲۰۰۷)، سانگ و سینگ (۲۰۱۰) و چن و جیو (۲۰۱۹) اشاره نمود. همچنین مطالعه‌ای از کاربرد تابع مفصل در تحلیل احتمالاتی شدت-مدت خشکسالی توسط امیدوی و همکاران (۱۳۸۹) برای استان تهران انجام شده است. این بخش به بررسی وابستگی بین متغیرهای خشکسالی در شهرهای مشهد، بندرانزلی و زاهدان با استفاده از معیارهای وابستگی معرفی شده در بخش ۵ می‌پردازد و برای تحلیل خشکسالی از شاخص  $SPI$  استفاده می‌شود. سری شاخص خشکسالی  $SPI$  برابر چندک‌های تابع توزیع نرمال استاندارد در نقاط تابع توزیع تجمعی سری بارش در شهر یا ایستگاه هواشناسی مورد نظر است. با برآورد سری  $SPI$ ، مقدار شدت خشکسالی ( $S_d$ ) به صورت تجمعی از  $S_d = -\sum_{i=1}^{D_d} Index_i$  بدست می‌آید، که در آن  $D_d$  معرف مدت خشکسالی است و به‌صورت مدت زمان وقوع مستمر خشکسالی تعریف می‌شود و همچنین  $L_d$  معرف مدت زمان بین دو خشکسالی و  $Index_i$  برابر مقدار شاخص خشکسالی  $SPI$  در زمان منفی بودن آن است.

نمودار سری  $SPI$  برای داده‌های بارش ماهیانه شهرهای مشهد، بندرانزلی و زاهدان برای بازه زمانی سال ۱۹۵۱ تا سال ۲۰۱۷ در مقیاس ۱۲ ماهه در شکل ۲ رسم شده است. آزمون ناپارامتری من-کندال برای بررسی روند، آزمون گشت برای بررسی تصادفی بودن و نمودارهای خودهمبستگی برای بررسی استقلال روی سری بارش هر شهر به‌کار گرفته شد و نتایج در سطح خطای ۵ درصد مورد تأیید قرار گرفت. هدف از این بخش، بررسی و تحلیل وابستگی بین متغیرهای  $S_d$ ،  $D_d$  و  $L_d$  در هر یک از سه شهر مورد نظر است که به صورت نمونه‌های تصادفی با حجم یکسان از سری زمانی  $SPI$  شهر مربوطه بدست می‌آیند. با توجه به حجم کم مشاهدات برای این متغیرها و نتایج بدست آمده در بخش نتایج عددی، می‌توان ادعا کرد که اندازه‌های وابستگی هلینجر دقت بیشتری در اندازه‌گیری وابستگی بین متغیرها دارند. در جدول ۴ برآورد ضریب وابستگی  $T$ -کندال و معیارهای وابستگی‌های کولبک-لیبلر، جفری و هلینجر در حالت استاندارد گزارش شده است. نتایج نشان می‌دهد که براساس اندازه‌های وابستگی معرفی شده و ضریب وابستگی  $T$ -کندال، وابستگی مثبتی بین شاخص‌های خشکسالی وجود دارد. براساس تمامی معیارهای وابستگی ارائه شده در جدول ۴، در هر سه شهر مشهد، بندرانزلی و زاهدان شدت وابستگی بین ( $S_d$ ,  $D_d$ ) بیشتر از

شدت وابستگی بین سایر متغیرها است.



شکل ۲. سری SPI برای داده‌های بارش ماهیانه شهرهای مشهد، بندرانزلی و زاهدان

## بحث و نتیجه‌گیری

برآورد ناپارامتری اندازه‌های وابستگی جغری و هلینجر بر مبنای تابع چگالی مفصل بر اساس روش تبدیل پروبیت درست‌نمایی موضعی ارائه و سازگاری مرتبه دوم برآوردگرها در حالت مجانبی اثبات شد. نشان داد معیار وابستگی بر مبنای اندازه‌های واگرایی هلینجر، در حجم نمونه‌های کم یا شدت وابستگی ضعیف،

جدول ۰۴. برآورد ضرایب  $\tau$ -کندال و معیارهای وابستگی‌های در حالت استاندارد

شهر	وابستگی	$(S_d, D_d)$	$(S_d, L_d)$	$(D_d, L_d)$
مشهد ( $n = ۳۶$ )	$\hat{\tau}$	۰,۷۷۱۴	۰,۵۲۳۸	۰,۴۳۴۹
	$\widehat{KLD}^*(c)$	۰,۸۴۶۶	۰,۶۲۹۷	۰,۵۵۴۱
	$\widehat{JD}^*(c)$	۰,۷۲۰۱	۰,۵۴۹۲	۰,۴۴۵۵
	$\widehat{HD}^*(c)$	۰,۴۵۹۱	۰,۳۲۲۰	۰,۲۵۰۵
بندرانزلی ( $n = ۴۴$ )	$\hat{\tau}$	۰,۶۸۷۱	۰,۵۹۴۱	۰,۵۳۴۸
	$\widehat{KLD}^*(c)$	۰,۸۲۲۰	۰,۶۷۱۲	۰,۷۶۹۱
	$\widehat{JD}^*(c)$	۰,۷۰۲۷	۰,۵۸۳۲	۰,۶۵۵۷
	$\widehat{HD}^*(c)$	۰,۴۴۴۲	۰,۳۴۸۲	۰,۴۰۴۱
زاهدان ( $n = ۳۳$ )	$\hat{\tau}$	۰,۸۱۸۲	۰,۵۲۲۷	۰,۴۹۲۴
	$\widehat{KLD}^*(c)$	۰,۹۰۸۳	۰,۶۴۰۹	۰,۶۲۱۸
	$\widehat{JD}^*(c)$	۰,۸۵۷۱	۰,۶۱۵۴	۰,۵۸۴۸
	$\widehat{HD}^*(c)$	۰,۵۹۶۲	۰,۳۶۵۷	۰,۳۴۳۱

بسیار دقیق‌تر از اندازه وابستگی کولیک-لیبلر عمل می‌کند. همچنین در شدت‌های وابستگی مختلف و حجم نمونه‌های متفاوت، دقت برآورد اندازه وابستگی هلینجر بیشتر از اندازه وابستگی جفری است. تحلیل شاخص‌های خشکسالی نشان داد که با توجه به حجم کم مشاهدات برای متغیرهای شدت، مدت و طول دوره خشکسالی، استفاده از اندازه وابستگی هلینجر اعتبار بیشتری دارد. با توجه به حجم کم مشاهدات برای متغیرهای مستخرج از شاخص خشکسالی  $SPI$  و نتایج بخش عددی، استفاده از معیار وابستگی هلینجر برای تحلیل وابستگی بین متغیرهای شدت، مدت و طول دوره خشکسالی توصیه می‌شود.

## تقدیر و تشکر

نویسندگان از پیشنهادات ارزنده داوران گرامی، هیئت تحریریه و ویراستار محترم مجله که باعث بهبود و ارائه بهتر این مقاله گردید، کمال تشکر و قدردانی را دارند.

## مراجع

- امیدی، م.، محمدزاده، م. و مرید، س. (۱۳۸۹)، تحلیل احتمالاتی شدت-مدت خشکسالی در استان تهران با استفاده از توابع مفصل، *مجله تحقیقات آب و خاک ایران*، ۴۱، ۹۵-۱۰۱.
- خسروی، س.، امینی، م. و محتشمی برزادران، غ. (۱۳۹۱)، مقایسه عددی برخی از اندازه های فی-واگرا برای مفصل های فارلی-گامبل-مورگنسترن تعمیم یافته، *مجله علوم آماری*، ۶، ۵۵-۶۷.
- Ahmad, I., and Lin, P. E. (1976), A Nonparametric Estimation of the Entropy for Absolutely Continuous Distributions, *IEEE Transactions on Information Theory*, **22**, 372-375.
- Belalia, M., Bouezmarni, T., Lemyre, F. C., and Taamouti, A. (2017), Testing Independence Based on Bernstein Empirical Copula and Copula Density, *Journal of Nonparametric Statistics*, **29**, 346-380.
- Blum, J. R., Kiefer, J., and Rosenblatt, M. (1961), Distribution Free Tests of Independence Based on the Sample Distribution Function, *The Annals of Mathematical Statistics*, **32**, 485-498.
- Blumentritt, T., and Schmid, F. (2012), Mutual Information as A Measure of Multivariate Association: Analytical Properties and Statistical Estimation, *Journal of Statistical Computation and Simulation*, **82**, 1257-1274.
- Bouezmarni, T., El Gouch, A., and Taamouti, A. (2013), Bernstein Estimator for Unbounded Copula Densities, *Statistics and Risk Modeling*, **30**, 343-360.
- Charpentier, A., Fermanian, J., and Scaillet, O. (2006), Copulas: From theory to application in Finance, Chapter The Estimation of Copulas: Theory and Practice, *Risk Books*, Torquay, UK, **1**, 35-62.

- Chen, L., and Guo, S. (2019), *Copulas and Its Application in Hydrology and Water Resources*, Springer.
- Geenens, G., Charpentier, A., and Paindaveine, D. (2017), Probit Transformation for Nonparametric Kernel Estimation of the Copula Density, *Bernoulli*, **23**, 1848-1873.
- Genest, C., and Favre, A. C. (2007), Everything You Always Wanted to Know About Copula Modeling but Were Afraid to Ask, *Journal of Hydrologic Engineering*, **12**, 347-368.
- Genest, C., Masiello, E., and Tribouley, K. (2009), Estimating Copula Densities Through Wavelets, *Insurance: Mathematics and Economics*, **44**, 170-181.
- Genest, C., and Remillard, B. (2004), Test of Independence and Randomness Based on the Empirical Copula Process, *Test*, **13**, 335-369.
- Joe, H. (1989), Relative Entropy Measures of Multivariate Dependence, *Journal of the American Statistical Association*, **84**, 157-164.
- Joe, H. (2014), *Dependence Modeling with Copulas*, Chapman and Hall/CRC.
- Kauermann, G., Schellhase, C., & Ruppert, D. (2013), Flexible Copula Density Estimation with Penalized Hierarchical B-Splines, *Scandinavian Journal of Statistics*, **40**, 685-705.
- Kullback, S., and Leibler, R. A. (1951), On Information and Sufficiency, *The Annals of Mathematical Statistics*, **22**, 79-86.
- Loader, C. R. (1996), Local Likelihood Density Estimation, *The Annals of Statistics*, **24**, 1602-1618.

- Ma, J., and Sun, Z. (2011), Mutual Information is Copula Entropy, *Tsinghua Science & Technology*, **16**, 51-54.
- Micheas, A. C., and Zografos, K. (2006), Measuring Stochastic Dependence Using  $\Phi$ -Divergence, *Journal of Multivariate Analysis*, **97**, 765-784.
- Mohtashami Borzadaran, G. R. and Amini, M. (2010), Information Measures via Copula Functions, *Journal of Statistical Research of Iran*, **7**, 47-60.
- Nagler T (2014), *Kernel Methods for Vine Copula Estimation*, Master's Thesis, Technische Universität München.
- Nagler, T. (2018), kdecopula: An R Package for the Kernel Estimation of Bivariate Copula Densities, *Journal of Statistical Software*, **84**, 1-22.
- Nelsen, R. B. (2007), *An Introduction to Copulas*, Springer Science and Business Media.
- Shiau, J. T. (2006), Fitting Drought Duration and Severity with Two-dimensional Copulas, *Water Resources Management*, **20**, 795-815.
- Sklar, M. (1959), Fonctions de Repartition an Dimensions et Leurs Marges, *Publications de l'Institut de Statistique de l'Université de Paris*, **8**, 229-231.
- Song, S., and Singh, V. P. (2010), Meta-Elliptical Copulas for Drought Frequency Analysis of Periodic Hydrologic Data, *Stochastic Environmental Research and Risk Assessment*, **24**, 425-444.
- Wen K and Wu X (2015), Transformation-Kernel Estimation of the Copula Density, *Working Paper*, <http://agecon2.tamu.edu/people/faculty/wu-ximing/agecon2/public/copula.pdf>.

Journal of Statistical Sciences, Spring and Summer, 2021  
Vol. 15, No. 1, pp 233-254  
DOI: 10.29252/jss.15.1.233

## **Bivariate Dependency Analysis using Jeffrey and Hellinger Divergence Measures based on Copula Density Estimation by Improved Probit Transformation**

Mohammadi, M., Emadi, M., Amini, M.  
Department of Statistics, Ferdowsi University of Mashhad, Mashhad, Iran.

**Abstract:** Divergence measures can be considered as criteria for analyzing the dependency and can be rewritten based on the copula density function. In this paper, Jeffrey and Hellinger dependency criteria are estimated using the improved probit transformation method, and their asymptotic consistency is proved. In addition, a simulation study is performed to measure the accuracy of the estimators. The simulation results show that for low sample size or weak dependence, the Hellinger dependency criterion performs better than Kullback-Libeler and Jeffrey dependency criteria. Finally, the application of the studied methods in hydrology is presented.

**Keywords:** Divergence, dependency, copula density, probit transformation.

**Mathematics Subject Classification (2010):** 62G10, 62G07, 62H15.