

بهبود کارایی شبیه‌سازی مونت کارلو برای تولید داده‌های همبسته: کاربرد ی در تغییر اقلیم

بیژن قهرمان^{۱*}، علی قهرمان^۲

چکیده

از شبیه‌سازی مونت کارلو به خوبی برای تولید گروه داده‌هایی که از توزیعی مشخص پیروی می‌کنند استفاده می‌شود. با این حال، تولید داده‌های همبسته تنها برای توزیعهای خاص (بهنجار دو-متغیره) امکان‌پذیر است، در حالی که در پاره‌ای از مواقع، همچون تحلیل فراوانی منطقه‌ای، پرداختن به توزیعهای غیربهنجار مورد نیاز است. برای حل مسأله‌ی مزبور، در این مقاله استفاده از الگوریتم ژنتیک به عنوان روش ابتکاری پیشنهاد گردیده است. برای تطابق الگوریتم ژنتیک با مسأله‌ی در دست بررسی، به جای تلاقی دو کروموزوم از دست کاری در هر کروموزوم استفاده شد، زیرا قرارگیری مقادیر در یک گروه زمانی به گونه‌ای است که تکرار اعداد مجاز نبوده، و تمامی اعداد باید در یک گروه مورد استفاده قرار گیرند. این شیوه برای یک متغیر تصادفی انجام گرفت و نشان داده شد که الگوریتم ژنتیک منجر به پاسخهای چندگانه می‌گردد، در حالی که در روش مونت کارلو، تنها پاسخی یکتا به دست می‌آید. با ارائه‌ی تابع هدف مناسب، روش پیشنهادی برای توزیعهای دو-متغیره نیز به کار رفت و شمول گسترده‌تر پاسخها در مقایسه با روش مونت کارلو (تنها برای توزیعهای بهنجار) ارایه شد. دلیل گستردگی پاسخ در روش پیشنهادی به این حقیقت باز می‌گردد که یک گروه زمانی را می‌توان به چندین شیوه‌ی گوناگون مرتب کرد که بتواند با یک گروه زمانی ثابت ضریب همبستگی یکسان را تولید کند. نشان داده شد که هرچه ضریب همبستگی کمتر باشد، شمار حالت‌های گوناگون افزایش می‌یابد.

واژه‌های کلیدی: بهینه‌سازی، توزیع غیربهنجار، توزیع یکنواخت، روشهای فراکاوشی.

^۱- استاد، گروه، مهندسی آب، دانشکده کشاورزی، دانشگاه فردوسی مشهد

^۲- دانشجوی مهندسی برق و کامپیوتر، دانشکده مهندسی، دانشگاه مانی‌توبا، وینیپگ، کانادا

*- نویسنده مسوول مقاله: bijangh@um.ac.ir

مقدمه

شبیه‌سازی مونت‌کارلو روشی شناخته شده است که از آن برای تولید داده‌هایی که از توزیع خاص با فراسنجهای معلوم پیروی می‌کنند استفاده می‌شود. در این روش، بر مبنای توزیع یکنواخت، معیار داده‌هایی بین صفر و یک تولید و با استفاده از رابطه‌ی چندک (یعنی $x(F)$) به فضای توزیع مورد نظر انتقال داده می‌شود. در پاره‌ای از مواقع (مثلاً در تحلیل فراوانی ناحیه‌ای) لازم است که چندین گروه همبسته تولید شوند. هاسکینگ و والیس (۱۹۹۷) روشی را پایه نهادند که در بخشی از آن باید داده‌ها (مثلاً سیلاب) برای چندین ایستگاه در ناحیه‌ای همگن به وسیله‌ی توزیع چهار فراسنجی کاپا (۱۹۹۴) با کاربرد روش مونت‌کارلو شبیه‌سازی شود. در ۵۰۰ بار شبیه‌سازی، میانگین آماره‌ی همگنی H با آنچه مربوط به داده‌های واقعی (اندازه‌گیری شده/مشاهداتی) است مقایسه و بر آن مبنای همگنی/غیرهمگنی ناحیه تشخیص داده می‌شود. در این روش چنانچه به همبسته بودن احتمالی ایستگاهها توجه نشود، آماره‌ی H منفی می‌شود (گروه شبیه‌سازی شده تغییرپذیری بیشتری دارد). کاسترلارین و همکاران (۲۰۰۸) با توجه به میانگین ضریب همبستگی بین آماره‌های سیلاب در ناحیه‌ی مورد نظر، رابطه‌ای را برپایه‌ی شبیه‌سازی مونت‌کارلو برای تصحیح آماره‌ی H روش هاسکینگ و والیس (۱۹۹۷) ارائه نمودند. نامبردگان تنها میانگین ضریب همبستگی را در نظر گرفته و متغیر بودن ضریب همبستگی (مثلاً وابستگی آن به فاصله‌ی مکانی ایستگاهها) را نادیده انگاشتند. هاسکینگ و والیس (۱۹۹۷) روش ساده‌ای را برای شبیه‌سازی داده‌های همبسته ارائه کرده‌اند که در آن نیز تنها به متوسط ضریب همبستگی در کل ناحیه و برپایه‌ی توزیع احتمال مشترک بهنجار پرداخته شده است. کاملاً آشکار است که برخی داده‌های محیطی (شامل سیلاب) به‌ندرت از توزیع بهنجار پیروی می‌کنند. استخراج توابع ریاضی توزیع مشترک دو و چندمتغیره برای توزیعهای غیربهنجار، توزیع شرطی آنها و امید ریاضی و انحراف معیار شرطی آنها دشوار است و راه‌حل ساده‌ای برای آن وجود ندارد؛ بنابراین، روش ساده‌ی هاسکینگ و والیس (۱۹۹۷) برای تولید داده‌های همبسته با دوگونه محدودیت روبه‌رو است؛ در نتیجه، یک

خلاء علمی و کاربردی برای شبیه‌سازی داده‌های همبسته‌ی غیربهنجار وجود دارد.

از دیرباز نشان داده شده است که روشهای بهینه‌سازی فراکوشی جایگزینی مناسب برای روشهای بهینه‌سازی سنتی به‌شمار می‌آیند. این روشها عمدتاً از طبیعت الهام گرفته و دامنه‌ی فضای متغیرها را به‌گونه‌ای تصادفی و در عین حال هدفمند جستجو می‌کنند تا تابع هدفی را بهینه ساخته و بویژه، برای توابع غیرخطی و پیچیده سودمندند. الگوریتم ژنتیک یکی از این روشهاست که کارایی آن بارها در زمینه‌های گوناگون دانش به وسیله‌ی پژوهشگران نشان داده شده است. با این حال برحسب شرایط مساله، امکان دست‌کاری در الگوریتم سنتی آن وجود دارد.

هدف این پژوهش، بهبود کارایی شبیه‌سازی مونت‌کارلو برای تولید داده‌های همبسته با استفاده از الگوریتم ژنتیک است.

روش پژوهش

شبیه‌سازی مونت‌کارلو تک - متغیره

اگر CDF متغیر تصادفی X ، $F_X(x)$ باشد، آن‌گاه مقدار متغیر تصادفی X متناسب با احتمال تجمعی معلومی همچون $F_X(x)=u$ برابر خواهد بود با:

$$x = F_X^{-1}(u) \quad (1)$$

چنانچه u یافته‌ای از متغیر یکنواخت معیار U با CDF یکنواخت بین صفر و یک باشد، آن‌گاه داریم:

$$F_U(x) = u \quad (2)$$

یعنی احتمال تجمعی $U \leq u$ برابر با u است؛ بنابراین، اگر u یافته‌ای از U باشد، احتمال تجمعی متغیر تصادفی X برابر است با (۱۹۸۴):

$$P(X \leq x) = P[F_X^{-1}(u) \leq x] = P[U \leq F_X(x)] = F_U[F_X(x)] = F_X(x) \quad (3)$$

یعنی اگر $\{u_1, u_2, \dots, u_n\}$ مجموعه‌ای از یافته‌هایی از متغیر تصادفی U باشد، مجموعه‌ی متناسب مقادیر (یافته‌ها) از متغیر تصادفی X که CDF آنها $F_X(x)$ می‌شود:

$$x_i = F_X^{-1}(u_i), \quad i = 1, \dots, n \quad (4)$$

$$\rho = \frac{Cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sigma_X \sigma_Y} \quad (10)$$

امید ریاضی و پراش متغیر تصادفی Y مشروط بر اتخاذ مقداری مشخص برای متغیر تصادفی X (یعنی $X=x$)، به ترتیب با روابط زیر ارائه گردند (۱۹۸۴):

$$E(Y|X=x) = \mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X) \quad (11)$$

$$Var(Y|X=x) = \sigma_Y^2(1 - \rho^2) \quad (12)$$

در ابتدا، همچون بخش ۱ (روابط ۱ تا ۴)، یافته‌ای برای متغیر تصادفی X تولید می‌شود، سپس از روابط ۱۱ و ۱۲ امید ریاضی و پراش شرطی Y محاسبه گردیده، و پس از تولید عدد تصادفی بر مبنای توزیع یکنواخت معیار، همچون روابط ۱ تا ۴، مقداری برای Y مشروط بر x تولید می‌شود.

بهبود روش مونت کارلو برای تولید متغیرهای همبسته با هر توزیع دلخواه

فرض کنید که توزیع و فراسنجهای توزیع دو متغیر تصادفی X و Y که ضریب همبستگی بین آنها ρ است معلوم باشد. همچون بخش ۱، گروه اعداد تصادفی یکنواخت معیار تولید و آنها را جداگانه در فضای دو متغیر تصادفی تحت مطالعه (مستقل از یکدیگر) ببرید. بسیار نادر است که ضریب همبستگی این دو گروه برابر با ρ باشد؛ بنابراین، ساختار همبستگی بین دو گروه را با ثابت نگاه داشتن یک گروه و مرتب کردن هدفمند ترتیب اعداد در گروه دیگر (یک گروه به طول n با n! حالت مجزا از هم مرتب می‌گردد) تأمین می‌کنیم. بنابراین، تابع هدف برای مسأله‌ی بهینه‌سازی به صورت زیر تعریف می‌شود:

$$\min: |\rho_{x,y} - \rho| \quad (13)$$

که در آن ρ مقدار دلخواه ضریب همبستگی بین متغیرهای تصادفی X و Y، و $\rho_{x,y}$ ضریب همبستگی بین دو گروه X و گروه در حال تغییر y (رابطه‌ی ۱۰) می‌باشد. به دلیل شمار زیاد حالات ممکن برای مرتب‌سازی گروه دوم (n معمولاً بزرگ است)، روشهای بهینه‌سازی سنتی را به دشواری می‌توان سودمند دانست؛ بنابراین، روشهای فراکاوشی سودمندترند. در این جا تأکید بر گونه‌ی روش فراکاوشی نداشته و به تصادف روش الگوریتم ژنتیک را برگزیدیم. با این حال ساختار الگوریتم ژنتیک سنتی با این

شبیه‌سازی اعداد با توزیع تصادفی توأم

X_1 تا X_n را مجموعه‌ای از n متغیر تصادفی در نظر می‌گیریم. CDF توأم این متغیرهای تصادفی این گونه نوشته می‌شود:

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = (F_{X_1}(x_1))(F_{X_2}(x_2|x_1)) \dots (F_{X_n}(x_n|x_1, \dots, x_{n-1})) \quad (5)$$

که در آن $F_{X_1}(x_1)$ و $F_{X_k}(x_k|x_1, \dots, x_{k-1})$ به ترتیب CDF حاشیه‌ای X_1 و CDF شرطی X_k است. چنانچه متغیرها همبسته باشند، نمی‌توان آنها را مستقل از هم شبیه‌سازی کرد. در چنین شرایطی، مطابق زیر اقدام می‌گردد: مجموعه‌ای از اعداد تصادفی با توزیع یکنواخت معیار تولید می‌شود (u_1, \dots, u_n). آن‌گاه مشابه با رابطه‌ی ۱، X_1 بطور مستقل تعیین می‌گردد:

$$x_1 = F_{X_1}^{-1}(u_1) \quad (6)$$

سپس X_2 با استفاده از CDF شرطی $F_{X_2}(x_2|x_1)$ محاسبه می‌شود:

$$x_2 = F_{X_2}^{-1}(u_2|x_1) \quad (7)$$

بطور مشابه، مقداری (یافته‌ای) برای X_n با استفاده از مقادیر X_1 تا X_{n-1} که پیش از این رابطه حاصل گردیده‌اند، به دست می‌آید:

$$x_n = F_{X_n}^{-1}(u_n|x_1, \dots, x_{n-1}) \quad (8)$$

کاربرد این روش، چنانچه توابع حاشیه‌ای و شرطی در دست بوده و وارون آنها را بتوان محاسبه، کرد آسان است؛ با این حال چنین چیزی همواره صادق نیست.

شبیه‌سازی مونت کارلو دو-متغیره‌ی بهنجار

pdf توأم دو-متغیره‌ی بهنجار X و Y عبارت است از (۲۰۰۷):

$$f_{x,y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}\right] \quad (9)$$

$-\infty < x < \infty, -\infty < y < \infty$

که در آن ρ ضریب همبستگی بین X و Y (رابطه‌ی ۱۰)، Cov و E به ترتیب نماد کوواریانس و امید ریاضی است، μ_X و σ_X به ترتیب میانگین و انحراف معیار متغیر تصادفی X و μ_Y و σ_Y مقادیر متناظر متغیر تصادفی Y می‌باشند:

نظام‌مندتر است؛ لذا، بیشتر در معرض تخریب قرار می‌گیرد (کلیدون، ۲۰۱۰؛ رودریگوئز-ایتورب و همکاران، ۱۹۹۲)؛ بنابراین، منطقی‌اً احتمال وقوع همبستگی‌های کمتر (عدم نظام) بایستی بیشتر باشد^۱. از مقادیر انحراف معیار نیز می‌توان ناپایداری سامانه را تحلیل کرد. متناظر با افزایش t ، هم انحراف معیار، و هم مقدار احتمال وقوع کاهش می‌یابد، ولی شیب کاهش انحراف معیار کمتر است (۶ در مقابل ۱۲) بنابراین، منجر به ناپایداری در سامانه می‌شود؛ در نتیجه، ضریب تغییرات نیز با افزایش قدرمطلق ضریب همبستگی بطور نمایی افزایش می‌یابد ($t=+0/842$).

شمار به هم ریختن تصادفی گروه دوم (y) را ۱۰۰۰ انتخاب کردیم. آشکار است که گروه دوم به ۳۰! حالت می‌تواند مرتب شود. عدد ۱۰۰۰ را صرفاً برای «عملی بودن» زمان اجرای رایانه برگزیدیم؛ با این حال، برای اعتماد بیشتر، شمار به هم ریختن گروه دوم را از ۱۰۰۰ به ۵۰۰۰ افزایش دادیم. تغییر نتایج معنی‌دار نبوده و نوسان در نتایج، در محدوده‌ی نوسانهای تصادفی ناشی از اجراهای متوالی بود. برای اطمینان بیشتر، تمامی ۱۲۰ حالت محتمل را برای نمونه‌ای به طول ۵ آزمودیم. شکل ۱ یکی از حالات ممکن توزیع فراوانی رخداد ضریب همبستگی را نشان می‌دهد. نتیجه‌ی پیشین تأیید می‌شود: رخدادن همبستگی‌های قوی کمتر از رخدادن همبستگی‌های ضعیف است. به‌جای توزیع یکنواخت، فرایند را برای اعداد تصادفی با توزیع بهنجار معیار $N(0,1)$ تکرار کردن و دریافتیم که نتایج به نوع توزیع بستگی ندارد.

برای بررسی حساسیت نتایج به طول گروه، گروه ۱۰ تایی را به‌جای گروه ۳۰ تایی تولید کردیم. در این حالت، احتمال وقوع همبستگی‌های کمتر به نفع همبستگی‌های بیشتر کاهش یافت (احتمال وقوع برای $0 < |r| < 0.1$ از ۴۰٪ به ۲۱/۸٪ رسید)؛ با این حال روند کاهش نمایی - گرچه با شیبی ملایمتر - همچنان برقرار بود ($t=-0/837$). احتمال وقوع همبستگی بیشتر از ۰/۸ (یعنی $P(|r| > 0.8)$) از صفر به ۰/۶٪ افزایش یافت. تفاوت با حالت پیش را این‌گونه تفسیر می‌کنیم: هرچه طول گروه

مساله‌ی بهینه‌سازی سازگاری ندارد. نخست آن که این بهینه‌سازی کمیته‌سازی است. دیگر آن که اعداد منتسب شده به هر ردیف متناظر با متغیر اول، گسسته اند؛ اعداد تنها می‌توانند با ردیف‌های دیگر تعویض شوند؛ بنابراین تلاقی بین دو کروموزوم در این‌جا درست نیست. به بیان دیگر، π تا عدد ثابت باید با هم جابه‌جا گردند. تلاقی بین دو کروموزوم در الگوریتم ژنتیک معیار را با این ترفند تغییر می‌دهیم: بخشی از هر کروموزوم را ثابت نگاه داشته و دو بخش دیگر آن را (که باید طول مساوی داشته باشند) با هم تعویض می‌کنیم. طول این بخشها را تصادفی، و مکان بخش ثابت کروموزوم را نیز تصادفی برمی‌گزینیم. جمعیت اولیه را ۵۰ و احتمال تعویض داخلی در هر کروموزوم را ۰/۷ برگزیدیم. مراحل چگونگی انتخاب کروموزومها برای نسل بعد و نخبه‌گرایی را مطابق روش معیار نگاه داشتیم.

نتایج و بحث

شبیه‌سازی دو گروه همبسته

دو گروه تصادفی (x و y) به طول ۳۰ و با توزیع یکنواخت تولید شدند. یکی از گروه‌ها را (x) ثابت نگاه داشتیم و گروه دوم (y) ۱۰۰۰ بار بطور تصادفی به هم ریخته شد. هر بار ضریب همبستگی بین این دو گروه محاسبه شد (رابطه ۱۰). فراوانیهای همبستگیهای مختلف در حول $r=0$ قرینه بودند؛ بنابراین، از قدرمطلق r استفاده شد. احتمال واقع شدن قدرمطلق همبستگی در بازه‌هایی به عرض ۰/۱ (مثلاً بین ۰/۲ تا ۰/۳) از روی فراوانی نسبی وقوع محاسبه شد. این فرایند ۱۰۰ مرتبه بطور تصادفی تکرار، سپس میانگین، انحراف معیار و ضریب تغییرات احتمال واقع شدن قدرمطلق ضریب همبستگی در بازه‌های ۰/۱ تایی به‌دست آمد.

نتایج نشان دادند که احتمال رخدادن همبستگی‌های قوی - مثلاً $P(|r| > 0.8)$ - در تمامی این ۱۰۰ تکرار، صفر بود. بیشترین احتمال وقوع (۰/۴) مربوط به بازه‌ی $0 < |r| < 0.1$ بود، و احتمال وقوع تقریباً بطور نمایی با افزایش $|r|$ کاهش می‌یافت ($t=-0/884$). ما احتمال کمتر برای رخدادن همبستگی‌های قویتر را به نظریه‌ی آنتروپی متناسب می‌کنیم: هرچه همبستگی قویتر باشد، سامانه

^۱ - هرکه بالاتر رود ابله‌تر است، کاستخوان او بتر خواهد شکست؛ مولوی؛ مثنوی معنوی، دفتر چهارم

U_y بر مبنای توزیع یکنواخت معیار، و در نتیجه داده‌های متناظر آنها بر مبنای توزیع بهنجار معیار (یعنی N_x و N_y) تصادفی است؛ در نتیجه، در هر بار اجرای برنامه تغییر می‌کنند. دوم این‌که ساختار به‌ریختن گروه N_y برای رسیدن به همبستگی مورد نظر در روش ژنتیک تصادفی است، در نتیجه، در اجراهای متوالی متغیر است. برای داشتن مبنایی قابل تکیه، گروه‌های U_x و U_y (و در نتیجه گروه‌های N_x و N_y) را ثابت در نظر گرفته؛ و روال الگوریتم ژنتیک را ۱۰۰۰ بار تکرار کردیم تا افزون بر ضریب همبستگی بین دو گروه $N_{y,M}^p$ و $N_{y,G}^p$ ، انحراف معیار آن نیز قابل بررسی باشد. جدول ۱ نتیجه‌ی به‌دست آمده را نشان می‌دهد. بر اساس این جدول، نتایج به‌دست آمده از دو روش متفاوت (شبیه‌سازی مونت کارلو و الگوریتم ژنتیک) برای تولید داده‌های همبسته یکسان نمی‌باشند؛ بنابراین، آیا می‌توان نتیجه‌گیری کرد که هدف این پژوهش برآورده نشده است؟

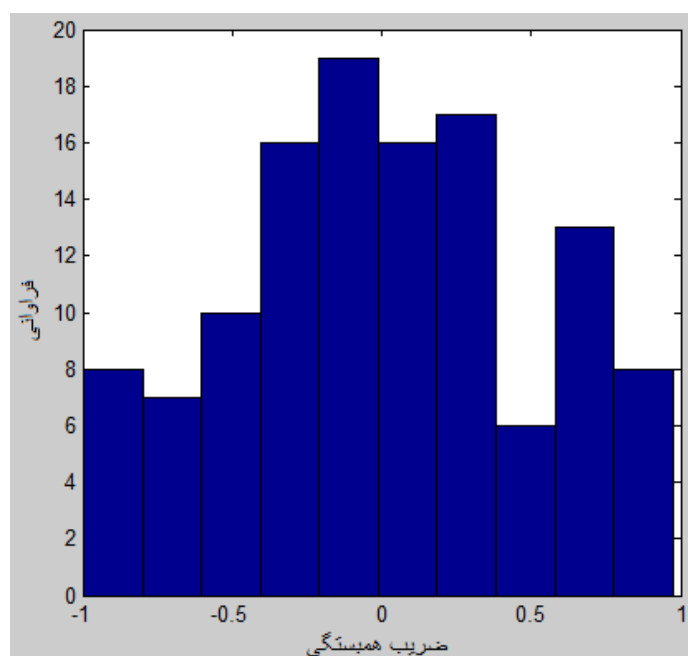
در روش شبیه‌سازی مونت کارلو بازاء مقادیری مشخص برای گروه‌های N_x و U_y و مقداری مشخص برای p ، تنها یک گروه یکتا برای N_y تولید می‌شود، در حالی‌که در روش الگوریتم ژنتیک این‌گونه نبوده و در هر مرتبه از اجرای این الگوریتم جوابی متفاوت به‌دست می‌آید. دلیل این تفاوت در این است که چندین ترتیب جابه‌جایی متفاوت در اعداد گروه N_y وجود خواهد داشت که منجر به مقدار مشخصی از ضریب همبستگی با گروه واحد N_x شود. با این حال نشان دادیم که هرچه این ضریب همبستگی قویتر باشد، شمار پاسخ‌های متفاوت کمتر خواهد شد (مثلاً شکل ۱)؛ در نتیجه، با توجه به این‌که در تولید گروه‌های همبسته ماهیت تصادفی آن نیز مدنظر می‌باشد، الگوریتم ژنتیک را کارا تر از شبیه‌سازی مونت کارلو می‌انگاریم.

کوتاهتر باشد، احتمال پیروی کردن اعداد تصادفی از توزیع یکنواخت کمتر می‌شود. (به‌عنوان مثالی حدی، احتمال بسیار کمی، $6/6^6 = 0.0154$ ، وجود دارد که در ۶ بار ریزش تاسی ناریب هر ۶ وجه متفاوت تاس مشاهده شود)؛ بنابراین، تفاوت بین احتمال‌های تجمعی گروه تولید شده با طول کوتاهتر بیشتر شده، و در نتیجه با تغییر ساختار گروه دوم (y)، احتمال وقوع همبستگی‌های بیشتر افزایش می‌یابد.

روش پیشنهادی خود را با الگوریتم سنتی مونت کارلو مقایسه کردیم. دو گروه تصادفی ۳۰ تایی بر مبنای توزیع یکنواخت معیار (U_x و U_y) تولید و آنها را به بهنجار معیار (N_x و N_y) تبدیل نمودیم؛ بنابراین، انتظار نداریم که بین دو گروه N_x و N_y همبستگی وجود داشته باشد. سپس بازاء هر مقداری از N_x ، و مقداری مشخص برای p ، امید ریاضی و انحراف معیار شرطی متغیر تصادفی Y مشروط به یافته‌ی مشخصی برای متغیر تصادفی X (یعنی $X=x$) محاسبه شد (روابط ۱۱ و ۱۲). آن‌گاه تابع چندک شرطی متغیر تصادفی Y را با استفاده از این مقادیر استخراج کردیم. سپس بازاء مقدار مشخص X مقدار متناظر گروه U_y را تعیین (ردیف‌های یکسان در تولید اعداد تصادفی یکنواخت معیار) و در نهایت مقداری (یافته‌ای) را برای Y محاسبه کردیم. گروه تولید شده برای Y بر مبنای روش مونت کارلو را، که با X بهنجار همبسته است، $N_{y,M}^p$ نامیدیم و آن‌را ملاکی برای مقایسه با روش پیشنهادی قلمداد کردیم. برای انجام مقایسه، گروه N_x را ثابت در نظر گرفته و گروه N_y را برای ارضاء تابع هدف ۱۳ تغییر دادیم. همبستگی این گروه- که ماهیتاً بهنجار است (چون N_y بهنجار است)- با گروه بهنجار N_x برابر با p است. این گروه را $N_{y,G}^p$ نامیدیم. سپس همبستگی دو گروه $N_{y,M}^p$ و $N_{y,G}^p$ مورد بررسی قرار گرفت. در این‌جا دو‌گونه تغییرپذیری وجود دارند: اول این‌که تولید داده‌های U_x و

جدول ۱. متوسط، $\bar{\rho}$ ، و انحراف معیار، std ، ضریب همبستگی بین دو گروه $N_{y,M}^p$ و $N_{y,G}^p$ متناظر با ضریب همبستگی مشخص (ρ) برای تولید دو گروه تصادفی بهنجار معیار همبسته به روش مونت‌کارلو. مقادیر ارایه شده ناشی از ۱۰۰۰ بار اجرای تصادفی الگوریتم ژنتیک است.

std	$\bar{\rho}$	ρ	std	$\bar{\rho}$	ρ
۰/۰۱۷	۰/۸۴	۰/۹	۰/۰۶۷	۰/۶۴	-۰/۹
۰/۰۴	۰/۶۵	۰/۸	۰/۰۸۲	۰/۵۵	-۰/۸
۰/۰۹۰	۰/۴۵	۰/۷	۰/۰۷۴	۰/۵۲	-۰/۷
۰/۱۱۰	۰/۳۶	۰/۶	۰/۱۳۲	۰/۲۹	-۰/۶
۰/۱۵۶	۰/۱۵	۰/۵	۰/۱۴۷	۰/۳۰	-۰/۵
۰/۱۵۵	۰/۱۴	۰/۴	۰/۱۹۰	۰/۱۴	-۰/۴
۰/۱۹۱	۰/۰۳	۰/۳	۰/۱۷۰	۰/۱۲	-۰/۳
۰/۱۶۴	۰/۰۳	۰/۲	۰/۱۶۱	۰/۰۸	-۰/۲
۰/۱۹۰	۰/۰۵	۰/۱	۰/۱۹۶	۰/۰۴	-۰/۱



شکل ۱. فراوانی ضریب همبستگی بین دو گروه از توزیع یکنواخت معیار به طول ۵ بطوری که گروه اول ثابت نگاه داشته شده و کلیه‌ی جایگشته‌های گروه دوم ($5! = 120$) تأمین شده باشد.

سه متغیر تصادفی X ، Y و Z را با یافته‌های x_i ، y_i و z_i (که $i = \{1, \dots, n\}$ است) در نظر می‌گیریم. این یافته‌ها بر مبنای توزیع متغیرهای تصادفی نظیر خود و مستقل از متغیرهای تصادفی دیگر و با استفاده از روش سنتی مونت‌کارلو برای گروه‌های منفرد تولید شده‌اند. می‌خواهیم سه گروه X' ، Y' و Z' را به‌گونه‌ای تولید کنیم که همبستگیهای متقابل آنها با یکدیگر به‌ترتیب برابر با مقادیر معلوم ρ_{xy} ، ρ_{xz} و ρ_{yz} باشند. برای رسیدن به این

شبیه‌سازی سه گروه همبسته

از نقطه‌ی نظر کاربردی، روش مونت‌کارلو سنتی قادر به انجام این شبیه‌سازی نمی‌باشد زیرا عموماً روابط نظری مربوط به توزیع مشترک سه متغیره، و امید ریاضی و پراش شرطی برای آنها نه برای هر توزیع دلخواه، بلکه برای توزیع بهنجار نیز در دست نیست؛ بنابراین پس از نشان دادن کارایی روش پیشنهادی در بخش ۱/۳، این شبیه‌سازی را با بهینه‌سازی ژنتیک پی می‌ریزیم.

تصادفی تغییر دادیم و در هر مرتبه $\rho_{x,z}$ و $\rho_{y,z}$ را تعیین کردیم. این کار را برای مقادیر محتمل $\rho_{x,y}$ (از $-0/9$ تا $+0/9$) تکرار کردیم. جدول ۲ نتیجه‌ی به‌دست آمده را نشان می‌دهد. بر اساس این جدول هرچه $\rho_{x,y}$ قویتر باشد همبستگی بین $\rho_{x,z}$ و $\rho_{y,z}$ نیز قویتر است؛ به بیان دیگر $\rho_{y,z}$ را می‌توان با دقت کافی از $\rho_{x,y}$ و $\rho_{x,z}$ محاسبه کرد، که به این مفهوم است که چنانچه تابع هدف به برابری تنها برای دو ضریب همبستگی $\rho_{x,y}$ و $\rho_{x,z}$ دست یابد کافی است. این نتیجه به این مفهوم است که برای همبستگیهای قوی $\rho_{x,y}$ می‌توان تابع هدف ۱۴ را ساده‌تر کرد، درحالی‌که هرچه همبستگی $\rho_{x,y}$ ضعیفتر باشد، به تمامی سه جمله در تابع هدف ۱۴ نیاز می‌باشد. در نتیجه بطور کلی احتمال خلاصه کردن تابع هدف وجود ندارد، گرچه خللی به الگوریتم پیشنهادی وارد نمی‌آید.

جدول ۲. ویژگیهای خط رگرسیون $\rho_{y,z} = a + b\rho_{x,z}$ (a و b) برای مقادیر از پیش تعیین شده‌ای از $\rho_{x,y}$ ناشی از ۱۰۰۰ بار اجرای تصادفی الگوریتم ژنتیک برای سه متغیر تصادفی یکنواخت معیار به طول ۳۰.

b	a	ρ	$\rho_{x,y}$	b	a	ρ	$\rho_{x,y}$
۱/۰۴	-۰/۰۲	۰/۹۰	۰/۹	-۰/۹۰	۰/۸۶	-۰/۷۸	-۰/۹
۰/۸۹	۰/۰۴	۰/۷۸	۰/۸	-۰/۸۵	۰/۸۴	-۰/۷۴	-۰/۸
۰/۸۲	۰/۰۷	۰/۷۱	۰/۷	-۰/۷۶	۰/۸۰	-۰/۶۶	-۰/۷
۰/۶۷	۰/۱۴	۰/۵۸	۰/۶	-۰/۷۰	۰/۷۷	-۰/۶۱	-۰/۶
۰/۵۷	۰/۱۹	۰/۵۰	۰/۵	-۰/۵۴	۰/۷۰	-۰/۴۷	-۰/۵
۰/۴۵	۰/۲۴	۰/۳۹	۰/۴	-۰/۵۰	۰/۶۸	-۰/۴۴	-۰/۴
۰/۳۶	۰/۲۸	۰/۳۲	۰/۳	-۰/۳۸	۰/۶۲	-۰/۳۳	-۰/۳
۰/۲۶	۰/۳۳	۰/۲۲	۰/۲	-۰/۲۰	۰/۵۴	-۰/۱۷	-۰/۲
۰/۱۳	۰/۳۹	۰/۱۶	۰/۱	-۰/۱۴	۰/۵۱	-۰/۱۲	-۰/۱

قدردانی

این تحقیق در زمان حضور نویسنده‌ی اول در دانشگاه مانی‌توبا، و در قالب فرصت مطالعاتی اعطا شده به وی به وسیله‌ی دانشگاه فردوسی مشهد، انجام گردید که بدین-وسیله تشکر می‌شود

هدف، یکی از این گروهها را به‌دلخواه ثابت در نظر می‌گیریم (مثلاً X) و دو گروه دیگر را آن‌قدر درون خود تغییر می‌دهیم که تابع هدف زیر را کمینه کند:

$$\min : \left\{ \max \left(\left| \rho_{x,y} - \rho'_{x,y} \right|, \left| \rho_{x,z} - \rho'_{x,z} \right|, \left| \rho_{y,z} - \rho'_{y,z} \right| \right) \right\} \quad (14)$$

که در آن $\rho_{a,b}$ همبستگی مورد نظر بین متغیرهای a و b و $\rho'_{a,b}$ همبستگی محاسبه شده بین این متغیرها (رابطه‌ی ۱۰) می‌باشند. این تابع هدف شکل تعمیم یافته‌ی رابطه‌ی ۱۳ است. آیا می‌توان تابع هدف را ساده‌تر کرد؛ به این صورت که اگر $\rho_{x,y}$ و $\rho_{x,z}$ معلوم باشند، آیا مقداری یکتا برای $\rho_{y,z}$ وجود خواهد داشت؟ برای بررسی این امر سه گروه ۳۰ تایی از توزیع یکنواخت معیار تولید کردیم (هر توزیع دیگری نیز امکان‌پذیر است؛ با توزیع بهنجار معیار به نتیجه‌ی یکسانی رسیدیم). گروه X را ثابت نگاه داشته، گروه Y را تغییر دادیم تا $\rho_{x,y}$ دلخواهی (مثلاً ۰/۲) به‌دست آید. سپس گروه Z را ۱۰۰۰ بار بطور

نتیجه‌گیری

استخراج توابع توزیع مشترک، توزیع شرطی و ویژگیهای آن (شامل امید ریاضی و انحراف معیار) برای توزیعهای همبسته‌ی غیربهنجار دشوار بوده و عموماً راه‌حل مشخصی ندارد. برای چیرگی بر این مشکل، روش بهینه‌سازی برپایه‌ی الگوریتم ژنتیک دست‌کاری گردیده پیشنهاد شد. نتایج نشان دادند که این شیوه را می‌توان برای دو و چند متغیر به‌کار برد، به‌گونه‌ای که ساختار همبستگی با متغیر(های) دیگر حفظ شود.

- منابع
4. Hosking, J.R.M. 1994. The four-parameter kappa distribution. *IBM J. Res. Develop.* 38, 251-258.
 5. Hosking, J.R.M., and J.R. Wallis. 1997. *Regional frequency analysis. An approach based on L-moments.* Cambridge University Press. 224 p.
 6. Kleidon, A. 2010. Non-equilibrium thermodynamics, maximum entropy production and earth-system evolution. *Philos. Tran. Royal Soc. A.* 368(1910), 181-196. doi: 10.1098/rsta.2009.01188.
 7. Rodriguez-Iturbe, I., A. Rinaldo, R. Rigon, R.L. Bras, A. Marani, and E. Ijjasz-Vasquez. 1992. Fractal structure as least energy patterns: The case of river networks. *Geophys. Res. Lett.* 19:889-892.
 1. Ang, A.H-S., and W.H. Tang. 1984. *Probability concepts in engineering and design. Volume II- Decision, Risk, and Reliability.* John Wiley and Sons, Inc. 562 p.
 2. Ang, A.H-S., and W.H. Tang. 2007. *Probability concepts in engineering. Emphasis on applications to civil and environmental engineering.* John Wiley and Sons, Inc. 2nd edition. 406 p.
 3. Casterllarin, A., D.H., Burn, and A. Brath, 2008. Homogeneity testing: how homogeneous do heterogeneous cross-correlated regions seem? *J. Hydrol.* 360,67-76.