فرایندهای خود مشابه

محمّد حسین علامت‌ساز*

۱ مقدمه

هدف از این مقاله معرفی فرایندهای تصادفی خود مشابه[1] و روشن نمودن انگیزهٔ اولّیه و اهمیّت کاربرد آنها در زمینه‌های مختلف آمار است. علاوه بر این، سعی می‌کنیم که خواننده را با ریاضیات مربوط به طریقی هر چه ساده‌تر آشنا سازیم. همان‌طور که خواهیم دید این گونه فرایندها در کارهای نظری و کاربردی، هر دو، سابقه‌ای نسبتاً طولانی داشته و در هر دو زمینه شهرت کافی دارند. گرچه ابتدا در مسائل نظری بود که خوش رفتاری و قابلیت این فرایندها معلوم گردید، لیکن امروزه آنها نقش به‌سزایی در تحلیل داده‌های فرایندهای تصادفی ایفا می‌کنند. این مفهوم به ویژه در رابطه با خاصیت حافظهٔ‌ـطولانی[2]، یعنی فرایندهایی که دارای ضرایب همبستگی دراز مدت[3] هستند، مورد استفاده قرار می‌گیرد. با این حال در مقالهٔ حاضر قصد نداریم وارد بحثهای فنی پیچیده مربوط به نظریه وکاربردها، آماره‌های مربوط و استنتاجهای آماری آنها شویم. بلکه، هدف اصلی جلب توجه کلیّهٔ تحلیلگران داده‌ها و نشان دادن اهمیّت این فرایندها برای آنهایی است که با مجموعه‌هایی بزرگ از داده‌ها سرو کار دارند.

به طور خلاصه، فرض استقلال یا وابستگی کوتاه مدت[4] که معمولاً در مسائل مختلف به طور خودکار و بیشتر برای سهولت کار فرض می‌شود (مثلاً در سیستمهای صف‌بندی عام یا در حالت پایا[5] و یا تعادل[6])، در بیشتر مجموعه داده‌های حقیقی مصداق نمی‌یابد و پیامد نامطلوبی بر آزمونها و فواصل بازه‌های اطمینان می‌گذارد. الگوهای خود مشابه بهترین و ساده‌ترین روشهای ریاضی برای تسخیر سیمای اصلی این گونه ضرایب همبستگی دراز مدّت هستند که در یک موقعیت واقعی از داده‌ها مشاهده می‌شوند.

۲ انگیزه

می‌دانیم که واریانس میانگین نمونه‌ای برابر با واریانس جامعه تقسیم بر اندازه نمونه است. به طور دقیقتر، فرض کنید $X_1, X_2, \ldots, X_n$ مشاهداتی نمونه‌ای با میانگین مشترک $\mu = E(X_i)$ و واریانس $\sigma^2 = \mathrm{Var}(X_i)$ باشند. در این صورت، واریانس $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ برابر با σ^2/n است، یعنی

$$\mathrm{Var}(\bar{X}) = n^{-1}\sigma^2 \tag{1}$$

علاوه بر این، $\bar{X}$ یک برآورد کنندهٔ μ است و برای نمونه‌های بزرگ یک فاصلهٔ اطمینان $(1-\alpha)100\%$ برای μ وقتی که σ^2 معلوم باشد عبارت است از

$$\bar{X} \pm Z_{\alpha/2}\sigma/\sqrt{n} \tag{2}$$

و وقتی که σ^2 نامعلوم باشد، به صورت زیر درمی‌آید:

$$\bar{X} \pm Z_{\alpha/2}S/\sqrt{n} \tag{3}$$

در اینجا مطابق استاندارد معمول، $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ واریانس

*) دکتر محمدحسین علامت‌ساز، گروه آمار، دانشگاه اصفهان

1) self-similar processes 2) long-memory property 3) long-range correlations 4) short-range dependence 5) stationary 6) in equilibrium

نمونه‌ای و $Z_{\alpha/2}$ نقطهٔ بحرانی $(1 - \alpha/2)$ام بالایی توزیع نرمال استانداردند.

فرمولهای (۱)، (۲) و (۳) بسیار ساده بوده و ما را به طور خودکار به استفاده از آنها ترغیب می‌کنند بدون اینکه شرایط مورد نیاز آنها را قبلاً در کنترل داشته باشیم. امّا، یک فرض مهّم در اینجا آن است که X_iها ناهمبسته باشند، یعنی داشته باشیم

$$\rho(i,j) = \mathrm{Corr}.(X_i, X_j) = 0, \forall i \neq j$$

یا به طور کلیتر، مستقل از یکدیگر باشند. گاهی تصور می‌شود که این فرض درست است و بعضاً درستی این فرض به طور تقریبی قابل توجیه است. ولی در مواقع دیگر، متأسفانه این خواسته به واقعیت نمی‌پیوندد. پس، مسأله این است که اگر این فرض برقرار نباشد، روابط (۱)، (۲) و (۳) چقدر نادرست‌اند و چگونه می‌توان آنها را تصحیح نمود؟

چون فرمولهای (۲) و (۳) به طور قابل توجهی به فرمول (۱) وابسته‌اند، در آنچه ذیلاً بحث خواهد شد توجه خود را تنها بر فرمول (۱) معطوف کرده خواهیم دید که وقتی مشاهدات همبسته باشند، این فرمول چه وضعی پیدا می‌کند. فرض کنید $\mu = E(X_i)$ ثابت باشد. آنگاه،

$$\mathrm{Var}(\bar{X}) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\gamma(i,j) = n^{-2}\sigma^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(i,j) \qquad (۴)$$

که در آن $\gamma(i,j)$ اتوکوواریانس بین X_i و X_j است. پس، داریم

$$\mathrm{Var}(\bar{X}) = \sigma^2 n^{-1}[1 + \delta_n(\rho)] = \sigma^2 n^{-1} C_n(\rho) \qquad (۵)$$

که در آن

$$\delta_n(\rho) = n^{-1}\sum_{i\neq j}^{n}\rho(i,j) \& C_n(\rho) = 1 + \delta_n(\rho) \qquad (۶)$$

بدیهی است که اگر مجموع ضریبهای همبستگی در (۶) برابر صفر شوند، یعنی،

$$\sum_{i\neq j}^{n}\rho(i,j) = 0 \qquad (۷)$$

آنگاه (۵) برابر σ^2/n شده و در این حالت فرمول (۱) برقرار می‌گردد. به ویژه، این امر وقتی صحّت می‌یابد که $X_1, X_2, \ldots, X_n$ دوبه‌دو ناهمبسته (و یا به طور کلیتر مستقل) باشند. علاوه بر این، اگر فرایند مانا باشد به طوری که $\rho(i,j)$ تنها به i و j از طریق کمیت $|i - j|$ وابسته باشد، برای سهولت می‌توان از نماد

$$\rho(i-j) = \rho(j-i) = \rho(i,j)$$

استفاده نمود و در این صورت، (۶) را ممکن است به شکل سادهٔ زیر نمایش داد:

$$\delta_n(\rho) = 2\sum_{k=1}^{n-1}(1 - k/n)\rho(k) \qquad (۸)$$

رابطهٔ (۵) نشان می‌دهد که در این حالت طرف راست (۱) باید در عامل $C_n(\rho)$ ضرب شود. از رابطهٔ (۶) برمی‌آید که اگر وابستگی بسیار قوی باشد، σ^2/n تقریب بسیار ضعیفی برای واریانس $\bar{X}_n$ است. در چنین حالتی، داشتن برآورد خوبی از $C_n(\rho)$ اهمیت می‌یابد. اگر

$$\delta(\rho) = \lim_{n\to\infty}\delta_n(\rho) = \lim_{n\to\infty} n^{-1}\sum_{i\neq j}^{n}\rho(i,j)$$

موجود، متناهی و بزرگتر از ۱− باشد، ممکن است (۵) را به طور مجانبی به صورت

$$\mathrm{Var}(\bar{X}) = \sigma^2 n^{-1}[1 + \delta(\rho)] = \sigma^2 n^{-1} C(\rho) \qquad (۹)$$

نوشت. اغلب الگوهای سری زمانی معروف این رفتار را نشان می‌دهند. که مشهورترین آنها فرایندهای ARMA و فرایندهای مارکوف هستند (بران، ۱۹۹۴، را ملاحظه کنید).

۳ فرایندهای تصادفی خود مشابه

برای بسیاری از مجموعه داده‌ها، ساده‌ترین تقریب منطقی واریانس میانگین نمونه‌ای به صورت یک قانون توانی به شکل $\sigma^2 n^{-\alpha}$، $0 < \alpha < 1$، است (مثلاً، به همپل و دیگران، ۱۹۸۶ فصل ۱-۸ مراجعه کنید). چنین قانون توانی را نمی‌توان از هیچ فرایند تصادفی با وابستگی کوتاه مدت به دست آورد (همپل ۱۹۸۷)، امّا این قانون به طور دقیق برای نموهای یک فرایند خود مشابه برقرار است. چون اختلاف فرمول (۱) با واقعیت در رشته‌های مختلف علوم نگران کننده بود، ابتدا کولموگوروف ردهٔ فرایندهای خود مشابه را به صورت مجرّد آن ابداع نمود. لیکن، در واقع بعدها مندلبرات[1] و همکاران تحقیقاتی‌اش در سالهای ۱۹۶۰ بودند که عملاً توجه آماردانان را به این فرایندها جلب کردند. برای نمونه، مندلبرات (۱۹۸۲) را ملاحظه نمایید.

سپس، آنها به سرعت در زمینه‌های مختلف از جمله نظریهٔ اغتشاش، مهندسی برق، اقتصاد، هواشناسی، ژئوفیزیک، سیستمهای مخابراتی و نظریهٔ صف‌بندی به کار گرفته شدند. برای جزئیات بیشتر در مورد کاربردها و مراجع مربوط، خوانندگان گرامی علاقمند را به تقّا[2] (۱۹۸۶) ارجاع می‌دهیم.

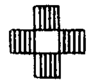
یکی از خصوصیات جذّاب فرایندهای خود مشابه این است که آنها در مقیاسهای مختلف یکسان به نظر می‌رسند (به رابطهٔ (۱۰) در زیر توجه کنید).

1) Mandelbort 2) Taqqu

به عبارت دیگر، این نوع فرایندها وقتی اتّفاق می‌افتند که هنگام بزرگ و بزرگتر شدن زیر ذرّه‌بین به صورت اجتماع تعدادی زیاد از «کپی»های کوچکتر از نوع خودشان به نظر برسند. مثال توصیفی در این مورد مجموعه‌های مشهور کانتور بر $\mathbb{R}$ و بر $\mathbb{R}^2$ هستند:

مرحلهٔ اول	مرحلهٔ دوم	مرحلهٔ سوم	مرحلهٔ چهارم و ...
$[0,1]$	$[0,\frac{1}{3}]\cup[\frac{2}{3},1]$	$[0,\frac{1}{9}]\cup[\frac{2}{9},\frac{1}{3}]\cup$ $[\frac{2}{3},\frac{7}{9}]\cup[\frac{8}{9},1]$	......

مجموعه‌های کانتور در R

مرحلهٔ اول	مرحلهٔ دوم	مرحلهٔ سوم	مرحلهٔ چهارم و ...
			......

مجموعه‌های شبه ـ کانتور در R^2

تعریف رسمی فرایندهای تصادفی خود مشابه به صورت زیر است.

تعریف فرایند تصادفی حقیقی $Z(t)$، $t \geq 0$، خود مشابه با پارامتر $H > 0$ است $(H-SS)$، اگر برای هر عدد $\alpha > 0$ داشته باشیم

$$Z(\alpha t) =^d \alpha^H Z(t), \quad t \geq 0 \qquad (10)$$

که در آن $=^d$ نشان دهندهٔ برابری تمام توزیعهای متناهی بُعد دو طرف است. پارامتر H به پارامتر هرست[1] موسوم است. به عبارت دیگر، $Z(t)$ یک فرایند $H-SS$ است اگر برای هر عدد $\alpha > 0$ و هر دنباله‌ای از نقاط زمانی t_1، t_2، ...، t_n توزیع توأم $Z(\alpha t_1)$، $Z(\alpha t_2)$، ...، $Z(\alpha t_n)$ با توزیع توأم $\alpha^H Z(t_1)$، $\alpha^H Z(t_2)$، ... و $\alpha^H Z(t_n)$ یکسان باشد. به همین ترتیب، می‌گوییم فرایند خود مشابه $Z(t)$ با پارامتر H دارای نموهای ماناست $(H-SSSI)$، اگر

$$Z(t_\circ + t) - Z(t_\circ) =^d Z(t) - Z(0), \quad \forall t, t_\circ \geq 0$$

به عنوان مثال، فرایند حرکت براونی[2] $B(t)$ از این نوع است. در واقع، این فرایند، یک فرایند $\frac{1}{2}-SSSI$ است. امّا فرایند $W(t) = |B(t)|$ تنها $\frac{1}{2}-SS$ است.

مثالهای کلاسیک فرایندهای $SSSI$ مثل فرایندهای پایدار محض[3] با اضافه کردن این فرض که نموها مستقل نیز هستند، به دست می‌آیند. فرایندهای خود مشابه مارکوف در زمرهٔ اولین فرایندهای خود مشابه بودند که مشخص سازی شدند. امّا، در عمل، یک خاصیت جالب فرایندهای $SSSI$ مربوط به هنگامی است که آمارگر تلاش در به حساب آوردن ضرایب همبستگی دراز مدت دارد و نه استقلال و یا حتّی ضرایب همبستگی کوتاه مدت (ویلینیگر و دیگران، ۱۹۹۵، را برای مثال ملاحظه کنید).

برای محاسبهٔ تابع خود همبستگی یک فرایند خود مشابه، ابتدا خاطر نشان می‌کنیم که (۱۰) نتیجه می‌دهد، با احتمال یک، $Z(0) = 0$ و

$$Z(t) =^d t^H Z(1), \quad \forall t > 0 \qquad (11)$$

برای سهولت در نمادگذاری، فرض کنید که $E(Z(t)) = 0$، $t \geq 0$. به‌علاوه، فرض کنید که فرایند نموها یعنی $X(t) = Z(t) - Z(t-1)$ دارای واریانس متناهی

$$\sigma^2 = E([Z(t) - Z(t-1)]^2)$$

باشد. در این صورت، بنابر مانایی $X(t)$، داریم $\sigma^2 = E(X^2(1))$. پس، برای هر $s < t$ با استفاده از (۱۱) می‌توان نوشت

$$\begin{aligned} E([Z(t) - Z(s)]^2) &= E([Z(t-s) - Z(0)]^2) \\ &= E(Z^2(t-s)) \\ &= (t-s)^{2H}\sigma^2 \end{aligned}$$

از طرف دیگر،

$$\begin{aligned} E([Z(t) - Z(s)]^2) &= E(Z^2(t)) + E(Z^2(s)) - 2E(Z(t)Z(s)) \\ &= \sigma^2 t^{2H} + \sigma^2 t^{2s} - 2\,\mathrm{Cov}(Z(t), Z(s)) \end{aligned}$$

بنابراین، از اینجا تابع کوواریانس یک فرایند $H-SS$ به صورت زیر حاصل خواهد شد.

$$\begin{aligned} \sigma(t,s) &= \mathrm{Cov}(Z(t), Z(s)) \\ &= \frac{1}{2}\sigma^2[t^{2H} - (t-s)^{2H} + s^{2H}] \qquad (12) \end{aligned}$$

پس، کوواریانس بین X_i و X_{i+k} $(k > 0)$ به سادگی به صورت زیر به دست می‌آید،

$$\begin{aligned} \gamma(k) &= \mathrm{Cov}(X_i, X_{i+1}) \\ &= \mathrm{Cov}(X_1, X_{1+k}) \\ &= E(X_1 X_{1+k}) \\ &= E([Z(1) - Z(0)][Z(1+k) - Z(k)]) \\ &= E(Z(1)Z(k+1)) - E(Z(1)Z(k)) \\ &= \sigma(k+1, 1) - \sigma(k, 1) \\ &= \frac{1}{2\sigma^2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \qquad (13) \end{aligned}$$

و $\gamma(k) = \gamma(-k)$ برای $k < 0$. در نتیجه، تابع خود همبستگی k قدمی[4] $(k > 0)$ نموهای یک فرایند $H-SS$ به شکل زیر خواهد بود:

1) Hurst parameter 2) Brownian motion process 3) strictly stable processes 4) lag of k

$$\rho(k) = \frac{1}{2}[(k+1)^{2H} - 2k^{2H} + (k-1)^{2H}] \qquad (14)$$

و برای $k < 0$، $\rho(k) = \rho(-k)$.

اینجا خاطر نشان می‌کنیم که اثبات عبارت $\gamma(k)$ که در (۱۳) ارائه داده‌ایم ساده‌تر از آن است که قبلاً در نوشته‌های مربوط ظاهر شده است. علاوه بر این، توجه کنید که در محاسبهٔ تابع ضریب همبستگی (۱۴) ما از قدرت کامل خاصیت خود مشابهی فرایند $Z(t)$ استفاده نکرده‌ایم. در واقع، علاوه بر خاصیت مانایی نموها، تنها از برابری در واریانسهای طرفین رابطهٔ (۱۰) یعنی رابطهٔ

$$\mathrm{Var}[Z(\alpha t)] = \alpha^{2H}\,\mathrm{Var}[Z(t)], \qquad \forall \alpha > 0, t \geq 0 \quad (15)$$

یا هم ارز آن

$$\mathrm{Var}[Z(t)] = t^{2H}\,\mathrm{Var}[Z(1)], \qquad \forall t > 0$$

سود برده‌ایم. این خود انگیزهٔ بررسی فرایندهای خود مشابه مرتبهٔ دوم [1] $(H-SSSO)$ توسط بعضی از مؤلفان مانند علامت‌ساز و لین (۱۹۹۷) شده است. فرایند $Z(t)$ را $H-SSSO$ گوییم هرگاه در رابطهٔ (۱۵) صدق نماید. بدیهی است که هر فرایند $H-SS$ یک فرایند $H-SSSO$ نیز هست ولی عکس این موضوع البته مصداق ندارد. فرایند پواسون، برای مثال، $\frac{1}{2}-SSSO$ است ولی SS نیست.

به طوری که در بران (۱۹۹۴) نشان داده شده است، وقتی کوواریانسها موجود باشند و $\lim_{k\to\infty}\rho(k) = 0$ مقادیر پارامتر H به 0 و 1 محدود می‌شوند یعنی، در این صورت، داریم $0 < H < 1$. برای $\frac{1}{2} < H < 1$، ضرایب همبستگی چنان آهسته به صفر می‌گرایند که $\sum_{k=-\infty}^{\infty}\rho(k) = \infty$. در این صورت، فرایند نموها X_i $(i = 1, 2, \ldots)$ دارای حافظهٔ طولانی یا فرایند $X(t)$ دارای وابستگی دراز مدت است. برای $H = \frac{1}{2}$، مشاهدات ناهمبسته‌اند. برای $0 < H < \frac{1}{2}$، فرایند دارای وابستگی کوتاه مدت است و مجموع ضرایب همبستگی برابر صفر می‌گردد، یعنی $\sum_{k=-\infty}^{\infty}\rho(k) = 0$.

ردهٔ مهم دیگری از فرایندهای $H-SS$، کلاس حرکت براونی کسری[2] $B_H(t)$ است. برای تعریف و بحث در مورد خود مشابهی این فرایندها به بران (۱۹۹۴، صفحهٔ ۵۵) مراجعه کنید. علامت‌ساز و لین (۱۹۹۷) رده‌ای از فرایندهای تصادفی مرکب $H-SSSO$ را ارائه می‌کنند. ضمناً لازم به تذکر است که فرایندهای خوشتعریف $H-SSSI$ ای با گشتاورهای مرتبهٔ دوم نامتناهی وجود دارند که برای آنها $H \geq 1$. برای تحقیقات بیشتر در این زمینه می‌توان به کتاب سامور و دنیتسکی و تقّا (۱۹۹۴) مراجعه نمود.

بالاخره، باید خاطر نشان کنیم که مطالعهٔ وسیعی در رابطه با مشخّص سازی فرایندهای خود مشابه و آماره‌های آنها انجام گردیده است. این مطالعات شامل برآورد کردن H و پارامترهای دیگر مربوط به مشخّص سازی وابستگی، آزمون فرضها، فواصل اطمینان، پیش‌بینی، رگرسیون و تحلیل واریانس می‌گردد. طبیعی است که بحث دربارهٔ این همه از عهدهٔ این مقاله خارج است. برای مطالعهٔ بیشتر در این زمینه‌ها می‌توان به همپل (۱۹۸۷) و یا بران (۱۹۹۴) مراجعه کرد. مقالهٔ حاضر را با مشاهدهٔ یک خاصیت مطلوب نموهای مانای

$$X_i = Z(i) - Z(i-1), \qquad i = 1, 2, 3, \ldots$$

از یک فرایند تصادفی $H-SS$ به پایان می‌بریم. به وضوح، میانگین نمونه را می‌توان به صورت زیر نوشت:

$$\begin{aligned}\bar{X} &= n^{-1}\sum_{i=1}^{n} X_i = n^{-1}(Z(n) - Z(0)) \\ &=^{d} n^{-1}n^{H}(Z(1) - Z(0))\end{aligned}$$

پس، به این ترتیب به دست می‌آوریم

$$\mathrm{Var}(\bar{X}) = n^{2H-2}\sigma^2$$

که برای $H = \frac{1}{2}$ به همان نتیجهٔ کلاسیک (۱) یعنی $\mathrm{Var}(\bar{X}) = \sigma^2/n$ دست می‌یابیم. به علاوه، اگر X_i فرایندی گاوسی[3] با میانگین μ و واریانس σ^2 باشد (که در این صورت، $Z(t)$ یک فرایند حرکت براونی کسری خواهد بود)، آنگاه $n^{1-H}(\bar{X} - \mu)/\sigma$ یک متغیر تصادفی نرمال استاندارد خواهد بود. بنابراین، از این آماره می‌توان در محاسبهٔ آزمونها و فواصل اطمینان μ سود جست.

1) self-similar of second order 2) fractional brownian motion 3) Gaussian process

مراجع

[1] Alamatsaz M. H. & Y. X. Lin (1994), Self-similarity under compounding. submitted.

[2] Beran J. (1994), Statistics for long-memory processes. Chapman & Hall, New York.

[3] Hampel F. R., E. M. Ronchetti, P. J. Rousseeuw & W. A. Stahel (1986), Robust statistics. Wiley, New York.

[4] Hampel F. R. (1987), Data analysis and self-similar processes. Proc. of 46 session of International Statistical Institute, Tokyo, p. 235-54.

[5] Kolmogorov A. N. (1940), Wienersche spiralen und einige andere interssante kurven in Hilbertschen Raum. C. R. (Doklady) Acad. Sci. USSR (N. S.) 26-115-18.

[6] Mandelbrot B. B. (1982), The fractal geometry of nature. San Francisco, W . H. Freeman.

[7] Taqqu M. S. (1986), A bibliographical guide to self-similar processes and long-range dependence. In "Dependence in probability and statistics", edited by E. Eberlein and M.S. Taqqu, Boston, Birkhäuser.

[8] Willinger W., M. S. Taqqu, W. E. Leland and D. V. Wilson (1995), Self -similarity in high-speed packet trafic: Analysis and modeling of ethernet trafic measurements. Statist. Sci., Vol. 10, No. 1, 67-85.

[9] Samorodnitsky G. & M. S. Taqqu (1994), Stable non-Gaussian random processes: stochastic models with infinite variance. Chapman & Hall, New York.