

آشنایی با فرایند نقطه ای

نیز اسمعیل زاده^۱

چکیده

یک فرایند نقطه ای، مدل‌هایی برای الگوهای بی قاعده نقاط ارائه می‌کند. در این مقاله فرایند نقطه ای را در فضای کلی R^d تعریف می‌کنیم و عمل‌هایی را روی فرایند نقطه ای مبدأ جهت سازگاری فرایند با وضعیت مورد بحث در عمل معرفی می‌کنیم. گشتاور مرتبه دوم کاهش یافته یا تابع K را همراه با چند قضیه که به آن جنبه کاملاً شهودی و کاربردی می‌دهند ارائه می‌کنیم. کلمات کلیدی: فرایند نقطه ای، گشتاور مرتبه دوم کاهش یافته

۱ مقدمه

داده‌هایی به شکل مجموعه ای از نقاط را که به طور نامنظم و بی قاعده داخل یک ناحیه توزیع شده‌اند، الگوهای نقطه ای^۲ گویند.

هدف، بنا کردن مدل‌هایی برای توزیع تصادفی نقاط در یک فضا است. فرایندهای نقطه ای، مدل‌های مناسبی برای چنین داده‌هایی هستند. فرایند نقطه ای در مدل سازی کاربردهای متعددی دارد. هر کدام از موارد زیر را می‌توان به عنوان یک الگوی نقطه ای در نظر گرفت و با استفاده از فرایند نقطه ای آنها را مدل سازی کرد.

(۱) مکانهای درختان در یک جنگل؛

(۲) مکانهای تانکها در میدان جنگ؛

(۳) مکانهای موردهای یک بیماری خاص در یک ناحیه جغرافیایی.

ابتدا بعضی موضوعات پایه ای و اساسی در رابطه با مدل سازی پخش تصادفی نقاط در یک فضا را بیان می‌کنیم. در ادامه اندازه گشتاور دوم کاهش یافته^۳ یا تابع K را که بسیاری جاها به عنوان ابزار مفیدی برای تحلیل الگوهای نقطه ای فضایی^۴ پذیرفته شده است، معرفی می‌کنیم [۴]، [۵].

۲ تعریف فرایند نقطه ای

یک فرایند نقطه ای، مدل‌هایی برای الگوهای بی قاعده نقاط ارائه می‌کند. یک فرایند نقطه ای روی R^d ، متغیری تصادفی است که مقادیرش را در فضای اندازه پذیر $[F, \mathcal{F}]$ می‌گیرد که F خانواده همه دنباله های ϕ از نقاط R^d است که در دو شرط زیر صدق می‌کنند:

(۱) دنباله ϕ به طور موضعی متناهی است (یعنی هر زیر مجموعه کراندار R^d باید شامل فقط تعدادی متناهی از

^۲ Reduced

^۴ Spatial

^۱ گروه ریاضی، دانشگاه کرمان

^۲ Point Patterns

نقاط $\varnothing$ باشد).

(۲) دنباله ساده است. (یعنی $x_i \neq x_j$ اگر $i \neq j$)

و $\mathcal{F}$ کوچکترین σ -میدان روی F است که همه توابع $\varnothing(B) \mapsto \varnothing(B)$ را اندازه پذیر می سازد. $\varnothing(B)$ تعداد نقاط دنباله $\varnothing$ داخل B متعلق به مجموعه های بسورل کراندار را نشان می دهد. نقاط فرایند x_i ، را پیشامدهای فرایند می نامیم، به این جهت که آنها را از نقاط دلخواه در R^d متمایز کنیم. بدین ترتیب یک فرایند نقطه ای N تابعی اندازه پذیر از فضای احتمال $[\Omega, \mathcal{A}, P]$ به $[F, \mathcal{F}]$ است. به طور شهودی فرایند نقطه ای، یک انتخاب تصادفی از یکی از دنباله های $\varnothing$ در F است.

فرایند نقطه ای را می توان یا به صورت مجموعه های تصادفی از نقاط گسسته در نظر گرفت یا به صورت اندازه های تصادفی که تعداد نقاط در هر ناحیه فضایی را شمارش می کند. مطابق این دو، نمادهای مختلف برای نمایش فرایند نقطه ای وجود دارند:

$x \in N$ بیان می کند که نقطه x متعلق به دنباله تصادفی $\varnothing$ است،

$N(B) = n$ بیان می کند که مجموعه B شامل n نقطه از N

است.

فرض کنید E یک زیر مجموعه فضای اقلیدسی باشد، می خواهیم یک نماد مناسب برای تابعی که تعداد تصادفی نقاط در یک مجموعه کراندار را شمارش کند، معرفی کنیم. فرض کنید که $\{X_n, n \geq 0\}$ عناصر تصادفی در E هستند که نقاط تصادفی در فضای وضعیت E را نشان می دهند. اگر اندازه زیر را تعریف کنیم:

$$I_{X_n}(A) = \begin{cases} 1 & ; X_n \in A \\ 0 & ; X_n \notin A \end{cases}$$

آنگاه با مجموعیابی روی n ، تعداد کل نقاط تصادفی را که در مجموعه A قرار می گیرند به دست می آوریم. حال اگر اندازه شمارشی N را با $N(\cdot) = \sum_n I_{X_n}(\cdot)$ تعریف کنیم آنگاه:

$$N(A) = \sum_n I_{X_n}(A)$$

تعداد تصادفی نقاطی است که در مجموعه A می افتند. N یک فرایند نقطه ای است و $\{X_n\}$ نقاط نامیده می شوند. نماد طوری طرح شده که به طور صریح وابستگی اندازه شمارشی را به نقاطش نشان می دهد.

به عنوان کاربردی عملی از یک فرایند نقطه ای فرض می کنیم که داده های موجود برای تحلیل شامل مکانهای همه مورد های یک بیماری خاص باشد. مکانها می توانند محل اقامت یا محل تولد یا محل کار بیمار باشند. این موردها از یک سرشماری کامل یا در صورت ممکن نبودن سرشماری، یک نمونه گیری تصادفی مناسب داخل یک ناحیه جغرافیایی مشخص در یک دوره زمانی $(0, T)$ به دست آمده اند. چارچوب مدل سازی مناسب برای چنین داده هایی فرایند نقطه ای فضایی است. در این چارچوب، مکان داده ها x_i ؛ $i = 1, 2, \dots$ را به عنوان یک تحقق جزئی فرایند در ناحیه A در نظر می گیریم.

روشی دیگر، گردآوری داده ها در اندازه های شمارشی $N(A_j)$ ، تعداد موردها در ناحیه $A_j \in A$ است. این عمل، ساختار فضایی خوب داده ها را از بین می برد و تا جایی که ممکن است باید از آن پرهیز شود. این نکته اهمیت استفاده از خود مکانهای واقعی موردها، x_i ، را به عنوان تحقق از فرایند نقطه ای در مقابل تجمع آنها در یک اندازه شمارشی نشان می دهد.

هر مکان x_i ممکن است با یک زمان t_i مثلاً زمان تولد یا زمان تشخیص بیماری تکمیل شود. در این صورت مجدداً می توان یک فرایند نقطه ای (x_i, t_i) روی $R^2 \times R$ برای این داده ها در نظر گرفت.

به عنوان یکی از رایجترین مدل های فرایند نقطه ای می توان به فرایند پواسون اشاره کرد.

۳ توزیع یک فرایند نقطه ای

توزیع یک فرایند نقطه ای N به وسیله احتمال

$$P(Y) = P(N \in Y)$$

$$= P(\omega \in \Omega : N(\omega) \in Y) \quad ; Y \in \mathcal{F}$$

و اندازه میانگین فرآیند نقطه ای را به صورت

$$\mu(A) = E[N(A)]$$

تعریف می کنیم. بدین ترتیب $\mu(A)$ تعداد نقاط مورد انتظار در ناحیه A است.

۴ مانایی و همسانگردی

فرآیند نقطه ای N را در نظر می گیریم. برای هر $x \in E$ فرآیند نقطه ای $N_x = \sum I_{x_n=x}$ را تعریف می کنیم، که در آن نقاط فرآیند N به وسیله بردار x انتقال داده می شوند. در این صورت، N ماناست هرگاه برای هر x ، N_x هم توزیع با N باشد. در حقیقت یک فرآیند را مانا^۵ گوئیم هرگاه هر گزاره احتمالی درباره فرآیند در هر ناحیه A تحت انتقال دلخواه A، ناورد^۶ باشد. اگر همین ناوردایی تحت دوران نیز برقرار باشد، فرآیند را همسانگرد^۷ گوئیم. یعنی در صورتی که هیچ اثر جهتی وجود نداشته باشد. در حقیقت همسانگرد بودن فرآیند دلالت بر این دارد که فرآیند در جهتهای مختلف ویژگیهای یکسانی از خود بروز می دهد.

۵ عملیاتی روی فرآیند نقطه ای

به منظور مطالعه مسائل واقعی، اغلب به ساختن مدل‌های فرآیند نقطه ای سازگار با وضعیت مورد بحث نیاز پیدا می کنیم. فرآیند های نقطه ای جدید با ترکیب یا تبدیل فرآیند های نقطه ای مبدأ حاصل می شوند. در اینجا سه عمل اساسی را که فرآیند نقطه ای جدید را از یک فرآیند مبدأ تولید می کنند بیان می کنیم. این سه عمل عبارت اند از:

(۱) برهمنهی^۸

(۲) فرآیند نقطه ای نشانه گذاری شده^۹

(۳) فرآیند نقطه ای تنک شده^{۱۰}

۱-۵ برهمنهی

فرض کنید N_1 و N_2 دو فرآیند نقطه ای با توزیع های P_1 و P_2 باشند. فرآیند نقطه ای N را به صورت

$$N = N_1 \cup N_2$$

در نظر بگیرید. N برهمنهی N_1 و N_2 نامیده می شود. فرض کنید با احتمال یک، مجموعه های نقاط N_1 و N_2 متداخل نباشند. در این صورت

$$N(A) = N_1(A) + N_2(A).$$

واضح است که اندازه میانگین و شدت (در حالت مانایی) فرآیند N عبارت اند از

$$\mu = \mu_1 + \mu_2$$

اگر توزیع N را با P نمایش دهیم آنگاه

$$P = P_1 * P_2$$

که * نماد پیچش P_1 و P_2 است.

۲-۵ فرآیند نقطه ای نشانه گذاری شده

یک فرآیند نقطه ای نشانه گذاری شده با ضمیمه کردن مشخصه ای (نشانه^{۱۱}) به هر نقطه فرآیند مبدأ حاصل می شود. بدین ترتیب یک فرآیند نقطه ای نشانه گذاری شده روی R^d دنباله ای تصادفی مانند $\psi = \{[x_n, m_n]\}$ است که نقاط x_n با هم تشکیل یک فرآیند نقطه ای در R^d می دهند و m_n ها نشانه های نظیر هر x_n هستند. نشانه های m_n ممکن است ساختاری پیچیده داشته باشند. آنها متعلق به یک فضای داده شده نشانه های M هستند. ماتس^{۱۲} در سال ۱۹۶۳ نظریه فرآیند نقطه ای نشانه گذاری شده را بنا نهاد. کاربردهای اولیه آن در نظریه صف بود. در زیر

Marked^۹

Thinned^{۱۰}

mark^{۱۱}

Mathes^{۱۲}

Stationary^۵

Invariant^۶

Isotropic^۷

Superposition^۸

$$\mu \times F(A_1 \times A_2) = \mu \times F((e_1, e_2): e_1 \in A_1, e_2 \in A_2) \\ = \mu(A_1)F(A_2).$$

برای اثبات گزاره ۱ به [۶] مراجعه شود.

۶ خاصیت‌های گشتاور مرتبه اول و دوم یک فرایند نقطه ای

فرض کنید $x_i; i = 1, 2, \dots$ نقاط یک فرایند نقطه ای در R^d باشد. A را هر ناحیه مسطح در نظر بگیرید، مساحت A را با $|A|$ نشان می دهیم و $N(A)$ تعداد پیشامدها در A و dx ناحیه بسیار کوچکی است که x را در بر می گیرد. تابع شدت^{۱۳} (مرتبه اول)، $\lambda(x)$ ، یک فرایند نقطه ای فضایی به صورت

$$\lambda(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{E[N(dx)]}{|dx|} \right\}$$

تعریف می شود. تابع $\lambda(x)$ متوسط تعداد پیشامدها در هر واحد مساحت در همسایگی نقطه x را نشان می دهد.

تابع شدت مرتبه دوم، $\lambda_2(x, y)$ ، به طور مشابه تعریف می شود:

$$\lambda_2(x, y) = \lim_{|dx| \rightarrow 0, |dy| \rightarrow 0} \left\{ \frac{E[N(dx)N(dy)]}{|dx||dy|} \right\}$$

اگر $N(dx)$ و $N(dy)$ ناهمبسته باشند، $\lambda_2(x, y) = \lambda(x)\lambda(y)$

یک فرایند نقطه ای مانا (مرتبه دوم) است اگر $\lambda(x) = \lambda$ و $\lambda_2(x, y) = \lambda_2(x - y)$ و همسانگرد است اگر $\lambda_2(x, y) = \lambda_2(\|x - y\|)$ که در آن $\|x\| = \sqrt{x_1^2 + x_2^2}$ فاصله اقلیدسی را نشان می دهد. کمیت

$$\gamma(x, y) = \lambda_2(x, y) - \lambda(x)\lambda(y)$$

را چگالی کوواریانس گویند که در حالت مانایی و همسانگردی فرایند، به صورت زیر در می آید:

چند حالت خاص از یک فرایند نقطه ای نشانه گذاری شده را بیان می کنیم:

- برای x مرکز یک ذره، m حجم ذره باشد،
 - برای x مکان یک درخت، m قطر تنه درخت یا ارتفاعش باشد،
 - برای x مرکز یک اتم، m نوع اتم باشد.
- نشانه ها اساساً می توانند متغیرهای پیوسته باشند، مانند دو مثال اول یا نوع را نشان دهند، مانند مثال سوم (که در این مورد اغلب اصطلاح فرایند نقطه ای چند متغیره به کار می رود). فرایند نقطه ای نشانه گذاری شده ψ را می توان مانند یک فرایند نقطه ای معمولی روی $R^d \times M$ در نظر گرفت. نماد اندازه شماری تصادفی را می توان مورد استفاده قرار داد. برای مجموعه بول B و L در M تعداد نقاط ψ در B با نشانه هایی در L ، با نشان $\psi(B \times L)$ داده می شود.

۳-۵ فرایند نقطه ای تنک شده

فرایند $N' = \sum Z_n I_{X_n}$ را در نظر بگیرید که در آن Z_n فقط مقادیر صفر و یک را می گیرد، بدین ترتیب N' یک فرایند نقطه ای است که نقاط آن زیر مجموعه نقاط N هستند. اگر Z_n ها مستقل و هم توزیع باشند و هم چنین مستقل از N با $P(Z_n = 1) = p$ ، آنگاه N' از حذف تصادفی مستقل نقاط فرایند مبدأ N نتیجه می شود و یک فرایند تنک شده از N نامیده می شود.

گزاره ۱: فرض کنید که $\{X_n\}$ عناصر تصادفی فضای اقلیدسی E_1 هستند به طوری که $\sum I_{X_n}$ یک فرایند نقطه ای است. فرض کنید J_n عناصر تصادفی i.i.d از فضای اقلیدسی E_2 با توزیع احتمال مشترک F هستند. همچنین فرض کنید که فرایند نقاط $\{X_n\}$ و دنباله J_n روی فضای احتمال یکسانی تعریف شده اند و مستقل اند. در این صورت فرایند نقطه ای $\sum I(X_n \in J_n)$ روی $E_1 \times E_2$ دارای اندازه میانگین $\mu \times F$ است، یعنی اگر $A_i \subset E_i; i = 1, 2$ در این صورت

$$\lambda'K'(s) = N'(s) \Rightarrow \lambda'K'(s) = pN(s)$$

$$p\lambda K'(s) = p\lambda K(s) \Rightarrow K'(s) = K(s).$$

حال یک فرایند نقطه ای دو متغیره را در بگیرید که در آن پیشامدها از دو نوع مجزا هستند. فرض کنید $N_j(A)$ تعداد پیشامدهای نوع j در ناحیه A باشد. در این صورت تابعهای شدت مرتبه اول و دوم را به ترتیب به صورت زیر تعریف می کنیم:

$$\lambda_j(x) = \lim_{|dx| \rightarrow 0} \left\{ \frac{E[N_j(dx)]}{|dx|} \right\}; j = 1, 2$$

$$\lambda_{ij}(x, y) = \lim_{|dx| \rightarrow 0, |dy| \rightarrow 0} \left\{ \frac{E[N_i(dx)N_j(dy)]}{|dx||dy|} \right\}.$$

در حالت مانایی و همسانگردی، $\lambda_j(x) = \lambda_j$ ، تعداد مورد انتظار پیشامدهای نوع j در هر واحد مساحت است، و

$$\lambda_{ij}(x, y) = \lambda_{ij}(\|x - y\|)$$

تابع K_{ij} دو متغیره به صورت زیر تعریف می شود:

$$K_{ij}(s) = 2\pi(\lambda_i\lambda_j)^{-1} \int_0^s \lambda_{ij}(u)udu$$

همچنین مشابه قضیه ۱ داریم:

قضیه ۳: $\lambda_j K_{ij}(s)$ تعداد مورد انتظار پیشامدهای نوع j

داخل ناحیه ای به فاصله s از یک پیشامد دلخواه نوع i است.

قضیه ۴: فرض کنید که پیشامدهای نوع ۱ و ۲ به وسیله

برچسب (نشان) زدن تصادفی یک فرایند مانای همسانگرد تولید می شوند، یعنی، پیشامدها به عنوان نوع ۱ یا ۲ در یک دنباله آزمایشهای برنولی مستقل مشخص می شوند. در این صورت:

$$K_{11}(s) = K_{12}(s) = K_{22}(s).$$

این قضیه از قضیه ۲ نتیجه می شود. زیرا تحت برچسب زدن

تصادفی، فرایند پیشامدهای نوع j یک فرایند تنک شده از فرایند نقطه ای تعریف شده از برهمه فرایندهای پیشامدهای نوع ۱ و ۲ است. چون توابع K_{ij} تحت تنک شدن تصادفی ناورد هستند قضیه اثبات می شود.

$$\gamma(u) = \lambda_v(u) - \lambda'$$

تحت مانایی و همسانگردی، $\lambda = \frac{E[N(A)]}{|A|}$ تعداد مورد انتظار

پیشامدها در هر واحد مساحت در هر ناحیه A است. ارائه تعبیر فیزیکی برای تابع شدت مرتبه دوم $\lambda_v(s)$ ساده نیست. در عوض اندازه گشتاور دوم کاهش یافته یا تابع K_j را به صورت زیر تعریف می کنیم:

$$K(s) = 2\pi\lambda^{-2} \int_0^s \lambda_v(u)udu$$

از آنجا که $\lambda_v(\|x\|)/\lambda$ شدت مرتبه اول شرطی پیشامدی در x به شرط وجود یک پیشامد در مبدأ است، نتیجه می گیریم که:

قضیه ۱: $K(s)$ تعداد مورد انتظار پیشامدها در داخل ناحیه ای به فاصله s از یک پیشامد دلخواه است.

اثبات^{۱۱}

این قضیه یک تعبیر شهودی از تابع K به عنوان یک امید ریاضی رامی دهد.

قضیه ۲: $K(s)$ تحت تنک شدن تصادفی ناورد است.

برهان:

موقعی که پیشامدهای یک فرایند به طور مستقل با احتمال p ثبت می شود و با احتمال $1-p$ ثبت نمی شود، در این صورت با استفاده از گزاره ۱ نرخ فرایند پیشامدهای ثبت شده برابر $p\lambda$ است. همچنین تعداد مورد انتظار پیشامدها در داخل ناحیه ای به فاصله s از یک پیشامد دلخواه نیز در p ضرب می شود. پس ناوردایی $K(s)$ از قضیه ۱ نتیجه می شود. یعنی اگر $K'(s)$ اندازه گشتاور دوم کاهش یافته و $\lambda' = p\lambda$ شدت فرایند پیشامدهای ثبت شده باشد و تعداد پیشامدهای مورد انتظار ثبت شده در داخل ناحیه ای به فاصله s از یک پیشامد ثبت شده را با $N'(s)$ و تعداد کل پیشامدهای مورد انتظار در داخل ناحیه ای به فاصله s از یک پیشامد دلخواه را با $N(s)$ نشان دهیم در این صورت از قضیه ۱ داریم:

^{۱۱} اسمعیل زاده، نبز، رساله کارشناسی ارشد آمار (دیماه ۱۳۷۸). مدل سازی فرایند نقطه ای در اپیدمیولوژی زیست محیطی، به راهنمایی دکتر محمد قاسم وحیدی اصل، دانشگاه شهید بهشتی.

۷ برآورد تابع K

فرض کنید که داده های موجود شامل دو نوع پیشامد در ناحیه A و در کل n پیشامد با n_1 تا از نوع ۱ و $n_2 = n - n_1$ تا از نوع ۲ باشند. یک برآوردگر نارایب برای شدت λ_j عبارت است از:

$$\hat{\lambda}_j = \frac{n_j}{|A|}, j = 1, 2.$$

با توجه به قضیه ۱ و قرار دادن مقدار میانگین مشاهده شده به جای مقدار مورد انتظار نظری، یک برآوردگر $\hat{K}_{11}$ به صورت زیر است:

$$n_1^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} I(d_{ij} \leq s)$$

که در آن $d_{ij} = \|x_i - x_j\|$. این برآوردگر یک اریبی ذاتی دارد که دلیل آن وجود اثر حاشیه ای پیشامدهای خارج A است که ثبت نمی شوند. روشهای متعددی برای تصحیح این اریبی پیشنهاد

مراجع

- [1] Barnett Vic and Turkman K.F.(1993). *Statistics for the Environment*. John Wiley and Sons Ltd.
- [2] Daley, D.J. and Vere - jones , D.(1988). *An Introduction to the Theory Of Point Processes*. Springer - Verlag.
- [3] Diggle , P.J. (1983). *Statistical Analysis of Spatial Point Patterns*. London : Academic Press.
- [4] Diggle , P.J. and Chetwynd , A.G.(1991). *Second-order Analysis of Spatial Clustering for Inhomogeneous Population*. Biometric 47 , 1155-1163.
- [5] Diggle , P.J. and Chetwynd ,A.G., Haggkvist.R and Morris ,S.E.(1995). *Second-order Analysis of Space-time Clustering*. Statistical Methods in Medical Research ,4,124-136.
- [6] Resnick I.S.(1992). *Adventures in Stochastic Processes*. Birkhauser Boston.
- [7] Ripley ,B.D.(1976). *The Second-order Analysis of Statinary Point Processes*. Journal of Applied Probability 13,255-266.
- [8] Ripley ,B.D.(1981) *Spatial Statistics*. Wiley and Sons,New York , Chichester.
- [9] Stoyan,D.Kendall,W.S. and Mecke , J.(1995). *Stochastics Geomety and its Application*.Wiley and Sons, New York, Chichester.

شده اند. رایجترین و کاربردی ترین آنها روش ریلی ([۷]) است. ریلی وزنه های w_{ij} را به کار برد. برآوردگر ریلی برای $\lambda_1 K_{11}(s)$ به صورت زیر است

$$n_1^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} w_{ij} I(d_{ij} \leq s)$$

که $w_{ii} = 0$ و برای $i \neq j$ ، $w_{ij} = w(x_i, d_{ij})$ معکوس نسبت محیط دایره ای به مرکز x_i و شعاع d_{ij} است که داخل A قرار می گیرد. در حالت کلی w_{ij} با w_{ji} برابر نیست.

برای تبدیل این تابع به برآوردگری برای $k_{11}(s)$ ، آن را بر $\tilde{\lambda}_1 = (n_1 - 1)/|A|$ تقسیم می کنیم. در این صورت داریم:

$$\hat{K}_{11}(s) = |A| \{n_1(n_1 - 1)\}^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n_1} w_{ij} I(d_{ij} \leq s).$$

به طور مشابه برآوردگر $K_{22}(s)$ به صورت زیر است:

$$\hat{K}_{22}(s) = |A| \{n_2(n_2 - 1)\}^{-1} \sum_{i=n_1+1}^{n_2} \sum_{j=n_1+1}^{n_2} w_{ij} I(d_{ij} \leq s).$$