پانزدهمین سمینار احتمال و فرآیندهای تصادفی
۸ و ۹ شهریور ۱۴۰۴
دانشگاه کردستان

15th Seminar On Probability and Stochastic Processes
August 30-31, 2025
University of Kurdistan

انجمن آمار ایران

یک روش جدید تنک‌سازی فرایندهای نقطه‌ای مبتنی بر مجموعه‌های تصادفی بولی

رضا اسدی [۱] ، مجتبی خزائی [۱] ، مجتبی گنجعلی [۱]

[۱] گروه آمار، دانشگاه شهید بهشتی

چکیده: فرایندهای نقطه‌ای فضایی از مدل‌های آماری برای تحلیل الگوهای نقطه‌ای اند. یکی از روش‌های ساخت فرایندهای نقطه‌ای جدید تنک‌سازی فرایندهای نقطه‌ای موجود است. در این مقاله یک الگوریتم تنک‌سازی جدید مبتنی بر یک مجموعه تصادفی بولی ارائه می‌شود. این الگوریتم را می‌توان تعمیمی از روش تنک‌سازی ارائه شده توسط اشتویان در معرفی فرایندهای نقطه‌ای وابریده دانست. بعضی ویژگی‌های فرایند نقطه‌ای تنک‌شده و فرایند نقاط حذف شده حاصل از اعمال این الگوریتم تنک‌سازی بر یک فرایند نقطه‌ای پواسون همگن بررسی شده و بعضی از آماره‌های خلاصه آن‌ها محاسبه شده است. از روش درستنمایی پالم برای برآورد پارامترها استفاده شده است. در نهایت، مدل‌های فوق برای مدل‌بندی دو مجموعه داده واقعی استفاده می‌شود.

واژه‌های کلیدی: فرایند نقطه‌ای پواسون؛ فرایندهای کاکس؛ مجموعه تصادفی بولی؛ تنک‌سازی وابسته؛ درستنمایی پالم.

کد موضوع‌بندی ریاضی (۲۰۲۰): xxAxx، xxBxx، xxCxx.

۱ مقدمه

فرایندهای نقطه‌ای فضایی از مدل‌های آماری برای مدل‌بندی الگوهای نقطه‌ای اند. به عنوان مثال، از فرایندهای نقطه‌ای فضایی برای مدل‌بندی محل انواع درختان جنگلی، محل وقوع زمین لرزه‌ها، لانه پرندگان و ... می‌توان استفاده کرد. رفتار الگوهای نقطه‌ای به سه دسته رفتار خوشه‌ای، رفتار تصادفی محض و رفتار منظم یا دفعی تقسیم‌بندی می‌شود. متناسب با رفتار هر الگوی نقطه‌ای دسته‌ای از مدل‌ها برای مدل‌بندی الگوهای نقطه‌ای ارائه شده است. روش‌های زیادی از جمله اعمال تبدیلات و یا اصلاح فرایندهای نقطه‌ای برای ساخت فرایند نقطه‌ای جدید وجود دارند. چهار رویکرد کلی برای این کار عبارتند از: نگاشت، برهم نهی، خوشه‌بندی و تنک‌سازی.

در رویکرد تنک‌سازی نقاط فرایند اولیه براساس یک الگوریتم حذف می‌شوند. در صورتی که احتمال حذف نقاط فرایند اولیه از هم مستقل باشند، تنک‌سازی مستقل و در صورتی که احتمال حذف نقاط فرایند اولیه به هم وابسته باشند، تنک‌سازی وابسته نامده می‌شوند.

[۱] سخنران، rezaasadi220@gmail.com

در این مقاله ابتدا یک الگوریتم تنک‌سازی معرفی می‌شود که در آن احتمال‌های حذف براساس تحققی از یک مدل بولی مشخص می‌شوند. سپس فرایندهای تنک‌شده و نقاط حذف شده حاصل از اعمال این الگوریتم روی یک فرایند اولیه پواسون همگن بررسی می‌شود. الگوریتم تنک‌سازی ارائه شده در این مقاله را می‌توان تعمیمی از الگوریتم تنک‌سازی ارائه شده توسط اشتویان (۱۹۷۹) و مولر و لاوانسیر (۲۰۱۶) دانست با این تفاوت که در اینجا احتمال حذف درون تحقق مجموعه تصادفی بولی می‌تواند هر مقداری بین صفر و یک باشد در حالی که اشتویان (۱۹۷۹) و مولر و لاوانسیر (۲۰۱۶) تنها احتمال حذف نقاط خارج از مجموعه تصادفی بولی را برابر یک در نظر گرفته‌اند. یک مجموعه تصادفی بولی به صورت زیر تعریف می‌شود.

تعریف ۱.۱. (اشتویان ، ۱۹۷۹) یک مجموعه تصادفی بولی از قرار دادن تحقق‌های یک مجموعه تصادفی بسته، معروف به دانه، روی نقاط یک فرایند نقطه‌ای پواسون، معروف به ریشه، حاصل می‌شود. در صورتی که Z فرایند پواسون مانا روی $\mathbb{R}^2$ با شدت θ و مستقل از M_s ها؛ یعنی مجموعه‌های تصادفی بسته مستقل و هم توزیع با M_o به‌طوری‌که $E\{\sup||x||, x \in M_o\} < \infty$ باشد، آنگاه مجموعه تصادفی بولی حاصل از دانه بسته تصادفی M_o و فرایند نقطه‌ای Z برابر است با: $\Xi = \bigcup_{s\in Z}(s + M_s)$

۲ معرفی الگوریتم تنک‌سازی پیشنهادی و فرایندهای مربوطه

فرایند نقطه‌ای اولیه Y و فرایند نقطه‌ای Z به عنوان فرایندی که بر اساس آن Y تنک می‌شود را در نظر بگیرید. در ادامه فرض می‌کنیم که Y و Z بر $\mathbb{R}^2$ تعریف شده‌اند. حول هر نقطه از فرایند نقطه‌ای فضایی Z دایره‌ای با شعاع r را در نظر گرفته و احتمال حذف نقطه‌ای از فرایند اولیه درون هر دایره را مستقل از سایر دوایر برابر با q در نظر می‌گیریم. به این ترتیب یک الگوریتم تنک‌سازی حاصل می‌شود که در آن:

۱- در ناحیه خارج از اجتماع دوایر احتمال حذف صفر است.

۲- در نواحی درون یک دایره احتمال حذف $q = ۱ - p$ است.

۳- در نواحی مشترک بین k دایره احتمال حذف $۱ - p^k$ است.

در صورتی که Z فرایند پواسون همگن و همسانگرد با تابع شدت θ باشد و $\Xi = \bigcup_{s\in Z} b(s, r)$ تعریف شود، یک مجموعه تصادفی بولی با دانه‌های دایره‌ای شکل به شعاع r است. الگوریتم تنک‌سازی فوق را می‌توان به صورت یک الگوریتم تنک‌سازی روی تحقق‌های این مجموعه بولی به صورت زیر تعریف کرد:

۱- در نواحی که خارج از مجموعه تصادفی بولی Ξ است نقاط Y حذف نمی‌شوند.

۲- در نواحی که محل تحقق فقط یک دانه از مجموعه تصادفی بولی Ξ است نقاط Y با احتمال $q = ۱ - p$ حذف می‌شوند.

۳- در نواحی که محل اشتراک k دانه از مجموعه تصادفی بولی Ξ است نقاط فرایند اولیه Y با احتمال $q = ۱ - p^k$ حذف می‌شوند.

در قضیه زیر توزیع حاشیه‌ای و توام احتمال بقای نقاط محاسبه می‌شود. اثبات قضیه‌های ارائه شده در این مقاله در ضمایم ارائه شده است.

قضیه ۱.۲. در الگوریتم تنک‌سازی پیشنهادی اگر $\Pi(\zeta)$ احتمال بقا در $\zeta \in \mathbb{R}^2$ باشد آن گاه:

(i) اگر $A_\zeta = b(\zeta, r)$ باشد برای هر $\zeta \in \mathbb{R}^۲$:

$$P[\Pi(\zeta) = p^k] = P[N(A_\zeta) = k] = \frac{e^{-\theta\pi r^۲}(\theta\pi r^۲)^k}{k!}, \quad k = ۰, ۱, ۲, \ldots, \tag{۱.۲}$$

(ii) توزیع توام احتمال بقا برای هر $\zeta_۱, \zeta_۲ \in \mathbb{R}^۲$ برابر است با:

$$P[\Pi(\zeta_۱) = p^n, \Pi(\zeta_۲) = p^m] = \begin{cases} P[\Pi(\zeta_۱] = p^n]P[\Pi(\zeta_۲) = p^m], & ||\zeta_۱ - \zeta_۲|| > ۲r, \\ C[\Pi(\zeta_۱, \zeta_۲)], & ||\zeta_۱ - \zeta_۲|| \leqslant ۲r, \end{cases} \tag{۲.۲}$$

که در آن

$$\begin{aligned} C[\Pi(\zeta_۱, \zeta_۲)] =& e^{-\theta(۲\pi r^۲ - A(t,r))}[\theta(\pi r^۲ - A(t,r))]^{m+n} \\ & \times \sum_{k=۰}^{\min(m,n)} \frac{۱}{k!(n-k)!(m-k)!}\left[\frac{A(t,r)}{\theta(\pi r^۲ - A(t,r))^۲}\right]^k, \end{aligned} \tag{۳.۲}$$

و $m, n \in \mathbb{N} \cup \{۰\}, r > ۰$

$$A(t,r) = ۲r^۲ Arc\cos(\frac{t}{۲r}) - \frac{t}{۲}[۴r^۲ - t^۲]^{\frac{۱}{۲}}, \tag{۴.۲}$$

مساحت ناحیه مشترک بین دو دایره به شعاع r با فاصله t ، ($t < ۲r$) است.

در صورتی که فرایند اولیه، Y ، یک فرایند نقطه‌ای پواسون با شدت δ باشد از اعمال الگوریتم فوق بر Y دو فرایند تنک‌شده‌ی X، Y، و نقاط حذف شده از Y، $Y - X$، فرایندهای نقطه‌ای تعریف می‌کنند که به پارامترهای (δ, θ, r, p) بستگی دارد. در ادامه ویژگی‌ها و رفتار این فرایندها بررسی می‌شود.

۳ آماره‌های خلاصه

آماره‌های خلاصه به عنوان ابزاری برای تحلیل اکتشافی داده‌ها، آزمون و بررسی اعتبار مدل توسط آماردان‌ها مورد استفاده قرار می‌گیرند. در تحلیل فرایندهای نقطه‌ای فضایی ویژگی‌های مرتبه اول و دوم متغیر شمارشی $N(B); B \subseteq \mathbb{R}^۲$ اطلاعات سودمندی در مورد رفتار فرایند نقطه‌ای به تحلیل‌گر می‌دهند. تابع شدت مرتبه اول و دوم این ویژگی‌ها را تبیین می‌کنند. اگر $\mu(B) = E[N(B)]; B \subseteq \mathbb{R}^۲$ ، $\mu(.)$ اندازه شدت نامیده می‌شود. در صورتی که بتوان نوشت $\mu(B) = \int_B \lambda(\zeta)d\zeta; B \subseteq \mathbb{R}^۲$ نوشت که در آن $\lambda(.)$ یک تابع نامنفی است، $\lambda(.)$ تابع شدت نامیده می‌شود. اگر $\alpha^{(۲)}(C) = E[\sum_{\zeta,\eta \in X}^{\neq} I[(\zeta, \eta) \in C]]; C \subseteq \mathbb{R}^۲ \times \mathbb{R}^۲$ ، $\alpha^{(۲)}(.)$ اندازه شدت فاکتوریلی مرتبه دوم نامیده می‌شود. بعلاوه اگر $\alpha^{(۲)}(C) = \int\int I[(\zeta, \eta) \in C]\lambda^{(۲)}(\zeta, \eta)d\zeta d\eta$ که در آن $\lambda^{(۲)}(.,.)$ یک تابع نامنفی است به $\lambda^{(۲)}(.,.)$ تابع شدت مرتبه دوم گفته می‌شود. در این‌جا با توجه به این‌که فرایند اولیه‌ی Y توسط یک میدان تصادفی تنک می‌شود لذا هر دو فرایند X و $Y - X$ فرایندهای کاکس هستند (مولر و واگه‌پیترسن ، ۲۰۰۳).

قضیه ۱.۳. برای فرایند تنک‌شده‌ی X، تابع شدت $\lambda(.)$ و تابع شدت مرتبه دوم $\lambda^{(۲)}(.,.)$ برابر است با:

$$(i)\ \lambda(u) = \delta e^{-\theta(۱-p)\pi r^۲}.$$

$$(ii)\ \lambda^{(۲)}(u,v) = \begin{cases} \delta^۲ e^{-۲\pi r^۲\theta(۱-p)+\theta A(t,r)(۱-p)^۲}, & \|t\| \le ۲r, \\ \delta^۲ e^{-۲(۱-p)\theta\pi r^۲}, & \|t\| > ۲r, \end{cases}$$

که در آن $t = ||u-v||$ است.

تابع همبستگی زوجی یکی دیگر از آماره‌هایی است که ویژگی‌های مرتبه دوم فرایندهای نقطه‌ای را بررسی می‌کند. تابع همبستگی زوجی به صورت $g(\zeta,\eta) = \frac{\lambda^{(۲)}(\zeta,\eta)}{\lambda(\zeta)\lambda(\eta)}$ تعریف می‌شود که در آن برای $a \geqslant ۰$ قرار می‌دهیم $\frac{a}{۰} = ۰$. برای فرایندهای مانا می‌توان نوشت $g(\zeta,\eta) = g(\zeta-\eta)$ و برای فرایندهای مانا و همسانگرد $g(\zeta,\eta) = g(t)$ که در آن $t = ||\zeta-\eta||$. در حالت کلی اگر $t \to \infty$ آنگاه $g(t) \to ۱$. ولی برای t های کوچک، $g(t) < ۱$ نشان‌دهنده فرایند با رفتار دفعی، $g(t) = ۱$ نشان‌دهنده فرایند با رفتار تصادفی محض و $g(t) > ۱$ نشان‌دهنده فرایند با رفتار خوشه‌ای است.

قضیه ۲.۳. تابع همبستگی زوجی برای فرایندهای X و $Y-X$ برابر است با:

$$(i)\ g_X(t) = \begin{cases} e^{\theta A(t,r)(۱-p)^۲}, & t \le ۲r, \\ ۱, & t > ۲r, \end{cases}$$

$$(ii)\ g_{Y-X}(t) = ۱ + \frac{a^۲}{(۱-a)^۲}(g_X(t)-۱),$$

که در آن $t = ||u-v||, a = E[\Pi(o)]$.

(i) نشان می‌دهد که $g_X(t) \geqslant ۱$. و از (ii) نتیجه می‌شود که $g_{Y-X}(t) \geqslant ۱$. نتایج فوق بیانگر این است که هر دو فرایند X و $Y-X$ دارای رفتار خوشه‌ای هستند. تابع شدت پالم یکی دیگر از آماره‌هایی است که در استنباط آماری فرایندهای نقطه‌ای کاربرد زیادی دارد. در صورتی که x نقطه‌ای دلخواه در $\mathbb{R}^۲$ با فاصله t از مبدا، o، باشد. آنگاه نرخ تحقق در x به شرط آنکه o یک نقطه از فرایند باشد برای مجموعه لبگ با اندازه بسیار کوچک dx برابر است با: $\lambda_o(x)dx = P[N(dx) = ۱|N\{o\} = ۱]$ به $\lambda_o(x)$ شدت پالم در نقطه x گفته می‌شود. در صورتی‌که فرایند مانا و همسانگرد باشد تابع شدت پالم صرفا به t وابسته است و می‌توان آن را به صورت $\lambda_o(t)$ نوشت. همچنین برای فرایندهای مانا و همسانگرد $\lambda_o(t) = \lambda g(t)$ (تاناکا و همکاران ، ۲۰۰۸).

قضیه ۳.۳. تابع شدت پالم برای فرایندهای X و $Y-X$ برابر است با:

$$(i)\ \lambda_{o,X}(t) = \begin{cases} \delta e^{-\theta(۱-p)[\pi r^۲ - A(t,r)(۱-p)]}, & t \le ۲r, \\ \delta e^{-\theta(۱-p)\pi r^۲}, & t > ۲r, \end{cases}$$

$$(ii)\ \lambda_{o,Y-X}(t) = \delta[۱ - e^{-\theta q\pi r^۲}]\{۱ + \frac{a^۲}{(۱-a)^۲}(g_X(t)-۱)\}.$$

۴ برآورد پارامترها و ارزیابی مدل

روش‌های زیادی برای برآورد پارامترهای فرایندهای نقطه‌ای وجود دارد، از جمله حداقل تضادها، ماکسیمم درستنمایی، شبه درستنمایی و درستنمایی پالم. در این مقاله برای برآورد پارامترهای فرایندهای X و $Y - X$ از روش درستنمایی پالم استفاده می‌شود. از روش‌های متداول بررسی نیکویی برازش در فرایندهای نقطه‌ای بررسی پوشش‌ها است. این روش مبتنی بر محاسبه و رسم نمودار چندک‌های آماره خلاصه به ازای t های متفاوت است. از روش بوت استرپ برای محاسبه چندک‌ها می‌توان استفاده کرد. این روش مبتنی‌بر این است که اگر مدل برازش داده شده صحیح باشد بروردگر ناپارامتریک آماره‌های خلاصه الگوی مشاهده شده به ازای t های متفاوت در اغلب اوقات رفتاری مشابه مدل مورد نظر (با پارامترهای برآورد شده) دارد.

۱.۴ درستنمایی پالم

درستنمایی پالم یک روش تقریبی برآورد پارامترهای فرایندهای نقطه‌ای است. این روش مبتنی بر فرایند تفاضل است. برای فرایند نقطه‌ای P که بر پنجره محدب W مشاهده شده است، فرایند تفاضل به صورت $D_{P,W} = \{x - y; x, y \in P_W, \ \ s.t. \ \ x \neq y\}$ تعریف می‌شود. برای فرایند مانا و همسانگرد P، فرایند تفاضل یک فرایند همسانگرد با تابع شدت $N(W)\lambda_o(t)$ است که در آن $\lambda_o(t)$ تابع شدت پالم فرایند P است. همچنین در صورتی که فرض شود فرایند P به خوبی توسط یک فرایند پواسون تقریب زده می‌شود آنگاه تابع درستنمایی فرایند برابر است با:

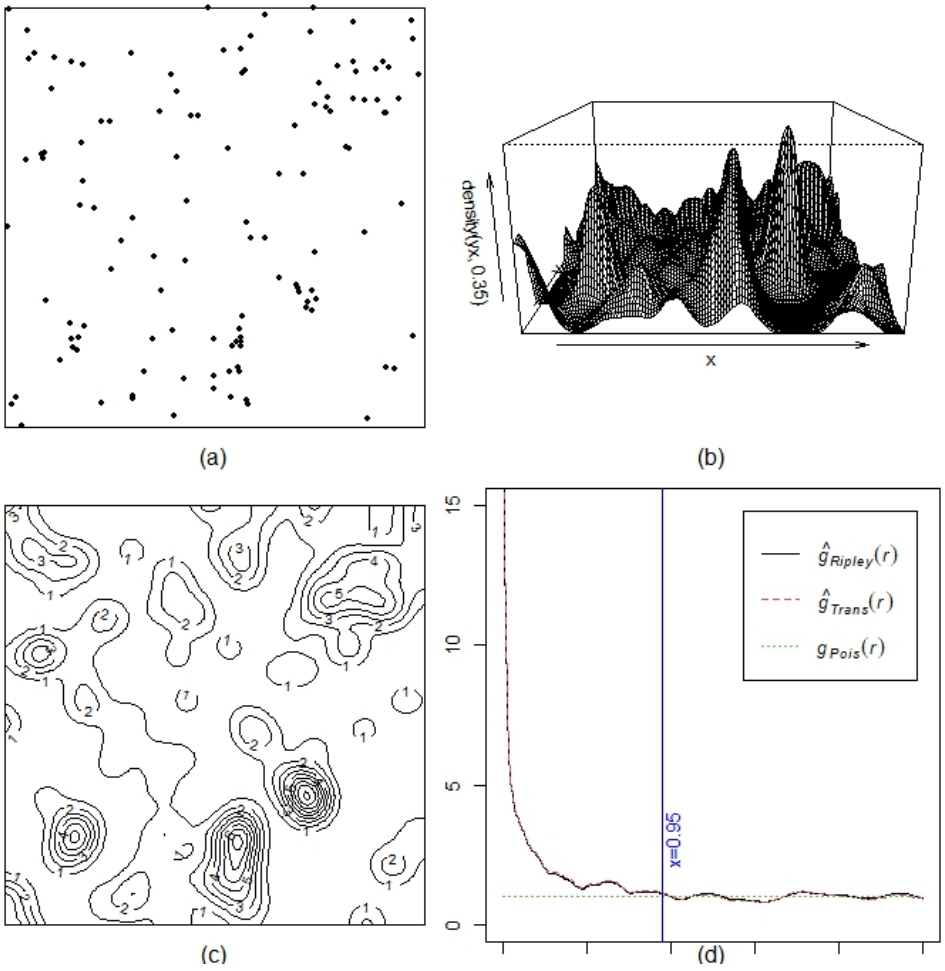
$$\log L(\delta, \theta, p, r, t) = \sum_{\{i,j;i\neq j;t_{ij}<T\}} \log\{N(W)\lambda_o(t_{ij})\} - N(W)\int_{\circ}^{T} \lambda_o(t)۲\pi t dt. \qquad (۱.۴)$$

که در آن t_{ij} فاصله بین نقاط p_i و p_j است و T آستانه‌ای است که فرض می‌شود نقاط با فاصله بیشتر از آن از هم مستقل هستند. به $(\hat{\delta}, \hat{\theta}, \hat{r}, \hat{p})$ که عبارت فوق را ماکسیمم کند برآوردگر ماکسیمم درستنمایی پالم گفته می‌شود. برای آشنایی بیشتر به تاناکا و همکاران (۲۰۰۸) به مراجعه شود.

۵ کاربرد

در این بخش دو مجموعه داده واقعی بررسی می‌شود و فرایندهای X و $Y - X$ به آن‌ها برازش داده می‌شود.

مثال ۱.۵. الگوی نقطه‌ای نهال‌های کاج که در بسته $spatstat$ نرم افزار $\mathbb{R}$ تحت عنوان $finpines$ موجود است، محل قرار گرفتن ۱۲۶ نهال کاج به همراه قطر و ارتفاع آن‌ها در جنگلی در فنلاند بر پنجره $[-۸, ۲] \times [-۵, ۵]$ را ثبت کرده است. در این‌جا صرفا محل قرار گرفتن نهال‌ها بررسی و مدل‌بندی می‌شود. در شکل ۱ نمودار (a) الگوی نقطه‌ای، (b) نمودار ۳ بعدی تابع شدت، (c) منحنی‌های تراز تابع شدت، (d) نمودار تابع همبستگی زوجی، نهال‌های کاج را نشان می‌دهند. نمودار تابع همبستگی زوجی نشان‌دهنده رفتار خوشه‌ای الگوی نقطه‌ای است همچنین منحنی‌های تراز ونمودار ۳ بعدی تابع شدت نشان‌دهنده این است که الگوی نقطه‌ای تمایل به تجمع در نواحی دایره‌ای شکل دارند. لذا برازش فرایند $Y - X$ به الگوی نقطه‌ای فوق پیشنهاد می‌شود. برای برازش فرایند $Y - X$ به الگوی نقطه‌ای فوق براساس روش درستنمایی پالم پارامترهای مدل فوق برآورد می‌گردد. با توجه به این‌که برآورد ناپارامتریک تابع همبستگی زوجی برای $t \geqslant ۰/۹۵$ کوچکتر از ۱.۱ است. (شکل ۱ مشاهده شود.) لذا T برابر با ۰/۹۵ در نظر گرفته می‌شود. همچنین



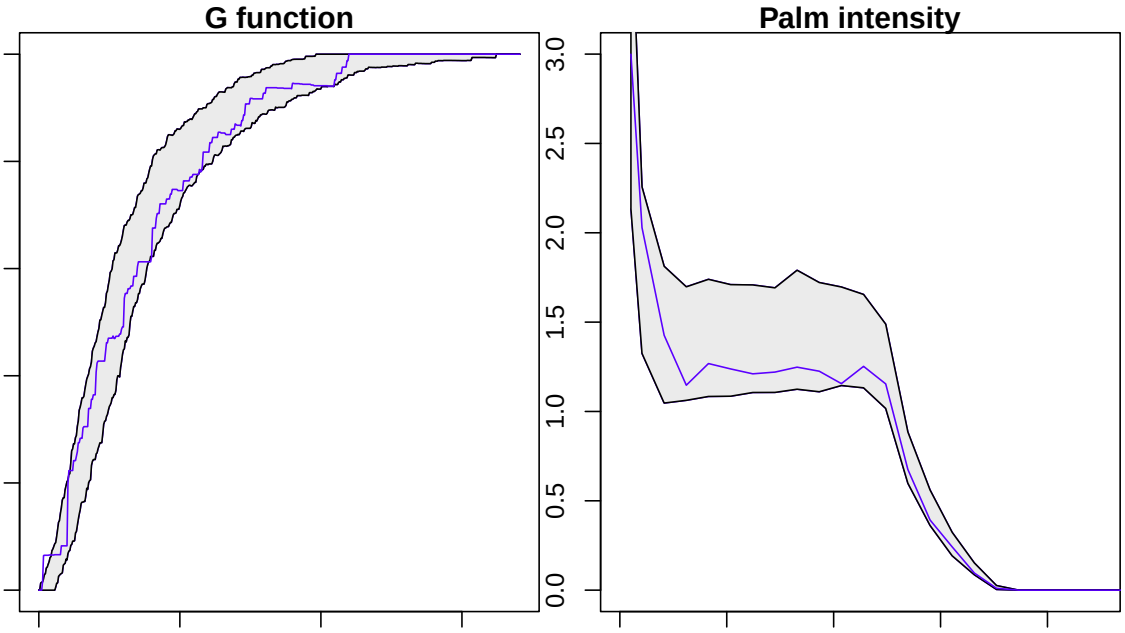
شکل ۱: (a) الگوی نقطه‌ای، (b) نمودار ۳ بعدی تابع شدت، (c) منحنی‌های تراز تابع شدت، (d) تابع همبستگی زوجی، نهال‌های کاج را نشان می‌دهد.

با جایگذاری $\lambda_{o,Y-X}(t)$ در (۱.۴) تابع درستنمایی پالم حاصل می‌شود. برآوردگر ماکسیمم درستنمایی پالم فرایند فوق در جدول ۱ ارائه شده است.

پارامتر	δ	θ	p	r
برآورد	۵/۶۸۳	۱/۷۱۳	۰/۰۴۶	۰/۲۳۸۸

جدول ۱: برآوردهای پارامترهای حاصل از برازش فرایند $Y-X$ به الگوی نهال‌های کاج

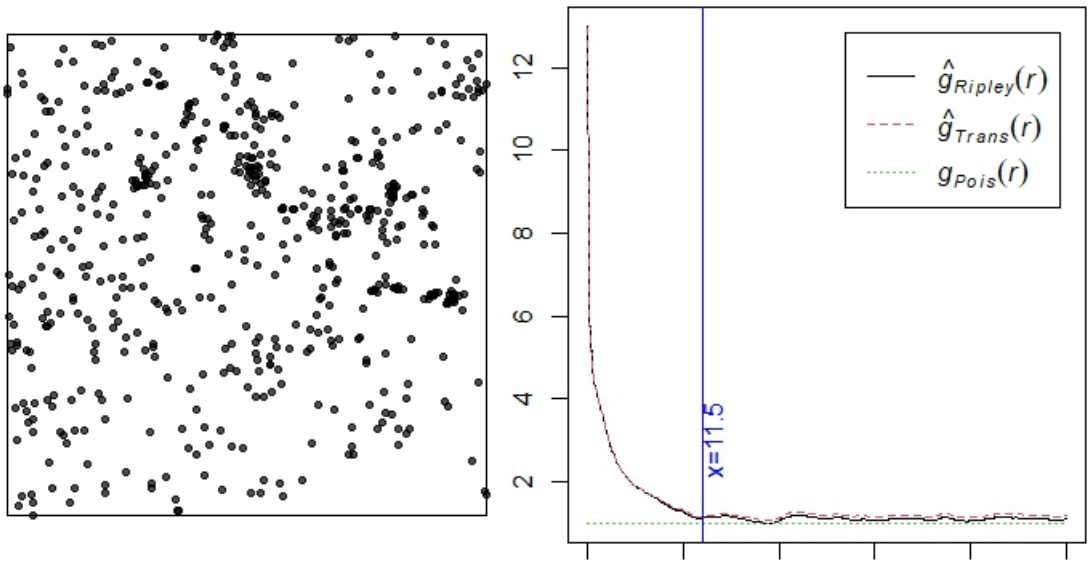
برای بررسی نیکویی برازش مدل برازش داده شده پوشش آماره‌های خلاصه بررسی می‌شود. از آن‌جایی که همه پارامترهای مدل در ساختار تابع شدت پالم وجود دارد پوشش تابع شدت پالم بررسی می‌شود همچنین پوشش تابع G که یکی از آماره‌های خلاصه مبتنی بر فاصله بین نقاط است نیز بررسی می‌شود. در شکل ۲ نمودار پوشش تابع شدت پالم رسم شده است که موید این است که فرایند برازش داده شده به خوبی الگوی نقطه‌ی فوق را مدل‌بندی می‌نماید. همچنین نمودار پوشش تابع G نیز رسم شده است. از آن‌جایی که روش



شکل ۲: نمودار پوشش تابع شدت پالم و تابعG برای مدل برازش داده شده

درستنمایی پالم یک روش تقریبی برآورد پارامترها است و این‌که، در اغلب نقاط تابع G مشاهده شده در فاصله اطمینان ۰٫۹۵ قرار گرفته است لذا می‌توان نتیجه گرفت که مدل برازش داده شده با پارامترهای فوق برازش مناسبی به الگوی فوق دارد.

مثال ۲.۵. الگوی نقطه‌ای درختان کاج برگ بلند در ناحیه‌ای در جنوب جورجیا (ایالات متحده امریکا) بر پنجره $[۰, ۲۰۰] \times [۰, ۲۰۰]$ محل قرار گرفتن و قطر ۵۸۴ درخت کاج برگ بلند را نشان می‌دهد. در این‌جا صرفا محل قرار گرفتن درخت‌های فوق بررسی می‌شود. این داده‌ها در پکیج $spatstat$ نرم افزار $\mathbb{R}$ با عنوان $longleaf$ موجود است.



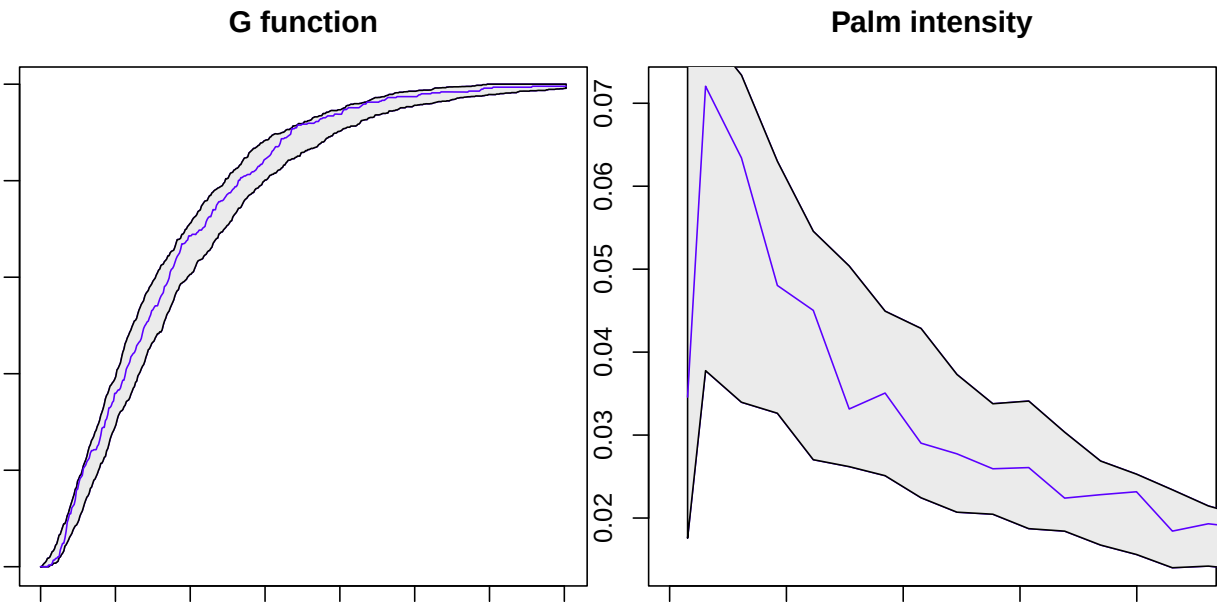
شکل ۳: نمودار الگوی نقطه‌ای محل قرارگرفتن درختان کاج برگ بلند و تابع همبستگی زوجی آن

در شکل ۳ نمودار محل قرار گرفتن درختان کاج و تابع همبستگی زوجی آن‌ها نشان داده شده. در شکل فوق به نظر می‌رسد بعضی نواحی دایره‌ای شکل تنک‌تر هستند. لذا برازش فرایند X به الگوی فوق پیشنهاد می‌شود. برای برازش برآورد پارامترهای مدل توسط روش درستنمایی پالم با توجه به رفتار تابع همبستگی زوجی در شکل ۳ و از آن‌جایی که برای $t \geqslant ۱۱/۵$ تابع همبستگی زوجی در اغلب نقاط کمتر از ۱/۱ است لذا T در رابطه (۱.۴) برابر با ۱۱/۵ در نظر می‌گیریم. با جایگذاری تابع شدت پالم فرایند X، $\lambda_{o,X}(t)$، از ۳.۳ در تابع درستنمایی پالم (۱.۴) برآورد درستنمایی پارامترهای فرایند که در جدول ۲ ارائه شده اند بدست می‌آید.

پارامتر	δ	θ	p	r
برآورد	۰/۵۴۴	۰/۱۰۹	۰/۳۴۹	۵/۵۱۱

جدول ۲: برآوردگر درستنمایی پالم حاصل از برازش فرایند X به الگوی درختان کاج برگ بلند

از طرفی در فرایند X با توجه به رفتار تابع همبستگی زوجی براساس رابطه $\hat{r}_T = \min_t\{g(۲t) < ۱ + \varepsilon \quad \& \quad g(۲t) > ۰\}$ پارامتر r را نیز می‌توان برآورد کرد. برای بهبود و تعدیل برآوردگر پارامتر r، آن را می‌توان به کمک رابطه $\hat{r}_{imp} = a\hat{r}_{Palm} + (۱-a)\hat{r}_T$ برآورد نمود. با در نظر گرفتن $\varepsilon = ۰/۱$ و $a = ۰/۴$ آنگاه $\hat{r}_{imp} = ۵/۶۵۴$.



شکل ۴: نمودار پوشش برای تابع شدت پالم و تابع G

برای بررسی نیکویی برازش مدل برازش داده شده، پوشش آماره‌های خلاصه بررسی می‌شود. برای این منظور پوشش تابع شدت پالم و تابع G بررسی می‌شود. در شکل ۴ نمودار پوشش برای تابع شدت پالم و تابع G رسم شده است. پوشش آماره‌های خلاصه نشان‌دهنده این است که دلیلی برای عدم برازش مناسب مدل برازش داده شده وجود ندارد و مدل برازش داده شده عملکرد مناسبی دارد.

۶ بحث و نتیجه‌گیری

در این مقاله یک الگوریتم جدید تنک‌سازی ارائه شد. همچنین فرایند تنک‌شده و نقاط تنک شده حاصل از اعمال الگوریتم فوق بر فرایند پواسون همگن بررسی شد و روشی برای برازش مدل‌های فوق ارائه شد. برای مطالعات آتی می‌توان به مطالعه فرایند تنک‌شده حاصل از اعمال الگوریتم فوق بر سایر فرایندهای اولیه اقدام نمود. همچنین پیشنهاد می‌شود سایر روش‌های برآورد پارامترهای مدل بررسی و با روش درستنمایی پالم مقایسه گردد.

مراجع

Stoyan, D. (1979), Interrupted point processes, *Biometrical Journal*, **21(7)**, 607-610.

Lavancier, F. and Møller, J. (2016), Modelling aggregation on the large scale and regularity on the small scale in spatial point pattern datasets, *Scandinavian Journal of Statistics*, **43(2)**, 587-609.

Møller, J., *Statistical inference and simulation for spatial point processes*, CRC Press.

Tanaka, U., Y. Ogata, and D. Stoyan (2008), Parameter estimation and model selection for neyman-scott point processes, *Biometrical Journal: Journal of Mathematical Methods in Bio- sciences* , **50(1)**, 43-57.

Neyman, J. and E. L. Scott (1958), Statistical approach to problems of cosmology, of the Royal Statistical Society: Series B (Methodological), **20(1)**, 1-29.

۷ ضمایم

قضیه ۱.۲ **اثبات.** (i) با استفاده از اینکه Z فرایند نقطه‌ای پواسون است به دست می‌آید. (ii) تساوی به ازای $||\zeta_۱ - \zeta_۲|| > ۲r$ با استفاده از مستقل بودن تعداد نقاط در ناحیه‌های مجزا در فرایند نقطه‌ای پواسون بدست می‌آید. در حالتی که $||\zeta_۱ - \zeta_۲|| \leqslant ۲r$ داریم:

$$
\begin{aligned}
P[\Pi(\zeta_۱) = p^n, \Pi(\zeta_۲) = p^m] &= P[N(A_{\zeta_۱}) = n, N(A_{\zeta_۲}) = m], \\
&= \sum_{k=۰}^{\min(m,n)} P[N(A_{\zeta_۱} - A_{\zeta_۲}) = n-k, N(A_{\zeta_۱} \cap A_{\zeta_۲}) = k, N(A_{\zeta_۲} - A_{\zeta_۱}) = m-k], \\
&= \sum_{k=۰}^{\min(m,n)} P[N(A_{\zeta_۱} - A_{\zeta_۲}) = n-k] P[N(A_{\zeta_۱} \cap A_{\zeta_۲}) = k] P[N(A_{\zeta_۲} - A_{\zeta_۱}) = m-k], \\
&= \sum_{k=۰}^{\min(m,n)} \frac{e^{-\theta||A_{\zeta_۱} - A_{\zeta_۲}||}[\theta||A_{\zeta_۱} - A_{\zeta_۲}||]^{n-k}}{(n-k)!} \\
&\times \frac{e^{-\theta||A_{\zeta_۱} \cap A_{\zeta_۲}||}[\theta||A_{\zeta_۱} \cap A_{\zeta_۲}||]^{k}}{k!} \times \frac{e^{-\theta||A_{\zeta_۲} - A_{\zeta_۱}||}[\theta||A_{\zeta_۲} - A_{\zeta_۱}||]^{m-k}}{(m-k)!}, \\
&= e^{-\theta[۲\pi r^۲ - A(t,r)]}[\theta(\pi r^۲ - A(t,r))]^{m+n} \sum_{k=۰}^{\min(m,n)} \frac{۱}{k!(n-k)!(m-k)!} \left[\frac{A(t,r)}{\theta(\pi r^۲ - A(t,r))^۲}\right]^k.
\end{aligned}
\tag{۱.۷}
$$

تساوی آخر با کمی محاسبات جبری بدست می‌آید. □

قضیه ۱.۳ **اثبات.** برای (i) با توجه به کاکس بودن فرایند X داریم:

$$
\begin{aligned}
\lambda(u) = E[\delta\Pi(u)] = \delta E[\Pi(u)] = \delta \sum_{k=۰}^{\infty} p^k P(\Pi(u) = p^k) = \delta \sum_{k=۰}^{\infty} p^k \frac{e^{-\mu(A_u)}\mu(A_u)^k}{k!} \\
= \delta e^{-\mu(A_u)} \sum_{k=۰}^{\infty} \frac{[p\mu(A_u)]^k}{k!} = \delta e^{-\mu(A_u)(۱-p)} = \delta e^{-(۱-p)\pi\theta r^۲}
\end{aligned}
\tag{۲.۷}
$$

در تساوی چهارم از رابطه (۱.۲) استفاده شده و $A_u = B(u,r)$. همچنین با توجه به اینکه $Z \sim Poisson(\theta)$ لذا $\mu(A_u) = \theta\pi r^۲$ است.

برای (ii) با توجه به کاکس بودن فرایند X داریم:

$$
\lambda^{(۲)}(u,v) = E[\delta\Pi(u)\delta\Pi(v)] = \delta^۲ E[\Pi(u)\Pi(v)]
$$

در صورتی‌که $||\zeta_۱ - \zeta_۲|| \geqslant ۲r$ از استقلال ناشی از الگوریتم تنک‌سازی داریم:

$$
E[\Pi(\zeta_۱)\Pi(\zeta_۲)] = E[\Pi(\zeta_۱)]E[\Pi(\zeta_۲)].
$$

و با توجه به قسمت (i) نتیجه حاصل می‌شود.

برای $|\zeta_۱ - \zeta_۲| < ۲r$ با استفاده از رابطه (۲.۲) داریم:

$$E[\Pi(\zeta_۱)\Pi(\zeta_۲)] = \sum_{m,n=۰}^{\infty} p^{n+m} P[\Pi(\zeta_۱) = p^n, \Pi(\zeta_۲) = p^m]$$

$$= \sum_{m,n=۰}^{\infty} p^{n+m} e^{-\theta[۲\pi r^۲ - A(t,r)]} [\theta(\pi r^۲ - A(t,r))]^{m+n} \sum_{k=۰}^{\min(m,n)} \frac{۱}{k!(n-k)!(m-k)!} \left[\frac{A(t,r)}{\theta(\pi r^۲ - A(t,r))^۲}\right]^k$$

$$= e^{-\theta[۲\pi r^۲ - A(t,r)]} \sum_{m,n=۰}^{\infty} p^{n+m} [\theta(\pi r^۲ - A(t,r))]^{m+n} \sum_{k=۰}^{\min(m,n)} \frac{۱}{k!(n-k)!(m-k)!} \left[\frac{A(t,r)}{\theta(\pi r^۲ - A(t,r))^۲}\right]^k$$

$$= e^{-\theta[۲\pi r^۲ - A(t,r)]} \sum_{k=۰}^{\infty} \sum_{n=k}^{\infty} \sum_{m=n}^{\infty} \frac{[p\theta(\pi r^۲ - A(t,r))]^{m+n}}{k!(n-k)!(m-k)!} \left[\frac{A(t,r)}{\theta(\pi r^۲ - A(t,r))^۲}\right]^k$$

$$+ e^{-\theta[۲\pi r^۲ - A(t,r)]} \sum_{k=۰}^{\infty} \sum_{m=k}^{\infty} \sum_{n=m+۱}^{\infty} \frac{[p\theta(\pi r^۲ - A(t,r))]^{m+n}}{k!(n-k)!(m-k)!} \left[\frac{A(t,r)}{\theta(\pi r^۲ - A(t,r))^۲)}\right]^k$$

$$= e^{-\theta[۲\pi r^۲ - A(t,r)]} \sum_{k=۰}^{\infty} \frac{۱}{k!} [\theta p^۲ A(t,r)]^k \sum_{n=۰}^{\infty} \frac{[p\theta(\pi r^۲ - A(t,r))]^n}{n!} \sum_{m=n}^{\infty} \frac{[p\theta(\pi r^۲ - A(t,r))]^m}{m!}$$

$$+ e^{-\theta[۲\pi r^۲ - A(t,r)]} \sum_{k=۰}^{\infty} \frac{۱}{k!} [\theta p^۲ A(t,r)]^k \sum_{m=۰}^{\infty} \frac{[p\theta(\pi r^۲ - A(t,r))]^m}{m!} \sum_{n=m+۱}^{\infty} \frac{[p\theta(\pi r^۲ - A(t,r))]^n}{n!}$$

$$= e^{-۲\pi r^۲\theta(۱-p)+\theta A(t,r)(۱-p)^۲} \sum_{n=۰}^{\infty} \frac{e^{-p\theta[\pi r^۲ - A(t,r)]} [p\theta(\pi r^۲ - A(t,r))]^n}{n!}$$

$$\times \sum_{m=n}^{\infty} \frac{e^{-p\theta[\pi r^۲ - A(t,r)]} [p\theta(\pi r^۲ - A(t,r))]^m}{m!}$$

$$+ e^{-۲\pi r^۲\theta(۱-p)+\theta A(t,r)(۱-p)^۲} \sum_{m=۰}^{\infty} \frac{e^{-p\theta[\pi r^۲ - A(t,r)]} [p\theta(\pi r^۲ - A(t,r))]^m}{m!}$$

$$\times \sum_{n=m+۱}^{\infty} \frac{e^{-p\theta[\pi r^۲ - A(t,r)]} [p\theta(\pi r^۲ - A(t,r))]^n}{n!},$$

اگر $T_۱, T_۲$ متغیرهای تصادفی مستقل و هم توزیع با $pos(p\theta(\pi r^۲ - A(t,r)))$ باشند آنگاه :

$$E[\Pi(\zeta_۱)\Pi(\zeta_۲)] = e^{-۲\pi r^۲\theta(۱-p)+\theta A(t,r)(۱-p)^۲} P(T_۱ \leqslant T_۲) + e^{-۲\pi r^۲\theta(۱-p)+\theta A(t,r)(۱-p)^۲} P(T_۱ > T_۲),$$

$$= e^{-۲\pi r^۲\theta(۱-p)+\theta A(t,r)(۱-p)^۲} [P(T_۱ \leqslant T_۲) + P(T_۱ > T_۲)] = e^{-۲\pi r^۲\theta(۱-p)+\theta A(t,r)(۱-p)^۲},$$

و در نتیجه:

$$\lambda^{(۲)}(u,v) = \begin{cases} \delta^۲ e^{-۲\pi r^۲\theta(۱-p)+\theta A(t,r)(۱-p)^۲}, & ||u-v|| \leq ۲r, \\ \delta^۲ e^{-۲(۱-p)\theta\pi r^۲}, & ||u-v|| > ۲r. \end{cases}$$

□

قضیه ۲.۳ **اثبات.** قسمت (i) از قضیه ۱.۳ به دست می‌آید. برای (ii) داریم:

$$g_{Y-X}(u,v) = \frac{E\{\delta[۱-\Pi(u)]\delta[۱-\Pi(v)]\}}{E\{\delta[۱-\Pi(u)]\}E\{\delta[۱-\Pi(v)]\}} = \frac{۱-۲E[\Pi(u)]+E[\Pi(u)\Pi(v)]}{\{۱-E[\Pi(u)]\}^۲}$$
$$= \frac{۱-۲E[\Pi(u)]+E[\Pi(u)]^۲+E[\Pi(u)\Pi(v)]-E[\Pi(u)]^۲}{\{۱-E[\Pi(u)]\}^۲} = ۱+\frac{E[\Pi(u)\Pi(v)]-E[\Pi(u)]E[\Pi(v)]}{\{۱-E[\Pi(u)]\}^۲}$$
$$= ۱+\frac{E[\Pi(u)]^۲}{\{۱-E[\Pi(u)]\}^۲}[g_X(u,v)-۱] = ۱+\frac{a^۲}{(۱-a)^۲}[g_X(t)-۱]$$

□

A New Thinning Method of Spatial Point Processes Based on using Boolean Random Set Models

Reza Asadi[1], **Mojtaba Khazaei**[1], **Mojtaba Ganjali**[1]

[1]Department of Statistics, Shahid Beheshti University, Tehran, Iran

Abstract: Spatial point processes are statistical models for the analysis of point patterns. One of the methods of making new point processes is thinning the pre-existing point processes. In this paper, a new dependent thinning algorithm based on a Boolean random set is proposed. This algorithm can be considered as a generalization of the Stoyan thinning algorithm in the interrupted point process. Some properties of the thinned processes and eliminated point processes resulted from the application of the proposed thinning algorithm on the homogeneous Poisson point process, have been studied. Some summary statistics of the both thinned and eliminated point processes are calculated. Palm likelihood method is used to estimate parameters of these models. Finally, the proposed models are used to model two real datasets.

Keywords: Homogeneous Poisson point process; Cox process; Boolean random set; Dependent thinning; Cox process; Palm likelihood.

Mathematics Subject Classification (2020): xxAxx, xxBxx, xxCxx.