Neparametrické odhady charakteristik bodových procesů

Budeme se zabývat odhadováním následujících dvou druhů charakteristik bodových procesů: charakteristikami druhého řádu (K, L, g) a vzdálenostními charakteristikami (F, G, J). Zaměříme se na odhady, které jsou implementovány v balíčku `spatstat`. V literatuře je možno najít několik jiných metod.

Uvažujme stacionární bodový proces X na $\mathbb{R}^d$ s intenzitou λ ($0 < \lambda < \infty$). Předpokládejme, že máme realizaci X v okně $W \in \mathcal{B}_0^d$, $|W| > 0$. V následujícím textu budeme používat úmluvu $a/0 = 0$ pro $a \geq 0$.

K-funkce

Definice: Druhá redukovaná momentová míra $\mathcal{K}$ je dána vztahem $\lambda \mathcal{K}(B) = \int x(B)\, P_0^!(\mathrm{d}x)$, $B \in \mathcal{B}^d$. Pokud je navíc X isotropní, tak $\mathcal{K}$ je rotačně invariantní a definujeme funkci $K(r) = \mathcal{K}(b(0,r))$, $r > 0$.

0. nepoužitelný odhad: Z přednášky víme, že

$$\lambda^2 K(r) = \mathrm{E} \sum_{\xi \in X \cap W} \frac{X(b(\xi, r) \setminus \xi)}{|W|}.$$

Teoreticky bychom tak mohli uvažovat nestranný odhad $\lambda^2 K(r)$ tvaru

$$\widehat{\lambda^2 K(r)} = \sum_{\xi \in X \cap W} \frac{X(b(\xi, r) \setminus \xi)}{|W|} = \sum_{\xi, \eta \in X}^{\neq} \frac{\mathbf{1}_{[\|\xi - \eta\| \leq r]} \mathbf{1}_{[\xi \in W]}}{|W|}.$$

Tento odhad je ale použitelný pouze, pokud máme dodatečnou informaci z vnějšku okna W o bodech η, které leží ve vzdálenosti nejvýše r od bodů procesu ξ ležících v okně (tzv. plusový výběr). Problém spočívá v okrajových efektech: nejsme schopni vyčíslit $X(b(\xi, r) \setminus \xi)$ jenom z informace uvnitř okna W.

1. border: Nejjednodušším postupem, jak se vyhnout okrajovým efektům, je uvažovat X v menším okně $W_{\ominus r} = \{\eta \in W : b(\eta, r) \subseteq W\}$ (tzv. minusový výběr). Pak

$$\widehat{\lambda^2 K_b(r)} = \sum_{\xi \in X \cap W_{\ominus r}} \frac{X(b(\xi, r) \setminus \xi)}{|W_{\ominus r}|} = \sum_{\xi, \eta \in X \cap W}^{\neq} \frac{\mathbf{1}_{[\|\xi - \eta\| \leq r]} \mathbf{1}_{[\xi \in W_{\ominus r}]}}{|W_{\ominus r}|}$$

je nestranný odhad $\lambda^2 K(r)$.

2. translate: Jiná možnost je využít korekčních faktorů, což jsou jakési váhy přiřazené tomu, že pozorujeme dva body v určité vzdálenosti. Na přednášce byl zmíněn tzv. translanční korekční faktor.

$$\widehat{\lambda^2 K_t(r)} = \sum_{\xi, \eta \in X \cap W}^{\neq} \frac{\mathbf{1}_{[\|\xi - \eta\| \leq r]}}{|W \cap (W + \xi - \eta)|}.$$

Bylo ukázáno, že je nestranný, pokud $|W \cap (W + \xi)| > 0$ pro všechna $\|\xi\| \leq r$ (věta 10). Předpoklad na isotropii procesu zde není podstatný, podobně můžeme definovat

$$\widehat{\lambda^2 \mathcal{K}(B)} = \sum_{\xi, \eta \in X \cap W}^{\neq} \frac{\mathbf{1}_{[\xi - \eta \in B]}}{|W \cap (W + \xi - \eta)|}.$$

3. isotropic/Ripley: Jiný korekční faktor navrhl B. Ripley:

$$\widehat{\lambda^2 K_R(r)} = \sum_{\xi, \eta \in X \cap W}^{\neq} \frac{\mathbf{1}_{[\|\xi - \eta\| \leq r]}}{|W|} \cdot \frac{|\partial b(\xi, \|\xi - \eta\|)|}{|\partial b(\xi, \|\xi - \eta\|) \cap W|}.$$

Lze ukázat, že se jedná o nestranný odhad pro $r < r_0 = \inf\{t > 0 : |W^{(t)}| < |W|\}$, kde $W^{(t)} = \{\xi \in W : \partial b(\xi, t) \cap W \neq \emptyset\}$. Někdy se uvádí Ohserova modifikace tohoto odhadu:

$$\widehat{\lambda^2 K_O(r)} = \sum_{\xi, \eta \in X \cap W}^{\neq} \frac{\mathbf{1}_{[\|\xi - \eta\| \leq r]}}{|W^{(\|\xi - \eta\|)}|} \cdot \frac{|\partial b(\xi, \|\xi - \eta\|)|}{|\partial b(\xi, \|\xi - \eta\|) \cap W|}.$$

Tento odhad je nestranný pro $r < r^* = \sup\{s > 0 : |W^{(s)}| > 0\}$. Pro $W = [0,1]^2$ je $r_0 = \sqrt{2}/2$, $r^* = \sqrt{2}$ a $W^{(s)} = W$ pro všechna $s \leq r_0$.

Vlastnosti: Při odhadování $K(r)$ je třeba dělit odhadem druhé mocniny intenzity, což má za následek porušení nestrannosti. Vychýlení a rozptyl se typicky zvětšují s rostoucím r. Pro obdélníkové okno se doporučuje odhady počítat pro r menší než čtvrtina kratší strany obdélníku. Výsledné odhady nemusí být monotónní funkce v r (na rozdíl od teoretické funkce).

S rostoucím r a dimenzí prostoru dochází u okrajové metody k významné ztrátě informace z dat. Statisticky lepší vlastnosti mají odhady založené na korekčních faktorech. Na druhou stranu výpočet $\hat{K}_b$ je rychlejší.

Poznámka: Lze zavést K-funkci i pro nestacionární bodový proces. Odhady jsou pak opět založené na uvedených korekčních faktorech.

L-funkce

Definice: $L(r) = \left(\frac{K(r)}{\omega_d}\right)^{1/d}$.

K odhadu L-funkce využijeme některý z odhadů K-funkce

$$\hat{L}(r) = \left(\frac{\hat{K}(r)}{\omega_d}\right)^{1/d}.$$

Vlastnosti: Důvodem pro zavedení L-funkce je snaha o stabilizaci rozptylu. Pro stacionární Poissonův proces je $L(r) = r$ pro všechna $r > 0$.

Párová korelační funkce

Definice: Pro stacionární isotropní proces souvisí párová korelační funkce s K-funkcí: $g(r) = \frac{K'(r)}{d\omega_d r^{d-1}}$.

Lze použít jádrový odhad s translačním nebo Ripleyho korekčním faktorem:

$$\hat{g}_t(r) = \frac{1}{\widehat{\lambda^2}} \sum_{\xi,\eta \in X \cap W}^{\neq} \frac{k(\|\xi - \eta\| - r)}{d\omega_d r^{d-1} |W \cap (W + \xi - \eta)|},$$

$$\hat{g}_R(r) = \frac{1}{\widehat{\lambda^2}} \sum_{\xi,\eta \in X \cap W}^{\neq} \frac{k(\|\xi - \eta\| - r)}{d\omega_d r^{d-1} |W|} \cdot \frac{|\partial b(\xi, \|\xi - \eta\|)|}{|\partial b(\xi, \|\xi - \eta\|) \cap W|},$$

kde k je vhodná jádrová funkce. V literatuře se doporučuje např. Epanečnikovova jádrová funkce:

$$k_h(u) = \frac{3}{4h}\left(1 - \left(\frac{u}{h}\right)^2\right), \quad |u| \leq h$$

s vhodně zvolenou šířkou pásma (vyhlazovacího okénka) h.

Další možnost je využít některý z odhadů K-funkce a aproximovat derivaci numerickými metodami (např. pomocí splinů). To obvykle není snadné, protože odhad K-funkce je po částech konstantní funkce.

Distribuční funkce nejbližšího souseda

Definice: $G(r) = P_0^!(d(0,X) \leq r)$, $r > 0$.

1. raw: Kdybychom pro každý pozorovaný bod procesu znali vzdálenost k nejbližšímu sousedu, tak můžeme odhadnout distribuční funkci nejbližšího souseda klasickým způsobem jako empirickou distribuční funkci

$$\hat{G}(r) = \frac{1}{X(W)} \sum_{\xi \in X \cap W} \mathbf{1}_{[e(\xi) \leq r]},$$

kde $e(\xi) = d(\xi, X \setminus \xi)$ je vzdálenost ξ k nejbližšímu sousedu. Z Campbellovy věty plyne, že tento odhad je podílově nestranný. Opět díky okrajovým efektům nejsme schopni získat $e(\xi)$ pro každé $\xi \in X \cap W$. Pokud nahradíme $e(\xi)$ vzdáleností $e^*(\xi) = d(\xi, (X \setminus \xi) \cap W)$, kterou jsme schopni pozorovat, dostáváme následující naivní odhad

$$\hat{G}_r(r) = \frac{1}{X(W)} \sum_{\xi \in X \cap W} \mathbf{1}_{[e^*(\xi) \leq r]}.$$

2. reduced sample: Přejdeme-li opět k erodovanému oknu $W_{\ominus r}$, dostaneme podílově nestranný odhad

$$\hat{G}_b(r) = \frac{1}{X(W_{\ominus r})} \sum_{\xi \in X \cap W_{\ominus r}} \mathbf{1}_{\{e(\xi) \leq r\}}.$$

3. Kaplan-Meier: Okrajové efekty lze chápat jako druh cenzorování. Můžeme zavést odhad Kaplan-Meierova typu:

$$\hat{G}_{KM}(r) = 1 - \prod_{s \leq r} \left(1 - \frac{\#\{\xi \in X : e(\xi) = s, e(\xi) \leq c(\xi)\}}{\#\{\xi \in X : e(\xi) \geq s, c(\xi) \geq s\}}\right),$$

kde $c(\xi) = d(\xi, \partial W)$ je vzdálenost ξ od hranice okna. Uvědomme si, že k výpočtu tohoto odhadu nám stačí informace, kterou máme z okna W.

4. hazard: Pokud máme absolutně spojitou distribuční funkci $H(t)$ s hustotou $h(t)$, tak riziková funkce je definována jako $\lambda(t) = h(t)/(1 - H(t))$. Prostorová Kaplan-Meierova metoda umožňuje odhadnout rizikovou funkci $\lambda(r)$ distribuční funkce $G(r)$. Musíme však být opatrní, protože G nemusí mít nutně hustotu.

Vlastnosti: Odhady G nemusí být distribuční funkce: $\hat{G}_b$ nemusí být monotónní a maximální hodnota může být větší nebo menší než 1, $\hat{G}_{KM}$ je neklesající, ale maximální hodnota může být menší než 1. Kaplan-Meierův odhad je vydatnější než minusový odhad.

Kontaktní distribuční funkce

Definice: $F(r) = P(d(0, X) \leq r)$, $r > 0$.

1. raw: Zvolme v okně W mříž I_r a pro každý bod mříže uvažujme nejbližší bod procesu, ten může ležet mimo okno. Tento okrajový efekt potlačíme opět naivním způsobem:

$$\hat{F}_r(r) = \frac{1}{\# I_r} \sum_{\xi \in I_r} \mathbf{1}_{[d(\xi, X \cap W) \leq r]}.$$

2. reduced sample: Zvolme v okně $W_{\ominus r}$ mříž I_r. Označme $d(\xi) = d(\xi, X)$ vzdálenost ξ od nejbližšího bodu procesu. Potom

$$\hat{F}_b(r) = \frac{1}{\# I_r} \sum_{\xi \in I_r} \mathbf{1}_{[d(\xi) \leq r]}$$

je nestranný odhad $F(r)$, neboť vzhledem ke stacionaritě $P(d(\xi) \leq r) = F(r)$. Spojitá verze tohoto odhadu má tvar

$$\hat{F}_b(r) = \frac{|W_{\ominus r} \cap X_r|}{|W_{\ominus r}|},$$

kde $X_r = \{\xi \in \mathbb{R}^d : d(\xi, X) \leq r\} = \cup_{\xi \in X} b(\xi, r)$.

3. Kaplan-Meier:

$$\hat{F}_{KM}(r) = 1 - \prod_{s \leq r} \left(1 - \frac{\#\{\xi \in I_r : d(\xi) = s, d(\xi) \leq c(\xi)\}}{\#\{\xi \in I_r : d(\xi) \geq s, c(\xi) \geq s\}}\right),$$

kde $c(\xi) = d(\xi, \partial W)$ je vzdálenost ξ od hranice okna.

4. hazard: Kontaktní distribuční funkce $F(r)$ stacionárního procesu je absolutně spojitá a riziková funkce $\lambda(r)$ existuje. Odhad je založen na Kaplan-Meierově odhadu $\hat{F}_{KM}(r)$.

Vlastnosti: Odhady F nemusí být distribuční funkce: $\hat{F}_b$ nemusí být spojitá ani monotónní a maximální hodnota může být větší nebo menší než 1, $\hat{F}_{KM}$ je neklesající, ale maximální hodnota může být menší než 1. Kaplan-Meierův odhad je vydatnější než minusový odhad.

J-funkce

Definice:

$$J(r) = \frac{1 - G(r)}{1 - F(r)}, \ r > 0, \ F(r) < 1.$$

Odhad J-funkce vychází z její definice:

$$\hat{J}(r) = \frac{1 - \hat{G}(r)}{1 - \hat{F}(r)}.$$

Rozlišíme následující odhady (podle toho, jaké odhady G a F použijeme): *uncorrected (raw)*, *reduced sample*, *Kaplan-Meier*.

Vlastnosti: I když nekorigované odhady $\hat{G}_r$ a $\hat{F}_r$ jsou výrazně vychýlené, tak jejich podílem se dostane přibližně nestranný odhad (alespoň když daný bodový proces je blízko Poissonovu procesu). Výhodou tohoto odhadu je necitlivost vzhledem k okrajovým efektům, měl by se tedy použít, pokud jsou okrajové efekty významné.

Další dva odhady jsou mírně vychýlené (podíl dvou přibližně nestranných odhadů). Logaritmus Kaplan-Meierova odhadu je nestranný odhad $\log J$.