Statistika Poissonových modelů pro sjednocení kruhů

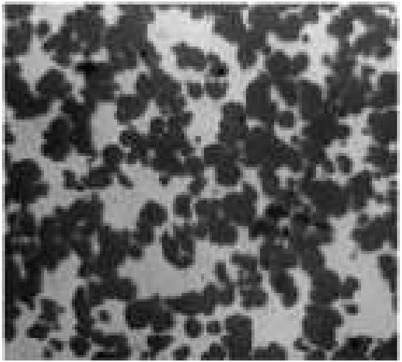
Zbyněk Pawlas

Katedra pravděpodobnosti a matematické statistiky
Matematicko-fyzikální fakulta
Univerzita Karlova v Praze

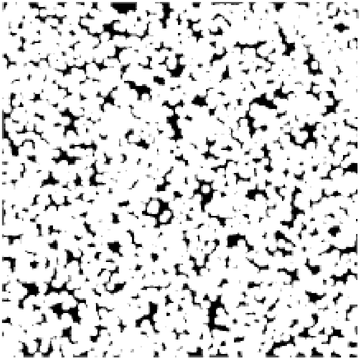
18. zimní škola JČMF ROBUST 2014

20. ledna 2014, Jetřichovice

Geometrické struktury v materiálových vědách

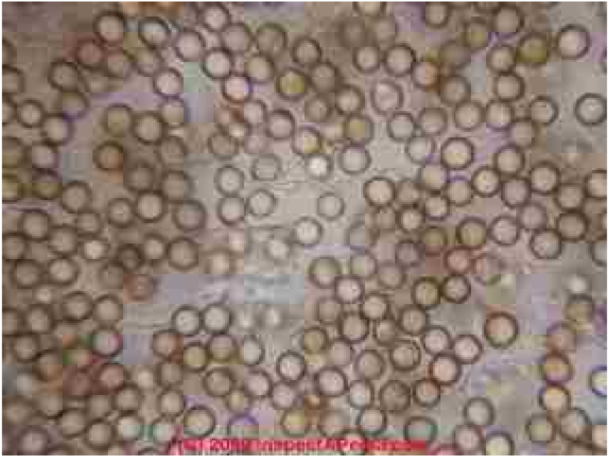


slitina Fe-Ag

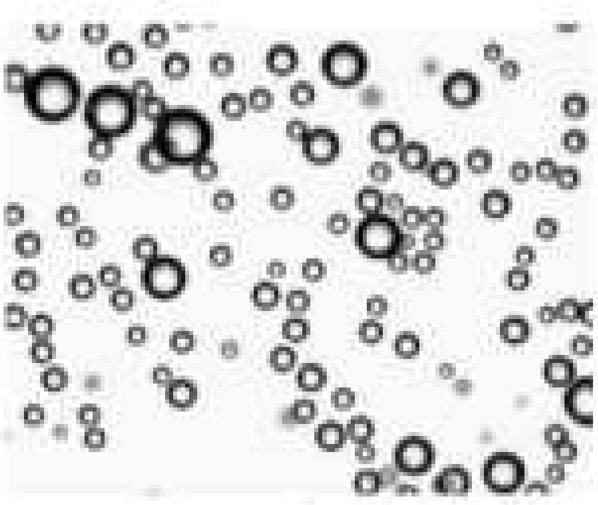


pórovitý materiál

Spory pýchavky

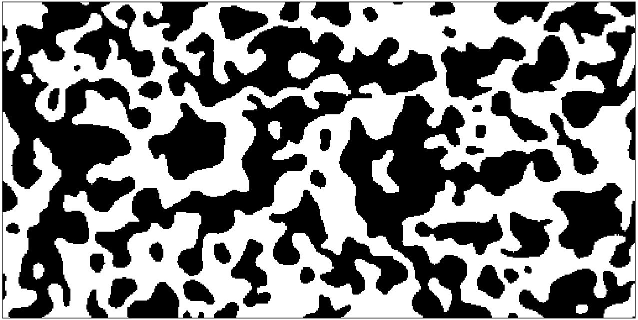


Vzduchové bubliny při flotaci



Vřes

Jädraås, Švédsko



Analýza vřesových dat

Diggle (1981) navrhl použít Booleův model kruhů, tj. kruhy mají náhodné poloměry a středy jsou generovány Poissonovým bodovým procesem.

Další studie:
Hall (1985, 1988), Ripley (1988), Cressie (1991)

Booleův model kruhů není vhodný.

Modely interagujících kruhů:
Møller a Helisová (2010), Dereudre et al. (2013)

Stacionární Booleův model koulí

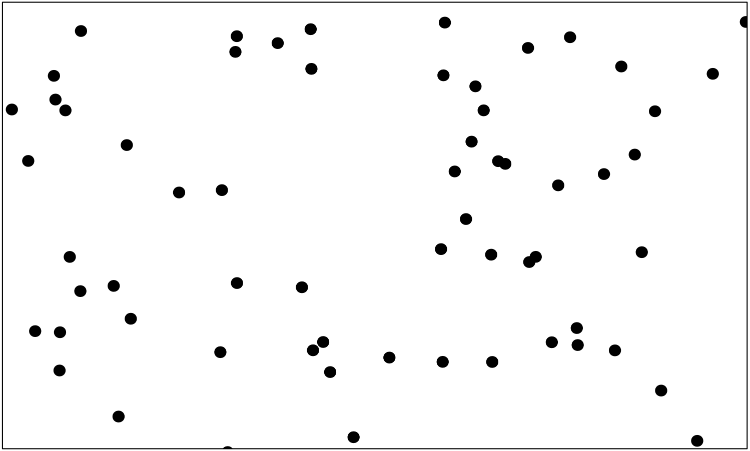
Náhodná uzavřená množina Z v $\mathbb{R}^d$ daná jako sjednocení náhodných koulí:

$$Z := \bigcup_{n \geq 1} B(\xi_n, R_n),$$

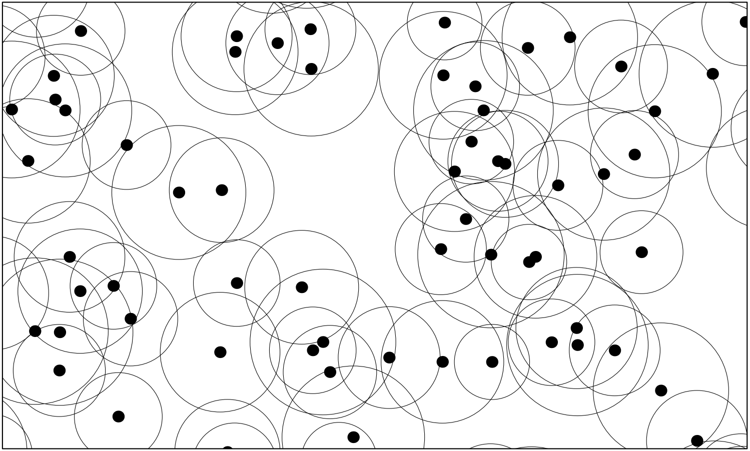
kde $B(x, r) := \{y \in \mathbb{R}^d : \|y - x\| \leq r\}$.

- $\Phi := \{\xi_n : n \geq 1\}$ stacionární Poissonův bodový proces v $\mathbb{R}^d$ s intenzitou $0 < \gamma < \infty$
- (R_n) posloupnost nezávislých, stejně rozdělených, kladných náhodných veličin se společným rozdělením $\mathbb{G}$
- (R_n) a Φ jsou nezávislé

Poissonův bodový proces



Poissonův bodový proces kruhů



Simulovaná realizace Booleova modelu kruhů

Typický poloměr

$$Z := \bigcup_{n \geq 1} B(\xi_n, R_n)$$

typický poloměr R je generická náhodná veličina se stejným rozdělením jako (R_n), tj. s rozdělením $\mathbb{G}$ (distribuční funkcí G)

$\mathbb{E}R^d < \infty$ zaručuje, že Z je s.j. uzavřená a různá od $\mathbb{R}^d$

budeme předpokládat, že

$$\mathbb{E}R^{2d} < \infty$$

Základní charakteristiky množiny Z

objemový podíl

$$p := \mathbb{E}|Z \cap [0,1]^d|_d = \mathbb{P}(o \in Z) = 1 - \exp\{-\gamma \mathbb{E}|B(o,R)|_d\}$$

kovariance

$$\begin{aligned} C(x) &:= \mathbb{P}(o \in Z, x \in Z) \\ &= 2p - 1 + \exp\{-\gamma \mathbb{E}|B(o,R) \cup B(x,R)|_d\} \end{aligned}$$

kapacitní funkcionál

$$T(K) := \mathbb{P}(Z \cap K \neq \emptyset) = 1 - \exp\{-\gamma \mathbb{E}|K \oplus B(o,R)|_d\}$$

Statistický problém

Cíl: získat informaci o γ a $\mathbb{G}$ na základě pozorování Z v omezeném okně $W \subseteq \mathbb{R}^d$

Problémy:

- překrývání koulí
- okrajové efekty

Parametrický přístup: Dupač (1980), Serra (1980), Diggle (1981), Hall (1988), ...

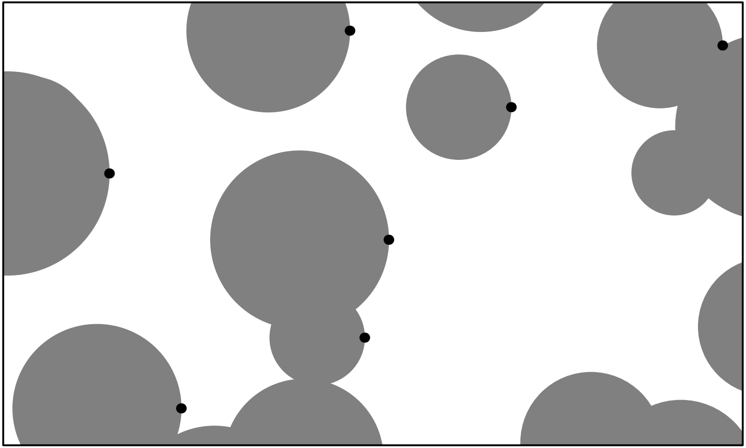
Odhad γ: Schmittova metoda, Laslettova transformace, ...

Neparametrické odhady $\mathbb{G}$

Neparametrické odhady rozdělení poloměru

- Hall (1988) – vážená empirická distribuční funkce pozorovaných poloměrů $[d = 2]$
- Molchanov (1990) – odhad $H(t) = \int_0^t G(s)\,\mathrm{d}s$ na základě odhadu kapacitního funkcionálu $T(K)$ pro vhodné kompaktní množiny K $[d \leq 3]$
- Molchanov a Stoyan (1994) a Heinrich a Werner (2000) – metoda tečných bodů
- Thovert et al. (2001), Thovert a Adler (2011) – odhad kovariance $C(x)$ $[d = 2]$
- Emery et al. (2012) – odhad druhé derivace funkce $C(\|x\|)$ $[d = 2]$

Exponované pravé tečné body



Metoda tečných bodů

pro pevné $u \in \mathbb{S}^{d-1}$:
Φ_u ... bodový proces exponovaných tečných bodů (tečné body nepřekryté jiným zrnem) ve směru u

Φ_u je stacionární (ale ne nutně izotropní) s intenzitou

$$\gamma_u = \gamma(1-p)$$

párová korelační funkce procesu Φ_u splňuje (pokud $\langle u, x\rangle \neq 0$)

$$g_u(x) = \frac{\gamma_u(1-2p+C(x))}{(1-p)^2} G\left(\frac{\|x\|^2}{2|\langle u, x\rangle|}\right)$$

jádrový odhad $g_u(x)$, empirické odhady p, $C(x)$ a γ_u

Kontaktní vzdálenosti

strukturní element: $B \neq \{o\}$ kompaktní konvexní množina obsahující počátek $o \in \mathbb{R}^d$

B-vzdálenost bodu $x \in \mathbb{R}^d$ k množině $A \subseteq \mathbb{R}^d$:

$$d_B(x, A) := \inf\{r \geq 0 : (x + rB) \cap A \neq \emptyset\} \in [0, \infty]$$

Příklady:

- sférický: $B = B(o, 1) = B^d$
- lineární: $B = [0, u]$, $\|u\| = 1$

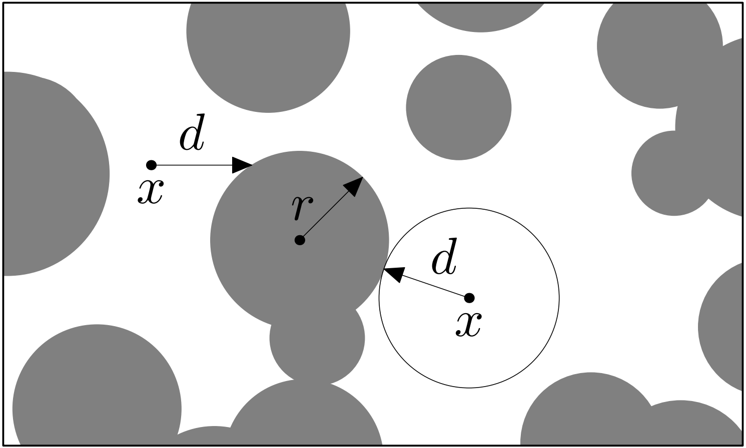
Pro dané $x \notin Z$, s.j. existuje jednoznačná koule $B(\xi_n, R_n)$ taková, že

$$(x + d_B(x, Z)B) \cap B(\xi_n, R_n) \neq \emptyset,$$

definujeme $r_B(x, Z)$ jako R_n.

Vzdálenosti pro sférický a lineární strukturní element

$d = d_B(x, Z)$, $r = r_B(x, Z)$



Kontaktní distribuční funkce

kontaktní distribuční funkce množiny Z:

$$\begin{aligned} F_B(t) &:= \mathbb{P}(d_B(o,Z) \le t) = \mathbb{P}(Z \cap tB \neq \emptyset) \\ &= 1 - \exp\left\{-\gamma \mathbb{E}|B_{t,R}|_d\right\} \\ &= 1 - \exp\left\{-\gamma \sum_{j=0}^{d} \kappa_{d-j} V_j(B) t^j \mathbb{E} R^{d-j}\right\}, \quad t \in \mathbb{R}^+, \end{aligned}$$

kde $B_{s,r} := sB \oplus rB^d$ pro $s, r \ge 0$, $\kappa_k := \frac{\pi^{k/2}}{\Gamma(1+\frac{k}{2})}$ je objem k-rozměrné jednotkové koule a $V_j(B)$ je j-tý vnitřní objem množiny B

doplňková kontaktní distribuční funkce:

$$\bar{F}_B(t) = 1 - F_B(t)$$

Kontaktní distribuční funkce druhého řádu

doplňková kontaktní distribuční funkce druhého řádu množiny Z:

$$\begin{aligned}\bar{F}_B^{(2)}(u; t_1, t_2) &:= \mathbb{P}(d_B(o, Z) > t_1, d_B(u, Z) > t_2) \\ &= \exp\left\{-\gamma \mathbb{E}\left|(t_1 B \cup (u + t_2 B)) \oplus R B^d\right|_d\right\}\end{aligned}$$

pro $u \in \mathbb{R}^d$, $t_1, t_2 \in \mathbb{R}^+$

Užitečný výsledek

Věta (Hug, Last, Weil, 2002): Pro všechny měřitelné funkce $g : [0, \infty] \times \mathbb{R}^+ \to \mathbb{R}^+$ takové, že $g(0, r) = g(\infty, r) = 0$, $r \in \mathbb{R}^+$, a všechny $x \in \mathbb{R}^d$:

$$\mathbb{E} g(d_B(x, Z), r_B(x, Z)) = \\ = \gamma \int_0^\infty \int_0^\infty g(t, r) h_B(t, r) \bar{F}_B(t) \, \mathrm{d}t \, \mathbb{G}(\mathrm{d}r),$$

kde

$$h_B(t, r) := \sum_{j=0}^{d-1} (j+1) \kappa_{d-1-j} V_{j+1}(B) r^{d-1-j} t^j, \quad t, r \geq 0.$$

Pozorování

kompaktní konvexní okno pozorování $W \subseteq \mathbb{R}^d$ takové, že $|W|_d > 0$

data ve tvaru

$$\{(d_B(x, Z), r_B(x, Z)) : x \in W \setminus Z\}$$

to může vyžadovat informaci mimo W
nejbližší koule k bodu $x \in W \setminus Z$ může ležet mimo W

Volíme měřitelnou funkci $f : [0, \infty] \to \mathbb{R}^+$ takovou, že $f(0) = f(\infty) = 0$ a předpokládáme, že

$$0 < \beta := \int_0^\infty f(t) \bar{F}_B(t) \, \mathrm{d}t < \infty.$$

Neparametrický odhad $\mathbb{G}$

Pro libovolnou $C \in \mathcal{B}(\mathbb{R}^+)$ definujeme

$$\widehat{\mathbb{G}}(C) := \frac{\int\limits_W \mathbf{1}\{r_B(x,Z) \in C\} \frac{f\left(d_B(x,Z)\right)}{h_B\left(d_B(x,Z), r_B(x,Z)\right)} \,\mathrm{d}x}{\int\limits_W \frac{f\left(d_B(x,Z)\right)}{h_B\left(d_B(x,Z), r_B(x,Z)\right)} \,\mathrm{d}x}.$$

Poznámky:

- Pokládáme $\frac{f(t)}{h_B(t,r)} := 0$ pro $t \in \{0, \infty\}$ a $r \geq 0$, takže integrace je přes doplněk množiny Z.
- Pokud je jmenovatel nula, tak i čitatel je nula. V tom případě používáme konvenci $0/0 := 0$.
- V praktických aplikacích aproximujeme integrál riemannovskou sumou.

Vážený součet

Nechť A_n je množina bodů $x \in W \setminus Z$ takových, že $d_B(x, Z) = d_B(x, B(\xi_n, R_n))$.

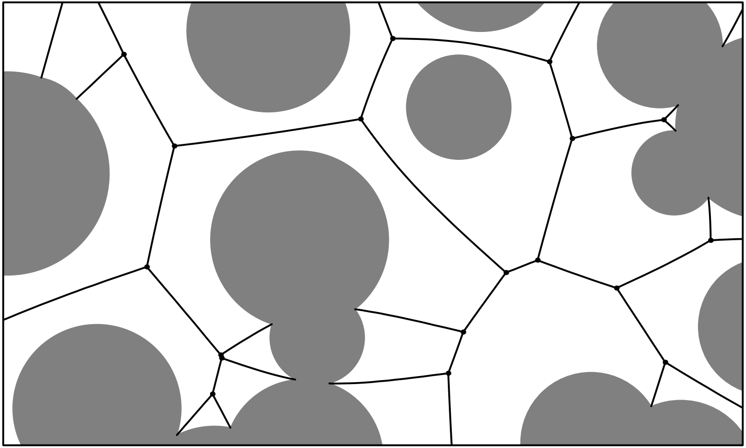
Pokud definujeme váhy

$$w_n = \int_{A_n} \frac{f(d_B(x, Z))}{h_B(d_B(x, Z), R_n)} \, \mathrm{d}x,$$

pak

$$\widehat{\mathbb{G}}(C) = \frac{\sum_n w_n \mathbf{1}\{R_n \in C\}}{\sum_n w_n}.$$

Zobecněná Poissonova-Voroného mozaika



Rovinný případ – sférický element

Uvažujme $d = 2$ a $B = B^2$, potom $h_B(t, r) = 2\pi(t + r)$.

Když $f(t) = \frac{1}{\varepsilon}\mathbf{1}\{t \leq \varepsilon\}$, pak

$$w_n = \frac{1}{2\pi\varepsilon} \int_{A_n \cap (C_n \oplus \varepsilon B^2)} \frac{1}{d_B(x, C_n) + R_n} \, \mathrm{d}x,$$

kde C_n je oblouk na hranici Z příslušející kouli $B(\xi_n, R_n)$, tj. $C_n = \partial B(\xi_n, R_n) \cap \partial Z$.

Pro $\varepsilon \to 0$ dostáváme

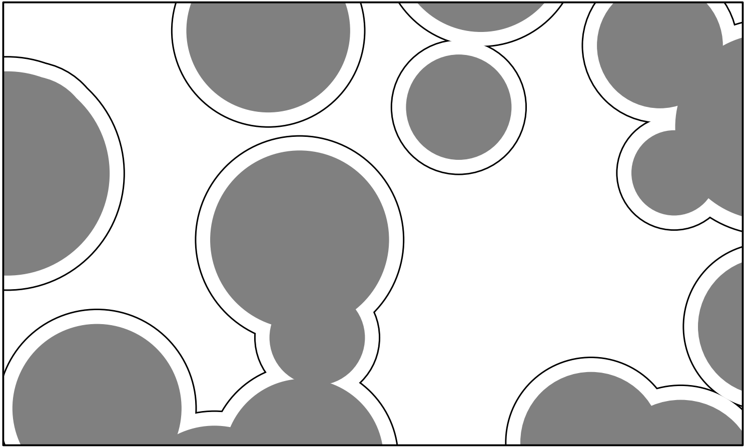
$$w_n = \frac{\mathcal{H}^1(\partial B(\xi_n, R_n) \cap \partial Z \cap W)}{2\pi R_n},$$

což vede na odhad

$$\widehat{\mathbb{G}}_0(C) = \frac{\sum_n w_n \mathbf{1}\{R_n \in C\}}{\sum_n w_n},$$

který byl navržen v knize P. Hall (1988).

Obor integrace pro f s omezeným nosičem



Hallův odhad

Uvažujme $d = 2$ a $B = B^2$, potom $h_B(t,r) = 2\pi(t+r)$.

Když $f(t) = \frac{1}{\varepsilon}\mathbf{1}\{t \le \varepsilon\}$, pak

$$w_n = \frac{1}{2\pi\varepsilon}\int_{A_n \cap (C_n \oplus \varepsilon B^2)} \frac{1}{d_B(x, C_n) + R_n}\,\mathrm{d}x,$$

kde C_n je oblouk na hranici Z příslušející kouli $B(\xi_n, R_n)$, tj. $C_n = \partial B(\xi_n, R_n) \cap \partial Z$.

Pro $\varepsilon \to 0$ dostáváme

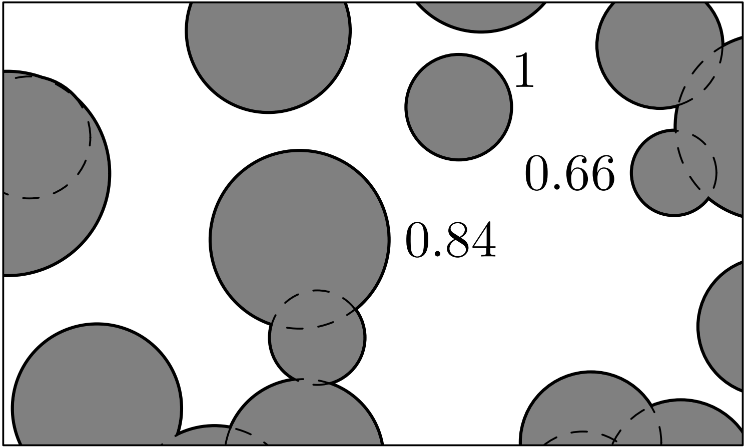
$$w_n = \frac{\mathcal{H}^1(\partial B(\xi_n, R_n) \cap \partial Z \cap W)}{2\pi R_n},$$

což vede na odhad

$$\widehat{\mathbb{G}}_0(C) = \frac{\sum_n w_n \mathbf{1}\{R_n \in C\}}{\sum_n w_n},$$

který byl navržen v knize P. Hall (1988).

Váhy pro Hallův odhad



Rovinný případ – lineární element

Uvažujme $d = 2$ a $B = [0, u]$, potom $h_B(t, r) = 2r$.

Když $f(t) = \frac{1}{\varepsilon}\mathbf{1}\{t \leq \varepsilon\}$, pak

$$w_n = \frac{1}{2\varepsilon R_n} \int_{A_n \cap (C_n \oplus \varepsilon[0,-u])} \mathrm{d}x = \frac{|A_n \cap (C_n \oplus \varepsilon[0,-u])|_2}{2\varepsilon R_n}.$$

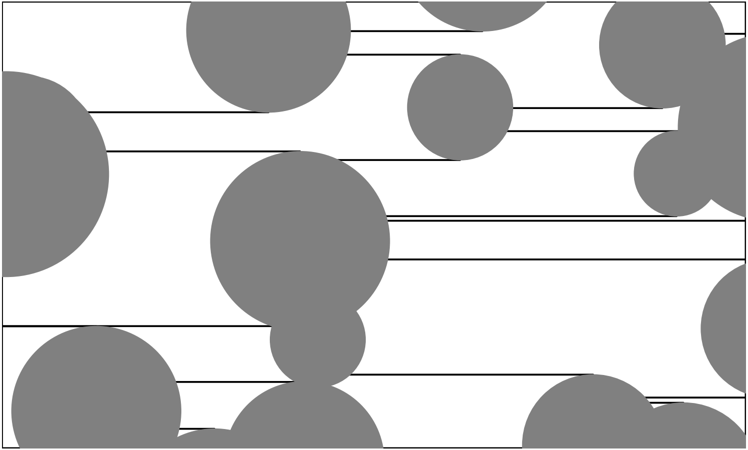
V limitě $\varepsilon \to 0$:

$$w_n = \frac{L_n(u)}{2\pi R_n},$$

kde $L_n(u)$ je délka projekce viditelné části oblouku C_n ve směru u.

Odhad lze vylepšit kombinováním různých směrů: $u = e_1, -e_1, e_2, -e_2$.

Zobecněná Poissonova-Voronого mozaika – lineární případ



Podílová nestrannost

Definujeme-li

$$\eta_W(C) := \int_W \mathbf{1}\{r_B(x,Z) \in C\} \frac{f(d_B(x,Z))}{h_B(d_B(x,Z), r_B(x,Z))} \, \mathrm{d}x,$$

potom

$$\widehat{\mathbb{G}}(C) = \frac{\eta_W(C)}{\eta_W(\mathbb{R}^+)}, \quad C \in \mathcal{B}(\mathbb{R}^+).$$

Protože

$$\mathbb{E}\eta_W(C) = \gamma\,\beta\,|W|_d\,\mathbb{G}(C),$$

máme

$$\frac{\mathbb{E}\eta_W(C)}{\mathbb{E}\eta_W(\mathbb{R}^+)} = \mathbb{G}(C),$$

tj. $\widehat{\mathbb{G}}(C)$ je podílově nestranný odhad $\mathbb{G}(C)$.

Vlastnosti druhého řádu

Tvrzení: Pro libovolnou $C \in \mathcal{B}(\mathbb{R}^+)$ a kompaktní konvexní W_1 a W_2 platí

$$\operatorname{Cov}\left(\eta_{W_1}(C), \eta_{W_2}(C)\right) = \int_{\mathbb{R}^d} |W_1 \cap (W_2 - u)|_d \left[\gamma\tau_1(C,u) + \gamma^2\tau_2(C,u)\right] \mathrm{d}u,$$

kde $\tau_1(C,u)$ a $\tau_2(C,u)$ lze vyjádřit pomocí integrálů obsahujících f, h_B a $\bar{F}_B^{(2)}$.

Asymptotický rozptyl

Věta: Nechť (W_n) je posloupnost konvexních množin taková, že $W_n \nearrow \mathbb{R}^d$. Pokud

$$\int_0^\infty f(t) e^{-ct}\,\mathrm{d}t < \infty \quad \text{pro libovolné} \quad c > 0,$$

potom

$$\frac{\operatorname{Var} \eta_{W_n}(C)}{|W_n|_d} \xrightarrow{n\to\infty} \sigma^2(C).$$

Asymptotický rozptyl je konečný a daný vztahem

$$\sigma^2(C) = \gamma \int_{\mathbb{R}^d} \tau_1(C,u)\,\mathrm{d}u + \gamma^2 \int_{\mathbb{R}^d} \tau_2(C,u)\,\mathrm{d}u.$$

Jestliže $0 < \mathbb{G}(C) < 1$, pak $\sigma^2(C) > 0$.

Konzistence

Důsledek:

$$\frac{\eta_{W_n}(C)}{|W_n|_d} \xrightarrow{n\to\infty} \gamma\,\beta\,\mathbb{G}(C) \quad \text{v pravděpodobnosti}$$

a $\widehat{\mathbb{G}}_n(C) := \frac{\eta_{W_n}(C)}{\eta_{W_n}(\mathbb{R}^+)}$ je slabě konzistentní odhad $\mathbb{G}(C)$

Věta: $\widehat{\mathbb{G}}_n(C)$ je silně konzistentní odhad $\mathbb{G}(C)$

Asymptotická normalita

Věta: Předpokládejme, že $\int_0^\infty (1+t^d) f(t)\,\mathrm{d}t < \infty$. Nechť $W_n = [-n,n)^d$. Potom pro libovolnou $C \in \mathcal{B}(\mathbb{R}^+)$ platí

$$\sqrt{|W_n|_d}\left(\frac{\eta_{W_n}(C)}{|W_n|_d} - \gamma\,\beta\,\mathbb{G}(C)\right) \xrightarrow{d} N(0,\sigma^2(C))$$

a

$$\sqrt{|W_n|_d}\left(\widehat{\mathbb{G}}_n(C) - \mathbb{G}(C)\right) \xrightarrow{d} N(0,\sigma^2_{\mathbb{G}}(C)),$$

kde

$$\sigma^2_{\mathbb{G}}(C) = \frac{1}{\gamma\beta^2}\int_{\mathbb{R}^d}\int_0^\infty\int_{\mathbb{R}^d} \frac{f\left(d_B(o,B(x,r))\right)}{h_B\left(d_B(o,B(x,r)),r\right)}\,\frac{f\left(d_B(u,B(x,r))\right)}{h_B\left(d_B(u,B(x,r)),r\right)}$$
$$\times\ \bar{F}_B^{(2)}\left(u; d_B(o,B(x,r)), d_B(u,B(x,r))\right)\left(\mathbf{1}\{r\in C\} - \mathbb{G}(C)\right)^2$$
$$\times\ \mathrm{d}x\,\mathbb{G}(\mathrm{d}r)\,\mathrm{d}u.$$

Děkuji za pozornost