

GEOGRAFICKÁ PROFILACE S VYUŽITÍM BAYESOVSKÝCH METOD

Robust 2018

Jana Faltýnková

Přírodovědecká fakulta

MASARYKOVA UNIVERZITA 2018



- 1 Úvod do matematiky v kriminologii
- 2 Bayesovský přístup
- 3 Zpracování dat
- 4 Simulace
- 5 Závěr
- 6 Literatura



Matematika v kriminologii

- Matematika jako nástroj pro
 - hledání pachatelů
 - předpovězení místa příštího trestného činu zkoumaného pachatele
- Matematici specializující se na kriminologii: *Rossmo, Mohler and Short, O'Leary*
- Geografická profilace – hledání místa obvyklého výskytu pachatele, tzv. kotevního bodu $\mathbf{z} = (z^{(1)}, z^{(2)}) \in \mathbb{R}^2$
 - místo bydliště
 - místo výkonu práce
 - oblíbená hospoda, bar
 - ...

Jak lokalizovat kotevní bod?

- Strategie založené na prostorovém rozložení
- Strategie založené na pravděpodobnostním rozložení
 - Skórová funkce

$$S(\mathbf{y}) = \sum_{i=1}^n f(d(\mathbf{x}_i, \mathbf{y}))$$

- f ... funkce vzdálenosti
- d ... vzdálenostní metrika
- Nejpopulárnější: *Rossmův model CGT*



Bayesovské metody v geografické profilaci

- Proč jsou užitečné?
 - implementace informací dostupných před analyzováním dat
 - výsledek: posteriorní rozdělení → obdržíme celou oblast, která s vysokou pravděpodobností obsahuje kotevní bod
- Aplikace Bayesova vzorce

$$p(\theta|D) = \frac{f(D|\theta) \cdot p(\theta)}{f(D)}$$

$$p(\theta|D) \propto f(D|\theta) \cdot p(\theta)$$

- $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ – známá místa trestných činů spáchaných v dané sérii
- model pachatelova chování: $p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} | \mathbf{z}, \theta)$.
- Obdržíme:

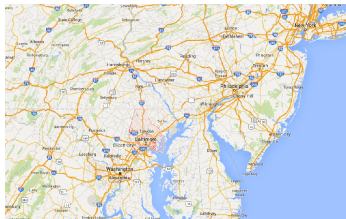
$$p(\mathbf{z}, \theta | \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}) = \frac{p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\} | \mathbf{z}, \theta) \cdot p(\mathbf{z}, \theta)}{p(\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\})}.$$

- S využitím některých předpokladů (např. nezávislosti) a dalším zjednodušením můžeme psát, že $p(\mathbf{z} | \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\})$ je úměrná

$$\int \dots \int_{M_\theta} p(\mathbf{x}_1 | \mathbf{z}, \theta) \cdot \dots \cdot p(\mathbf{x}_n | \mathbf{z}, \theta) \cdot h(\mathbf{z}) \cdot g(\theta) d\theta$$

Dataset kriminálních sérií pro analýzu

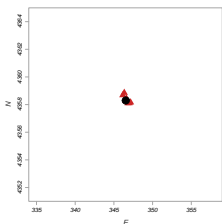
- 88 sériových pachatelů, 962 trestných činů
- Baltimore County, 1993–1997
- Dataset obsahuje informace o: zeměpisné šířce a délce pachatelova kotevního bodu, místa trestného činu a typ trestného činu



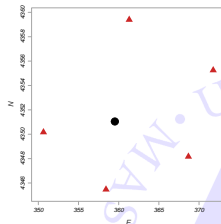
Obrázek 1: *Baltimore County, USA (Google maps)*

Typy pachatelů a modely pro ně určené

■ Místní pachatelé (*marauders, residents*):



(a) Bez ochranné zóny.



(b) S ochrannou zónou.

Obrázek 2: Dva základní podtypy místních pachatelů (červený trojúhelník značí místo trestného činu, černé kolečko kotevní bod).

Typy pachatelů a modely pro ně určené

■ Místní pachatelé:

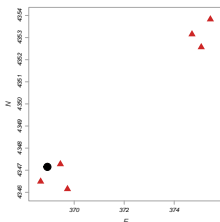
- *Bez ochranné zóny* → normální rozdělení se střední hodnotou v místě kotevního bodu
- *S ochrannou zónou* → model je úměrný

$$\exp\left(-\frac{1}{2\sigma^2} \left[\sqrt{(x_i^{(1)} - z^{(1)})^2 + (x_i^{(2)} - z^{(2)})^2} - \alpha \right]^2\right)$$

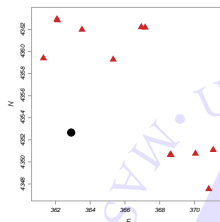
- *S ochrannou zónou nebo bez ní, ale s výskytem shluků* → multimodelová inference

Typy pachatelů a modely pro ně určené

■ Místní se shluky:



(a) Bez významné ochranné zóny.



(b) S ochrannou zónou.

Obrázek 3: *Místní, jejichž trestné činy vytvářejí shluky (červený trojúhelník značí místo trestného činu, černé kolečko kotevní bod).*

Typy pachatelů a modely pro ně určené

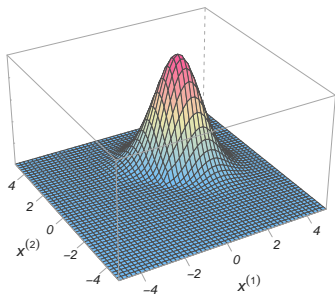
■ Dojíždějící (*commuters, non-residents*):

- páchají trestné činy relativně daleko od svého kotevního bodu
- preferují určitý úhel pro volbu místa trestného činu
- model je kombinací funkce vzdálenosti $q_1(\mathbf{x}_i|\mathbf{z}, \alpha, \sigma_1)$ a funkce preferovaného úhlu $q_2(\mathbf{x}_i|\mathbf{z}, \vartheta, \sigma_2)$

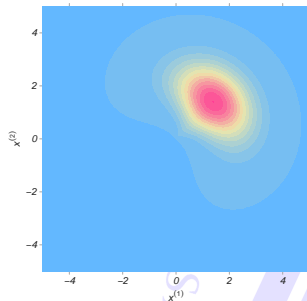
$$q_1 : \exp \left(-\frac{1}{2\sigma_1^2} \left[\sqrt{(x_i^{(1)} - z^{(1)})^2 + (x_i^{(2)} - z^{(2)})^2} - \alpha \right]^2 \right)$$

$$q_2 : \exp \left(-\frac{1}{2\sigma_2^2} \left[\arg \left((x_i^{(1)} - z^{(1)}) + i(x_i^{(2)} - z^{(2)}) \right) - \vartheta \right]^2 \right) .$$

Typy pachatelů a modely pro ně určené



(a) Trojrozměrný graf.



(b) Vrstevnicový graf.

Obrázek 4: Model pro dojíždějící pachatele s kotevním bodem $\mathbf{z} = [0, 0]$, $\alpha = 2$, $\sigma_1 = \frac{4}{5}$, $\vartheta = \frac{\pi}{4}$ and $\sigma_2 = \frac{\pi}{6}$.

Typy pachatelů a modely pro ně určené

■ Dojíždějící (*commuters, non-residents*):

■ další variace modelu:

- Funkce vzdálenosti q_1 inspirovaná log-normálním rozdělením:

$$q_1 : \exp \left(-\frac{1}{2\sigma_1^2} \left[\ln \sqrt{(x_i^{(1)} - z^{(1)})^2 + (x_i^{(2)} - z^{(2)})^2} - \mu \right]^2 \right)$$

- Zavedení tzv. zavedení tzv. hybridního vícerozměrného log-normálního a normálního rozdělení

Typy pachatelů a modely pro ně určené

■ Dojíždějící (*commuters, non-residents*):

- Hybridní vícerozměrné log-normální a normální rozdělení:

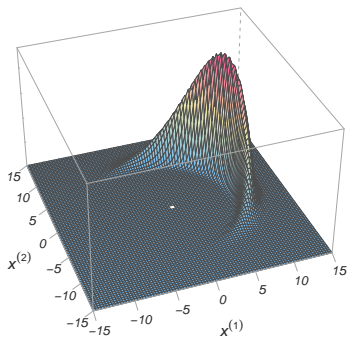
$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \frac{1}{y_1} \cdot \exp\left\{-\frac{1}{2(1-\rho^2)} \cdot m(y_1, y_2)\right\},$$

kde $m(y_1, y_2)$ odpovídá

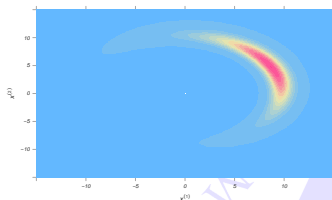
$$\frac{(\ln y_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(\ln y_1 - \mu_1)(y_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(y_2 - \mu_2)^2}{\sigma_2^2}.$$

- uvážením $y_1 = \sqrt{(x_i^{(1)} - z^{(1)})^2 + (x_i^{(2)} - z^{(2)})^2}$ a $y_2 = \arctg 2(x_i^{(2)} - z^{(2)}, x_i^{(1)} - z^{(1)})$ dostaneme požadovaný tvar pro náš případ.
- možnost uvážení korelace ρ mezi vzdáleností a úhlem

Typy pachatelů a modely pro ně určené



(a) Trojrozměrný graf.



(b) Vrstevnicový graf.

Obrázek 5: Hybridní log-normální a normální model pro dojíždějící pachatele s kotevním bodem $\mathbf{z} = [0, 0]$, $\alpha = 10$, $\vartheta = \frac{\pi}{6}$, $s_1 = 0,8$, $s_2 = \frac{\pi}{6}$
 $a \rho = \frac{1}{3}$.

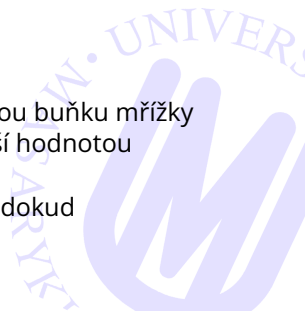
Volba priorní funkce

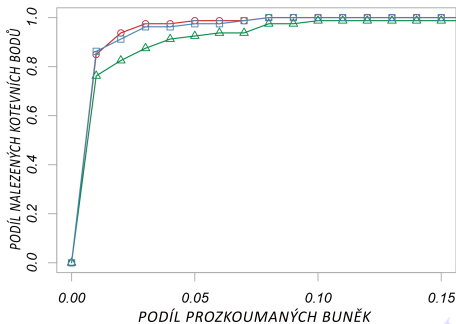
- Pro každý parametr potřebujeme zvolit vhodný prior
- Jádrové vyhlazování logsplínové odhady hustoty
- Předpoklad: Všechny informace o každém pachateli jsou nám známy → vyloučíme pouze data o vyšetřovaném viníkovi



Zpracování dat

- Grafické znázornění míst trestných činů u každého pachatele
- Volba vhodného modelu
- Rozdělení plochy na mřížku
- Pouze místní pachatelé
 - Posterior a Rossmova metoda pro každou buňku mřížky
 - Uspořádání všech buněk od té s nejvyšší hodnotou posterioru
 - Efektivita: Kolik buněk musíme vyšetřit, dokud nenalezneme kotevní bod





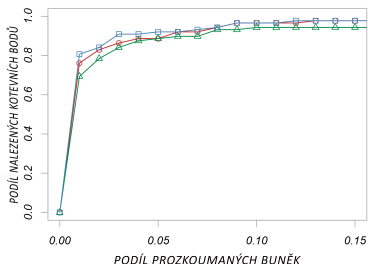
Obrázek 6: Vztah mezi proporcí nalezených kotevnických bodů a proporcí vyšetřované oblasti, pokud se zabýváme pouze místními pachateli (červená kolečka a modré čtverečky - naše metoda; zelené trojúhelníčky - Rossmova metoda).

Zpracování dat

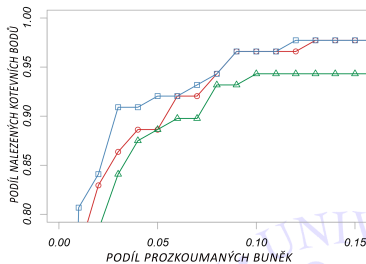
■ Všichni pachatelé

- Posterior pro místní, posterior pro dojíždějící → v každé buňce multimodelová inference
- Vyhodnocení v každé buňce prostřednictvím Rossmovy metody
- Stejný způsob zhodnocení efektivity



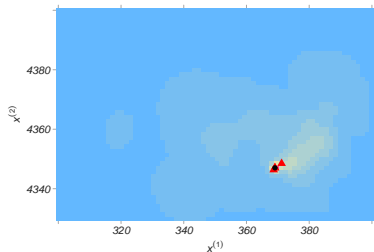


(a) Normální měřítko.

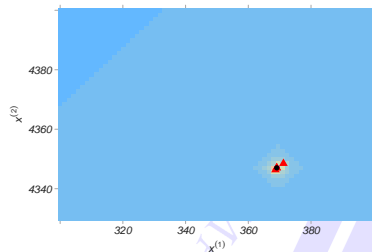


(b) Detail.

Obrázek 7: *Vztah mezi proporcí nalezených kotevních bodů a proporcí vyšetřované oblasti, pokud se zabýváme všemi pachateli, aniž bychom znali předem jejich typ (červená kolečka a modré čtverečky - naše metoda; zelené trojúhelníčky - Rossmova metoda).*

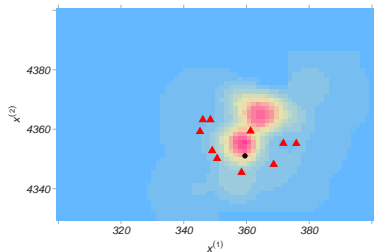


(a) Náš model.

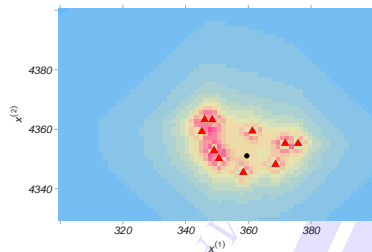


(b) Rossmův model.

Obrázek 8: *Místní pachatelé bez ochranné zóny.*

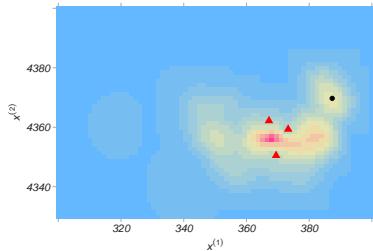


(a) Náš model.

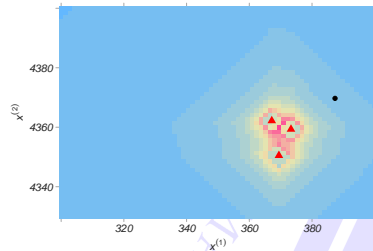


(b) Rossmův model.

Obrázek 9: *Místní pachatelé s ochrannou zónou*



(a) Náš model.



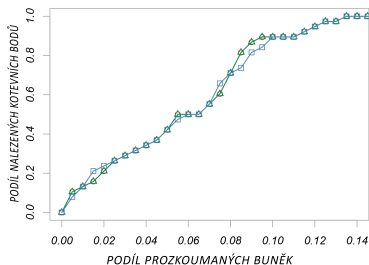
(b) Rossmův model.

Obrázek 10: *Dojíždějící.*

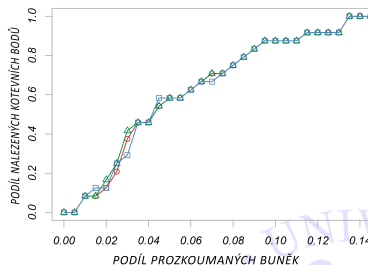
Simulace

- Celkem 30 sérií pachatelů, minimálně 8 pachatelů, maximálně 49
- Dohromady 839 pachatelů se 14 087 trestnými činy
- Aplikace tří variací téhož modelu:
 - (1) Korelace ρ odhadnuta z dat
 - (2) Korelace ρ nulová
 - (3) Přesně taková hodnotu korelace ρ , která byla vygenerována (tedy známá hodnota)





(a) Série 8.



(b) Série 25.

Obrázek 11: Vztah mezi podílem nalezených kotevních bodů vzhledem k podílu prozkoumané plochy – tři varianty hybridního vícerozměrného normálního a lognormálního rozdělení: (1) ρ odhadnuté z dat (červená kolečka), (2) ρ nulové (zelené trojúhelníčky), (3) ρ známé (modré čtverečky).

Závěr

- Náš přístup - komplexnější uchycení problému
- Uvažované modely zachycují pachatelovo chování lépe než modely *O'Leary*
- Jsou jednodušší a interpretovatelnější než modely *Mohlera* and *Shorta*
- Tento přístup ukazuje, jak rozlišit typy pachatelů na základě prostorového rozložení míst trestných činů
- Vyšší efektivita oproti využití Rossmova přístupu, obzvláště v přítomnosti dojíždějících pachatelů.

Další problémy a možnosti zlepšení

- Jak bude metoda fungovat na jiném datovém souboru?
- Je možné použít tento přístup v České republice?
- Jaké vhodnější modely pro místní i dojíždějící pachatele lze najít?
- Bude mít korelace ρ v reálném případě vliv na zlepšení odhadu?
- Lze využít modely i pro teroristické organizace?

- [1] MOHLER, George O. and Martin B. SHORT. Geographic Profiling from Kinetic Models of Criminal Behavior. *SIAM Journal on Applied Mathematics*. 2012, vol. 72, n. 1, p. 163–180. DOI: 10.1137/100794080.
- [2] O'LEARY, Mike. The mathematics of geographic profiling. *Journal of Investigative Psychology and Offender Profiling*. 2009, vol. 6, n. 3, p. 253–265. DOI: 10.1002/jip.111.
- [3] O'LEARY, Mike. Multimodel Inference and Geographic Profiling. In: *Crime mapping: A Journal of Research and Practice*. 2011, p. 50–64. ISSN 2152-9876.
- [4] ROSSMO, D. *Geographic profiling*. Boca Raton, Fla: CRC Press, c2000, xxiii, 347 p. ISBN 08-493-8129-0.