Odhad parametrů v pravděpodobnostních modelech velkých systémů

Martin Janžura, ÚTIA, Praha

1 Úvod

Systémem nazýváme soubor veličin (elementů systému), které nabývají hodnoty (stavy) z nějakých obvykle konečných množin. Stavem systému pak rozumíme konfiguraci stavů jeho elementů, tedy prvek příslušného kartézského součinu. Typickým znakem systému (neboť zejména tehdy mluvíme o "systémech") je "velký počet" elementů, tedy vysoká dimenze příslušného stavového prostoru.

Stav systému chápeme jako realizaci diskrétní náhodné veličiny s rozdělením daným nějakou pravděpodobnostní mírou na příslušném kartézském součinu.

Z hlediska matematického je situace tedy poměrně jednoduchá, jedná se prostě o sdružené rozdělení konečného vektoru diskrétních náhodných veličin. Problém nastane, když si uvědomíme, že pouze pro velmi omezenou dimenzi jsme schopni všechny hodnoty pravděpodobnosti uchovat a vlastně i vyčíslit, protože pro vyšší dimenze neumíme spočítat ani potřebnou normalizační konstantu. I když dotyčné rozdělení není obvykle zcela obecné, neboť závislostní struktura často odráží nějakou grafovou strukturu, ať už obecnou (při zpracování rozsáhlých vektorů pozorování konaných na jednotlivých individuích, případně dotazníků, apod.), nebo pravidelnou mřížovou (při zpracování prostorových dat, obrazů apod.), nelze si zpravidla (s výjimkou tzv. "rozložitelných modelů" - viz Hájek, Havránek, Jiroušek (1992)) vypomoci nějakými rekurentními formulemi, které např. existují pro časové řady nebo Markovovy řetězce.

Při řešení problému identifikace systému, který budeme chápat jako úlohu odhadu parametru příslušného rozdělení systému, musíme z praktických důvodů volit takové metody, které jsou numericky vůbec realizovatelné. Zde existují v zásadě dvě možnosti. Ta první radí vyjít od metody maximální věrohodnosti, která je teoreticky optimální, a při implementaci nahradit hodnoty, které nelze vyčíslit, jejich aproximacemi, ať již deterministickými (např. Strauss (1975)) nebo stochastickými Younès (1988)).

My však budeme volit druhou cestu. Použijeme metody odhadu, která sice není optimální, ale míru její "ne-optimality" (tj. zejména ne-eficience a ne-robustnosti) budeme mít pod kontrolou a přitom implementace metody nebude vyžadovat již žádné další simplifikace a aproximace.

Ukážeme také, jak tento postup souvisí s metodou maximální pseudo-věrohodnosti, někdy užívanou (viz Besag (1986)), avšak bez podrobné analýzy statistických vlastností, pro zpracování prostorových dat.

2 Exponenciální rozdělení velkých systémů

Označme

$$A_N = \bigotimes_{i \in N} A_i,$$

kde $1 \ll |N| < \infty$, $2 \leq |A_i| < \infty$ pro každé $i \in N$. Tedy N je konečná ("velká") množina elementů systému, A_i je konečná množina stavů i-tého elementu pro každé $i \in N$, A_N je pak množina stavů systému.

Nechť P je pravděpodobnostní míra na A_N. Pro jednoduchost budeme uvažovat pouze míry kladné, tj.

$$P(a_N) > 0 \qquad \text{pro každé } a_N = (a_i)_{i \in N} \in A_N.$$

Zvolme nyní nějaký pevný "bázický" stav $b_N \in A_N$ a pro každé $V \subset N$, $a_V \in A_V = \bigotimes_{i \in V} A_i$ definujme

$$\Phi_V(a_V) = \sum_{W \subset V} (-1)^{|V \setminus W|} \log P(a_W, b_{N \setminus W}).$$

Funkce Φ_V nazýváme interakcemi a systém interakcí $\Phi = \{\Phi_V\}_{V \subset N}$ potenciálem.

Díky formuli $\sum_{W \subset V} (-1)^{|W|} = 0$ pro $V \neq \emptyset$ snadno ověříme, že

$$P(a_N) = \exp\left\{ \sum_{V \subset N} \Phi_V(a_V) \right\},$$

kde $\Phi_\emptyset = \log P(b_N)$ je normalizační konstanta. Vztah mezi pravděpodobností P a potenciálem Φ se nazývá Möbiova a inverzní Möbiova formule.

Nechť nyní $i \in V \subset N$. Potom můžeme psát

$$\Phi_V(a_V) = \sum_{W \subset V \setminus \{i\}} (-1)^{|V \setminus W|} \log \frac{P^i\left(b_i | a_W, b_{N \setminus (W \cup \{i\})}\right)}{P^i\left(a_i | a_W, b_{N \setminus (W \cup \{i\})}\right)},$$

jak snadno ověříme. Odtud vidíme:

i) Jestliže $a_i = b_i$ pro nějaké $i \in V$, potom $\Phi_V(a_V)$. Přidání této normalizační podmínky již zaručí jednoznačnost potenciálu Φ.

ii) Pomocí systému podmíněných rozdělení $\{P^i\}_{i \in N}$, která nazýváme lokálními charakteristikami, lze zkonstruovat sdružené rozdělení P. Existuje tedy vzájemně jednoznačný vztah mezi sdruženým rozdělením a systémem lokálních charakteristik.

V praktických úlohách můžeme obvykle předpokládat, že lokální charakteristiky závisejí vždy pouze na hodnotách některých veličin, přičemž jejich výběr je dán nějakou grafovou strukturou. Tedy

$$P^i\left(a_i | a_{N \setminus \{i\}}\right) = P^i\left(a_i | a_{\partial i}\right),$$

kde $j \in \partial i$ pokud $(i,j) \in G$, kde G je nějaký graf na N. Uvědomme si, že symetrie, tj. $j \in \partial i$ právě tehdy, když $i \in \partial j$, je nutná vlastnost lokálních charakteristik.

Nechť nyní $i, j \in V \subset N$. Můžeme psát

$$\Phi_V(a_V) = \sum_{W \subset V \setminus \{i,j\}} (-1)^{|V \setminus W|} \log \left[\frac{P^i\left(b_i | b_j, a_W, b_{N\setminus(W \cup \{i,j\})}\right)}{P^i\left(b_i | a_j, a_W, b_{N\setminus(W \cup \{i,j\})}\right)} \frac{P^i\left(a_i | a_j, a_W, b_{N\setminus(W \cup \{i,j\})}\right)}{P^i\left(a_i | b_j, a_W, b_{N\setminus(W \cup \{i,j\})}\right)} \right],$$

jak opět snadno ověříme. Odtud pak plyne

$$\Phi_V \equiv 0 \qquad \text{pokud} \quad (i,j,) \notin G,$$

a proto

$$\Phi_W \equiv 0 \qquad \text{pokud } V \text{ není klika grafu } G.$$

Můžeme tedy uzavřít, že typický pravděpodobnostní model systému bude dán v exponenciálním tvaru

$$P(a_N) = \exp\left\{ \sum_{V \in \mathcal{V}} \Phi_V(a_V) \right\},$$

Kde $\mathcal{V}$ je množina klik nějakého grafu. Jestliže budeme interakce Φ_V považovat za známé až na multiplikativní konstantu, dostaneme tímto přirozeným způsobem klasickou exponenciální rodinu.

3 Maximálně pseudo-věrohodný odhad parametru

Uvažujme nyní obecný model exponenciálních pravděpodobnostních rodin, tj. nechť

$$P_\theta(a_N) = \exp\left\{\theta^T f(a_N)\right\} / c(\theta),$$

kde

$$f = \{f_k : A_N \to R\}_{k=1}^K$$

je vektor známých reálných funkcí,

$$\theta \in R^K$$

je vektor neznámých parametrů a $c(\theta)$ je příslušná normalizační konstanta.

Známý maximálně věrohodný odhad (MLE) θ_n^* parametru $\theta \in R^K$ je dán systémem rovnic

$$\int \left\{ E_{\theta_n^*}[f] - f \right\} d\hat{P}_n = 0,$$

kde $\hat{P}_n$ je empirické rozdělení odpovídající rozsahu výběru n.

Jak již bylo naznačeno v "Úvodu", narazili bychom při výpočtu tohoto odhadu na neschopnost vyčíslit

$$E_\theta[f] = \sum_{a_N \in A_N} f(a_N)\, P_\theta(a_N).$$

Můžeme však vyčislit podmíněné střední hodnoty

$$E_\theta[f|a_V] = \sum_{b_{N\setminus V} \in A_{N\setminus V}} f(a_V, b_{N\setminus V})\, P_\theta^{N\setminus V}(b_{N\setminus V}|a_V)$$

pro každé $a_V \in A_V$, pokud $|N \setminus V| \leq r$, kde r je "přiměřeně malé", a tyto podmíněné střední hodnoty využijeme.

Definujme maximální pseudo-věrohodný odhad (MPLE) obecně předpisem

$$\int \left\{ \sum_{V \subset N} \alpha_V\, E_{\hat\theta_n}[f|a_V] \right\} d\hat{P}_n = 0,$$

kde $\alpha = \{\alpha_V\}_{V \subset N}$ je pevně daný soubor konstant. V praxi budou mít význam pouze takové verze, kde $\alpha_V = 0$ pro $|N \setminus V| > r$. Uvědomme si, že pro MLE je $\alpha_\emptyset = -\alpha_N = 1$, $\alpha_V = 0$ jinak.

První přirozený požadavek na odhad je, aby platilo

$$\hat\theta_n = \theta^0 \quad \text{pokud } \hat{P}_n = P_{\theta^0}.$$

Odtud plyne podmínka

(P I.)
$$\sum_{V \subset N} \alpha_V = 0.$$

Dále obrátíme pozornost k problému existence a konzistentnosti odhadu MPLE. Označme

$$S^\theta(a_N) = \sum_{V \subset N} \alpha_V E_\theta[f|a_V]$$

a současně

$$D^\theta(a_N) = \left(\frac{\partial}{\partial \theta_k} S^\theta(a_N) \right)_{k=1}^{K} = \sum_{V \subset N} \alpha_V \mathrm{var}_{P_\theta(\cdot|a_V)}[f].$$

Budeme předpokládat

(P II.)
$$D^\theta(a_N) > 0$$

(tj. pozitivně definitní) pro každé $\theta \in R^K$, $a_N \in A_N$.

Lemma 3.1

Nechť platí (P II.). Potom pro každou pravděpodobností míru Q na A_N je zobrazení $G_Q : R^K \to R^K$ definované předpisem

$$G_Q(\theta) = \int S^\theta\, dQ$$

prosté a regulární na R^K. Navíc je matice jeho prvních parciálních derivací v každém bodě $\theta \in R^K$ pozitivně definitní.

Důkaz:

Zobrazení G_Q má pozitivně definitní matici prvních parciálních derivací $\nabla G_Q(\theta)$ v každém bodě $\theta \in R^K$, neboť

$$\nabla G_Q(\theta) = \int D^\theta \, dQ > 0$$

díky předpokladu (P II.). Odtud plyne i to, že G_Q je prosté, přičemž spojitost ∇G_Q je zřejmá.

Pro případ MLE má podmínka (P II.) jednoduchý tvar

$$\mathrm{var}_{P_\theta}[f] > 0 \qquad \text{pro každé } \theta \in R^K,$$

která je splněna právě tehdy, pokud $1, f_1, \ldots, f_K$ jsou lineárně nezávislé. Pro obecný MPLE je obtížné podmínku (P II.) charakterizovat. Platí však

$$\mathrm{var}_{P_\theta(\cdot|a_V)}[f] > 0$$

právě když

(P II(a_V)) $\qquad 1, f_1(\cdot, a_V), \ldots, f_K(\cdot, a_V)$ jsou lineárně nezávislé jako funkce na $A_{N \setminus V}$.

Tedy triviálně $\mathrm{var}_{P_\theta(\cdot|a_N)}[f] = 0$.

Stačí tedy pro splnění podmínky (P II.) vzít

$$\alpha_N < 0, \quad \alpha_V \geq 0 \quad \text{pro} \quad V \neq N, \ \sum_{V \neq N} \alpha_V = -\alpha_N,$$

přičemž musí být $\alpha_V > 0$ pro alespoň jedno $V \subset N$ takové, že podmínka (P II(a_V)) je splněna pro každé $a_V \in A_V$.

Hlavním důsledkem Lemmatu 3.1 je to, že pro každou Q existuje spojité inverzní zobrazení

$$(G_Q)^{-1}$$

definované na otevřené množině $G_Q(R^K)$.

Můžeme tedy odhad MPLE přepsat pomocí předpisu

$$\hat{\theta}_n = (G_{\hat{P}_n})^{-1}(0).$$

Odhad $\hat{\theta}_n$ je tedy reálně definován pokud je 0 v oboru hodnot zobrazení $G_{\hat{P}_n}$.

Lemma 3.2

Nechť $\|G_n - G_0\| \to 0$ pro $n \to \infty$, kde pro každé $n = 0, 1, \ldots$ je $G_n : R^K \to R^K$ prosté regulární zobrazení s pozitivně definitní maticí prvních parciálních derivací v každém bodě. Nechť $G_0(\theta^0) = 0$. Potom existuje n^0 tak, že $G_n^{-1}(0)$ je definováno pro $n \geq n^0$, a $G_n^{-1}(0) \to \theta^0$.

Důkaz:

Zvolme $\delta > 0$. Potom

$$\min_{\|\theta-\theta^0\|=\delta} \|G_0(\theta)\| = \varepsilon > 0.$$

Nechť $\|G_n - G_0\| < \frac{\varepsilon}{3}$. Potom $\|G_n(\theta^0)\| < \frac{\varepsilon}{3}$ a $\min_{\|\theta-\theta^0\|=\delta} \|G_n(\theta)\| > \frac{2\varepsilon}{3}$.

Funkce $\|G_n(\theta)\|^2$ má tedy uvnitř koule $\{\theta;\ \|\theta - \theta^0\| < \delta\}$ lokální minimum θ^n, a proto platí

$$0 = \nabla\|G_n(\theta^n)\|^2 = 2\,[\nabla G_n(\theta^n)]\; G_n(\theta^n).$$

Jelikož $\nabla G_n(\theta^n) > 0$, musí být $G_n(\theta^n) = 0$. Přitom $\|\theta^n - \theta^0\| < \delta$ a G_n je prosté, proto $\theta^n = G_n^{-1}(0)$.

Věta 3.3. Odhad

$$\hat{\theta}_n = \left(G_{\hat{P}_n}\right)^{-1}(0)$$

je pro dostatečně velká n definován a je konzistentní, tj.

$$\hat{\theta}_n \to \theta^0 \quad \text{s.j. } [P_{\theta^0}].$$

Důkaz.

Jelikož $G_{P_{\theta^0}}(\theta^0) = 0$ a $\|G_{P_{\theta^0}}(\theta) - G_{\hat{P}_n}(\theta)\| \leq \text{konst.}\ \|P_{\theta^0} - \hat{P}_n\|$, je tvrzení věty přímým důsledkem Lemmatu 3.2 a silného zákona velkých čísel, podle kterého $\|P_{\theta^0} - \hat{P}_n\| \to 0$ s.j. $[P_{\theta^0}]$.

4 Asymptotická normalita a eficience

Vyšetříme nyní další vlastnosti navrženého odhadu.

Věta 4.1

Odhad $\hat{\theta}_n = \left(G_{\hat{P}_n}\right)^{-1}(0)$ je při skutečné hodnotě parametru θ^0 asymptoticky normální s asymptotickým rozptylem

$$\sigma^2_{\theta^0} = \left[\int D^{\theta^0}\, dP_{\theta^0}\right]^{-1} \left[\operatorname{var}_{P_{\theta^0}}[S^{\theta^0}]\right] \left[\int D^{\theta^0}\, dP_{\theta^0}\right]^{-1}.$$

Důkaz:

Podle definice platí

$$0 = \int S^{\hat{\theta}_n} d\hat{P}_n = \int S^{\theta^0} d\hat{P}_n + \int D^{\gamma_n \hat{\theta}_n + (1-\gamma_n)\theta^0} d\hat{P}_n,$$

kde $\gamma_n \in [0, 1]$.

Přitom

$$\int D^{\gamma_n\hat{\theta}_n+(1-\gamma_n)\theta^0} d\hat{P}_n \longrightarrow \int D^{\theta^0}.dP_{\theta^0} = F_{\theta^0} \quad \text{s. j. } [P_{\theta^0}],$$

neboť $\hat{\theta}_n \to \theta^0$ a $\hat{P}_n \to P_{\theta^0}$ s. j. $[P_{\theta^0}]$ podle Věty 3.3 a vše je spojité.

Současně

$$n^{1/2}\int S^{\theta_0} d\hat{P}_n \longrightarrow N_K(0, V_{\theta^0}) \quad \text{v distribuci } [P_{\theta^0}],$$

kde $V_{\theta^0} = \mathrm{var}_{P_{\theta^0}}[S^{\theta^0}]$, podle centrální limitní věty (připomeňme $\int S^{\theta^0} dP_{\theta^0} = 0$). Proto

$$n^{1/2}(\hat{\theta}_n - \theta^0) \longrightarrow N_K\left(0, F_{\theta^0}^{-1} V_{\theta^0} F_{\theta^0}^{-1}\right) \quad \text{v distribuci } [P_{\theta^0}].$$

Pro stále pevné $\alpha = \{\alpha_V\}_{V\subset N}$ označíme $\mathcal{V} = \{V;\ \alpha_V \neq 0\}$. Nechť

$$\overline{V} \supset \bigcup_{V\in\mathcal{V}} V, \qquad \underline{V} \subset \bigcap_{V\in\mathcal{V}} V.$$

Označme dále

$$T^{\theta}(a_N) = E_{\theta}[f|a_{\underline{V}}] - E_{\theta}[f|a_{\overline{V}}].$$

Lemma 4.2

Platí

i) $\int D^{\theta} dP_{\theta} = \mathrm{cov}_{P_{\theta}}(S^{\theta}, T^{\theta}) = \mathrm{cov}_{P_{\theta}}(T^{\theta}, S^{\theta})$;

ii) $[\int D^{\theta} dP_{\theta}]^{-1}[\mathrm{var}_{P_{\theta}} S^{\theta}][\int D^{\theta} dP_{\theta}]^{-1} - \left[\mathrm{var}_{P_{\theta}}[T^{\theta}]\right]^{-1} \geq 0.$

Důkaz:

Jelikož pro každé $V \in \mathcal{V}$ je

$$\int E_{\theta}[f|a_V]\, E_{\theta}[f|a_V]^T\, dP_{\theta} = \int E_{\theta}[f|a_V]\, E_{\theta}[f|a_{\overline{V}}]^T\, dP_{\theta}$$

a také

$$\int E_{\theta}[f|a_V]\, E_{\theta}[f|a_{\underline{V}}]^T\, dP_{\theta} = \int f E_{\theta}[f|a_{\underline{V}}]^T\, dP_{\theta},$$

obdržíme

$$\begin{aligned}\int D^{\theta}\, dP_{\theta} &= \int \left\{\sum_{V\in\mathcal{V}} \alpha_V \mathrm{var}_{P_{\theta}(\cdot|a_V)}[S^{\theta}]\right\} dP^{\theta} = \\ &= \mathrm{cov}(S^{\theta}, T^{\theta}) + \sum_{V\in\mathcal{V}} \alpha_V \cdot \left[\mathrm{var}_{P_{\theta}}[f] - \int f E_{\theta}[f|a_{\underline{V}}]^T\, dP_{\theta}\right].\end{aligned}$$

Přitom $\sum_{V\in\mathcal{V}} \alpha_V = 0$, a tím jsme dokázali i), když symetrie je zřejmá.

Tvrzení ii) pak plyne ze známé vlastnosti kovariančních matic

$$\mathrm{var}_{P_{\theta}}[S^{\theta}] - \mathrm{var}_{P_{\theta}}(S^{\theta}, T^{\theta}) \left\{\mathrm{var}_{P_{\theta}}[T^{\theta}]\right\}^{-1} \mathrm{cov}_{P_{\theta}}(T^{\theta}, S^{\theta}) \geq 0.$$

Věta 4.3

Žádná volba $\alpha = \{\alpha_V\}_{V \subset N}$ nezaručí větší eficienci než má maximálně věrohodný odhad.

Důkaz:

V předchozím lemmatu můžeme samozřejmě volit $\overline{V} = N$, $\underline{V} = \emptyset$. Tvrzení ii) pak přímo poskytne požadovanou nerovnost.

Matice v tvrzení ii) Lemmatu 4.2 je zřejmě nulová pokud $S^\theta = M_\theta T^\theta$ s.j. $[P_{\theta^0}]$, kde M^θ je nějaká deterministická matice. Je striktně pozitivně definitní, pokud pro žádnou dvojici deterministických vektorů neplatí $a(\theta)^T S^\theta = b(\theta)^T T^\theta$ s.j.$[P_\theta]$.

Je zřejmé, že náš odhad nebude obvykle eficientní, neboť z důvodu numerické realizovatelnosti, jak už bylo uvedeno výše, bude nutné volit $\alpha_V = 0$ pro V malá.

5 Robustnost

Pro zkoumání robustnosti definujme funkci $\theta(\varepsilon, Q)$ pro libovolnou míru Q na A_N předpisem

$$0 = \int S^{\theta(\varepsilon,Q)} d\left((1-\varepsilon) P_{\theta^0} + \varepsilon Q\right).$$

Označme

$$\theta'(\varepsilon, Q) = \left(\frac{d\theta_i(\varepsilon, Q)}{d\varepsilon}\right)_{i=1}^{K}.$$

Lemma 5.1

Platí

$$\theta'(0, Q) = -\left[\int D^{\theta^0} dP_{\theta^0}\right]^{-1} \int S^{\theta^0} dQ.$$

Důkaz:

Můžeme psát

$$0 = \int S^{\theta(\varepsilon,Q)} d(Q - P_{\theta^0}) + \left[\int D^{\theta(\varepsilon,Q)} d\left((1-\varepsilon) P_{\theta^0} + \varepsilon Q\right)\right] \theta'(\varepsilon, Q).$$

Odtud pro $\varepsilon = 0$ obdržíme

$$0 = \int S^{\theta^0} dQ + \left[\int D^{\theta^0} dP_{\theta^0}\right] \theta'(0, Q).$$

Uvažujme nyní místo obecné míry Q míru δ_{a_N} degenerovanou v jednom bodě $a_N \in A_N$. Potom

$$q(a_N) = \theta(0, \delta_{a_N}) = -\left[\int D^{\theta^0} dP_{\theta^0}\right]^{-1} S^{\theta^0}(a_N).$$

Vidíme, že platí

$$E_{\theta^0}[q] = 0$$
$$\mathrm{var}_{P_{\theta^0}}[q] = \left[\int D^{\theta^0}\, dP_{\theta^0}\right]^{-1} \mathrm{var}_{P_{\theta^0}}[S^{\theta^0}] \left[\int D^{\theta^0}\, dP_{\theta^0}\right]^{-1}.$$

Věta 5.2

Pro žádnou volbu $\alpha = \{\alpha_V\}_{V \subset N}$ není odhad robustnější než maximálně věrohodný, neboť influenční funkce q je rozptýlenější.

Důkaz: Plyne z předchozích úvah a nerovností části 4.

Vidíme tedy, že náš odhad je nejenom méně eficientní, ale i méně robustní než maximálně věrohodný. Rozhodující kritérium však musí být stále jeho numerická realizovatelnost, která je zaručena pouze vhodnou volbou parametrů $\alpha = \{\alpha_V\}_{V \subset N}$.

6 Maximálně pseudo-věrohodný odhad z prostorových dat

Mějme nyní $N = \{t \in Z^2;\ 0 \le t_i < T,\ i = 1,2\}$, $A_t = A_0 = \{0,1\}$ pro každé $t \in N$.

Pro $a_N \in A_N$ definujme

$$f_0(a_N) = \sum_{t \in N} a_t$$
$$f_1(a_N) = \sum_{\substack{t,s \in N \\ t-s=[0,1]}} a_t \cdot a_s$$
$$f_2(a_N) = \sum_{\substack{t,s \in N \\ t-s=[1,0]}} a_t \cdot a_s.$$

Tak obdržíme tzv. Isingův model, kde θ_1, θ_2 mají význam vertikální, resp. horizontální interakce. θ_0 je tzv. chemický potenciál.

Pro vnitřní body $t \in N$ $(1 \le t_i \le T_2,\ i = 1,2)$ platí

$$P_\theta^t\left(a_t | a_{N\setminus\{t\}}\right) = \frac{\exp\{a_t(\theta_0 + \theta_1(a_{t+[0,1]} + a_{t-[0,1]}) + \theta_2(a_{t+[1,0]} + a_{t-[1,0]}))}{U(a_{\partial t})},$$

kde $U(a_{\partial t})$ je příslušná normalizační konstanta a $\partial t = \{s \in N;\ \|t-s\| = 1\}$.

Budeme-li nyní (viz Besag (1986)) definovat maximálně pseudo-věrohodný odhad

$$\hat{\theta}_n = \mathrm{argmax}_\theta \int \sum_{t \in T} \log P_\theta^t(a_t | a_{\partial t})\, d\hat{P}_n,$$

dostaneme vlastně náš výše definovaný odhad při

$$\alpha_N = -|N|, \quad \alpha_{\{i\}} = 1 \qquad \text{pro každé } i \in N.$$

Literatura

Besag, J. (1986). On the statistical analysis of dirty pictures (with discussion). J. Roy. Statist. Soc. Ser. B 48, 259–302.

Hájek, P., Havránek, T., Jiroušek, R. (1992). Uncertain Information Processing in Expert Systems. CRC Press, Londýn.

Strauss, D. J. (1975). Analysis binary lattice data with the nearest-neighbor property. J. Appl. Probab. 12, 702–712.

Younès, L. (1988): Estimation and annealing for Gibbsian fields. Ann. Inst. H. Poincaré Lect. 8 (N. S.) 24, 269–294.