

# ODHADY PARAMETRŮ GAUSSOVSKÝCH NÁHODNÝCH POLÍ

Martin JANŽURA<sup>1</sup>

ÚTIA ČAV

**Abstract.** Two methods of parameter estimation for Gibbs–Markov random fields are introduced. Their theoretical background and statistical properties are studied.

**Резюме:** В статье представлены два метода оценки параметров гауссовских марковских случайных полей. Изучаются теоретические основы и статистические свойства.

## 1. ÚVOD

Uvažujme náhodné pole

$$\{X_t\}_{t \in T},$$

kde  $T = Z^d$ ,  $d \geq 1$ , je  $d$ -rozměrná celočíselná síť a  $X_t$  je pro každé  $t \in T$  reálná náhodná veličina. Pro  $d = 1$  se takové procesy obvykle nazývají časovými řadami. Pro  $d \geq 2$  pak náhodná pole popisují náhodné procesy, které probíhají ne v čase, ale v prostoru (časoprostoru), s přímou aplikací v prostorové statistice a zpracování obrazu.

Nejpřirozenějším a nejčastěji užívaným parametrickým modelem je autoregresní (AR) model, který bychom v našem případě zapsali vztahem

$$\sum_{t \in L} a_t X_{t+s} = \varepsilon_s \quad \text{pro každé } s \in T,$$

kde  $L \subset T$ ,  $|L| < \infty$ , je konečný „řád“ autoregrese a  $\{a_t\}_{t \in L}$  jsou autoregresní koeficienty (pro jednoduchost předpokládáme nulovou střední hodnotu a jednotkový rozptyl „chyb“  $\varepsilon_t$ ,  $t \in T$ ).

Podstatným předpokladem AR modelu je pak nekorelovanost chyb

$$\varepsilon_t \perp \varepsilon_s \quad \text{pro } t \neq s.$$

Je však několik důvodů, pro které je užití AR modelu pro náhodná pole nevhodné. Abychom totiž mohli využít jeho výhody a dosáhli skutečné analogie s časovými řadami, museli bychom uvažovat „kauzální“ tvar, tj. např.

$$0 = \max_{t \in L} t,$$

kde maximum je chápáno ve smyslu nějakého lineárního (např. lexikografického) uspořádání. Jedině potom můžeme sestavit Yule–Walkerovy rovnice, apod. Takový předpoklad je však pro náhodná pole zcela umělý, těžko interpretovatelný a obvykle neodpovídající realitě, neboť prostorové vazby zpravidla závisejí na „vzdálenosti“ a ne na „pořadí“ (pro  $d = 1$  to ovšem

<sup>1</sup>Tento příspěvek vznikl za přispění grantu GA ČR 202/96/0731.

víceméně splývá). Ještě závažnější problém však vyplývá z odlišných vlastností komplexních polynomů jedné a více proměnných. Spektrální hustota AR modelu má totiž známý tvar

$$\begin{aligned} f^{\text{AR}}(\omega) &= \left| \sum_{t \in L} a_t e^{i\langle t, \omega \rangle} \right|^{-2} \\ &= \left( \sum_{p \in L-L} b_p e^{i\langle p, \omega \rangle} \right)^{-1} \quad \text{pro každé } \omega \in [-\pi, \pi]^d, \end{aligned}$$

kde  $b_p = b_{-p} = \sum_{s \in L \cap (L-p)} a_s a_{s+p}$ .

Na rozdíl od  $d = 1$  pro  $d \geq 2$  taková třída polynomů zdaleka nevyčerpá prostor všech konečných polynomů se symetrickými koeficienty.

Např. pro  $d = 2$ ,  $\omega = (\omega_1, \omega_2)$ , nelze polynom

$$P(\omega_1, \omega_2) = A + B \cos \omega_1 + C \cos \omega_2$$

faktorizovat ve tvaru

$$P = |Q|^2$$

pro žádný konečný polynom  $Q$ .

Uvažováním AR modelů bychom se tedy zbytečně ochudili o celou řadu přirozených jednoduchých modelů. Je tedy přirozené ve výše uvedeném vztahu

$$\sum_{t \in L} a_t X_{t+s} = \varepsilon_s$$

předpokládat spíše

$$\varepsilon_s \perp X_t \quad \text{pro } s \neq t.$$

Takové modely budeme nazývat markovskými poli (MP); jejich spektrální hustoty mají přímo tvar

$$f^{\text{MP}}(\omega) = \left( \sum a_t e^{i\langle t, \omega \rangle} \right)^{-1} \quad \text{pro každé } \omega \in [-\pi, \pi]^d.$$

Markovská pole tvoří tedy širší třídu, která v sobě AR modely zahrnuje. Budeme se proto zabývat přímo jimi.

## 2. MARKOVSKÁ POLE

Uvažujme spektrální hustotu náhodného pole ve tvaru

$$f_U(\omega) = \left[ 2 \sum_{k \in M} U_k \cos \langle k, \omega \rangle \right]^{-1} \quad \text{pro každé } \omega \in [-\pi, \pi]^d,$$

kde  $M \subset T$ ,  $|M| < \infty$ , splňuje  $M \cap (-M) = \{0\}$  a

$$U = (U_k)_{k \in M} \in \mathcal{D}_m = \{U \in \mathcal{R}^M; 0 < f_U < \infty\}$$

je vektor koeficientů zajišťující podmínku kladnosti a konečnosti spektrální hustoty na celém definičním oboru.

Potom obdržíme příslušný člen kovarianční funkce (nekonečné matice)  $\mathcal{R}^U$  ve tvaru

$$(I) \quad \mathcal{R}^U(p, q) = \mathcal{R}^U(p - q) = \int_{[-\pi, \pi]^d} e^{i(p-q, \omega)} f_U(\omega) d\omega$$

pro každé  $p, q \in T$ .

Snadno se ověří, že platí

$$(II) \quad \mathcal{R}^U = (A^U)^{-1},$$

kde

$$A^U = 2 \sum_{k \in M} U_k I^k$$

a

$$I^k(p, q) = \frac{1}{2} (\delta(p, q + k) + \delta(p, q - k))$$

pro každé  $p, q \in T$ . Zde  $\delta(\cdot, \cdot)$  je známý Kroneckerův symbol, a proto je matice  $I^k$  nulová až na symetricky  $k$ -tou a  $(-k)$ -tou vedlejší diagonálu, kde jsou poloviny. Pro  $k = 0$  dostaneme přímo jednotkovou matici  $I^0 = I$ .

Uvedené dva způsoby zavedení kovarianční funkce (I) a (II) jsou ekvivalentní. Mohli bychom též začít s definicí (II) a získali bychom spektrální hustotu  $f_U$ .

Případnou nenulovou střední hodnotu pak zapíšeme ve tvaru

$$\mu^\theta = -h \cdot f_U(0) = -h \cdot \left[ 2 \sum_{k \in M} U_k \right]^{-1},$$

kde  $h \in \mathcal{R}$  a proto

$$\theta = (h, U) \in \mathcal{R} \times \mathcal{D}_M$$

je tedy celým parameterem modelu.

Takové zavedení střední hodnoty sice vypadá poněkud uměle, je však zcela obecné, neboť  $h$  je libovolné a  $f_U(0) > 0$ , a jak uvidíme v dalším, má i dobrou interpretaci a svoji užitečnost.

### 3. ODHAD PARAMETRŮ

Je zřejmé, že relevantní odhady parametru budou založeny na přirozených statistikách, tj. výběrových kovariancích a středních hustotách. Potřebujeme pak nalézt vhodný vztah mezi parametry modelu a těmito statistikami. Jako východisko se zde přímo nabízí vztahy (I) a (II) z předchozího odstavce, které vedou na dva různé odhady.

I. „Momentová metoda“. V nekonečné soustavě rovnic (I) stačí zřejmě při známé kovarianční funkci  $\mathcal{R}(\cdot)$  k určení parametru  $U \in \mathcal{D}_M$  konečná soustava

$$\mathcal{R}(m) = \mathcal{R}^U(m) \quad \text{pro } m \in M.$$

Lze dokonce ukázat, že zobrazení  $U \mapsto \{\mathcal{R}^U(m)\}_{m \in M}$  je regulární. Stačí tedy dosadit empirické hodnoty (odhady) kovariancí  $\widehat{\mathcal{R}(m)}$ ,  $m \in M$ , a řešení

$$\widehat{U}^I = \{\widehat{U}_k^I\}_{k \in M}$$

soustavy

$$(S_I) \quad \widehat{\mathcal{R}(m)} = \mathcal{R}^U(m) = \int_{[-\pi, \pi]^d} e^{i\langle m, \omega \rangle} f_U(\omega) d\omega \quad \text{pro } m \in M$$

pak budeme považovat za odhad parametru  $U$ .

Zdůrazněme, že soustava (SI) je značně nelineární a její řešení, vyžadující např. numerické integrování, je složitá a zdlouhavá numerická záležitost.

Odhad parametru  $h$  pak jen dopočteme jako

$$\widehat{h}^I = - \sum_{k \in M} \widehat{U}_k^I \widehat{\mu}_k,$$

kde  $\widehat{\mu}_k$  pro  $k \in M$  jsou odhady (které mohou, ale nemusejí být různé) střední hodnoty.

Jestliže jsou odhady  $\widehat{\mathcal{R}(m)}$ ,  $\widehat{\mu}_k$  konzistentní, je takový i odhad

$$\widehat{\theta}^I = (\widehat{h}^I, \widehat{U}^I).$$

Přístup založený na řešení soustavy (SI) může být chápán jako momentová metoda.

II. „Yule – Walkerovy rovnice“. Vztah (II) můžeme přepsat jako soustavu

$$2 \sum_{k \in M} U_k I^k \mathcal{R}^U = I^0,$$

kde

$$(I^k \mathcal{R}^U)(p, q) = \frac{1}{2} (\mathcal{R}^U(p - q + k) + \mathcal{R}^U(p - q - k)).$$

Označíme-li  $p - q = j$ , pak k jednoznačnému určení vektoru  $U = \{U_k\}_{k \in M}$  opět stačí konečná soustava

$$2 \sum_{k \in M} U_k \frac{\mathcal{R}^U(j + k) + \mathcal{R}^U(j - k)}{2} = \delta(j) \quad \text{pro } j \in M.$$

Můžeme nyní dosadit empirické hodnoty místo neznámých kovariancí a obdržíme odhad jako řešení

$$\widehat{U}^{II} = \{\widehat{U}_k^{II}\}_{k \in M}$$

lineární soustavy

$$(S_{II}) \quad \sum_{k \in M} U_k (\mathcal{R}(j + k) + \mathcal{R}(j - k)) = \delta(j) \quad \text{pro } j \in M.$$

Řešení takovéto lineární soustavy je přirozeně numericky mnohem snazší než soustavy (SI).

Odhad parametru  $h$  pak opět dopočteme podobně jako nahoře, tj.

$$\widehat{h}^{II} = - \sum_{k \in M} \widehat{U}_k^{II} \widehat{\mu}_k,$$

a celý odhad

$$\hat{\theta}^{\text{II}} = (\hat{h}^{\text{II}}, \hat{U}^{\text{II}})$$

je konzistentní, pokud jsou konzistentní příslušné výběrové momenty.

I když soustava (S<sub>II</sub>) není úplnou analogií Yule – Walkerových rovnic, můžeme pracovně použít tento název vzhledem k lineárnímu charakteru vazeb mezi neznámými parametry a kovariancemi.

Máme tedy dva konzistentní odhady neznámých parametrů, zbývá však celá řada dalších otevřených problémů. V následujících odstavcích se budeme zabývat zejména těmito dvěma otázkami:

A). Jak definovat empirické kovariance a střední hodnoty tak, aby byly konzistentní, případně jaké další vlastnosti by mohly mít?

B). Lze odvodit výše uvedené odhady z některých obecných principů statistického odhadování (jako je např. princip maximální věrohodnosti) a jaké z toho případně plynou další jejich vlastnosti?

#### 4. EMPIRICKÉ POLE

Předpokládejme, že pozorovaná data určují nějakým způsobem empirické pole

$$\hat{P},$$

tj. pravděpodobnostní míru na prostoru  $\mathcal{R}^T$ .

Příslušné empirické momenty pak chápeme jako integrály počítané s pomocí této míry, tj. např.

$$\begin{aligned} \widehat{\mathcal{R}(m)} &= \int X_0 X_m d\hat{P} - \widehat{\mu}_0 \widehat{\mu}_m \\ \widehat{\mu}_m &= \int X_m d\hat{P} \end{aligned}$$

apod. Na empirické pole  $\hat{P}$  můžeme mít některé požadavky, od nichž se pak odvine i konkrétní definice.

**i) Stacionární empirické pole.** Požadujeme, aby pro každou  $f \in L_1(\hat{P})$  platilo

$$\int f d\hat{P} = \int f \circ \tau_t d\hat{P},$$

kde

$$\tau_t : \mathcal{R}^T \rightarrow \mathcal{R}^T$$

je pro každé  $t \in T$  příslušné posunutí dané předpisem

$$\tau_t(x)_s = x_{t+s} \quad \text{pro každé } s \in T, x \in \mathcal{R}^T.$$

Potom musíme předpokládat, že pozorovací oblast  $W \subset T$  je pravoúhlá (tj. obdélník), a položíme

$$\int f d\hat{P} = |W|^{-1} \sum_{t \in W} f \circ \tau_t(x_W^{\text{PER}}),$$

kde  $x_W^{\text{PER}} \in \mathcal{R}^T$  je periodické prodloužení konfigurace  $x_W \in \mathcal{R}^W$ .

Je ovšem zřejmé, že takový odhad není nevychýlený, tj. obecně

$$\mathbb{E} \int f d\hat{P} \neq \mathbb{E} f.$$

Současně je zřejmé, že pro lokální (cylindrické, závisující pouze na konečném počtu souřadnic) funkce  $f$  je toto vychýlení úměrné velikosti hranice množiny  $|\partial W|$ , kde

$$\partial W = \bigcup_{\|s\|=1} W \cap (W - s)^c.$$

Abychom tedy vychýlení eliminovali alespoň asymptoticky, musíme požadovat, aby pozorovací oblast  $W_n$  rostla způsobem, který zaručuje

$$\lim_{n \rightarrow \infty} |W_n|^{-1} |\partial W_n| = 0,$$

což je ekvivalentní s podmínkou

$$\lim_{n \rightarrow \infty} |W_n|^{-1} |W_n \cap (W_n - s)^c| = 0 \quad \text{pro všechna } s \in T.$$

Toto je standardním předpokladem splněným např. pro posloupnost krychlí nebo obecněji obdélníků, jejichž všechny strany rostou do nekonečna. Tak je zaručena platnost ergodické věty a tedy konzistentnost výběrových momentů. Pro platnost centrální limitní věty a tedy asymptotické normality odhadů bychom ovšem potřebovali

$$\lim_{n \rightarrow \infty} |W_n|^{-\frac{1}{2}} |\partial W_n| = 0,$$

což jednoduše platí pro  $d = 1$ , kdy pro interval  $W_n$  platí  $|\partial W_n| = 2$ , ale obecně, je-li  $W_n$  krychle o straně  $n$ , platí

$$|W_n| = n^d, \quad |\partial W_n| = 2d n^{d-1}$$

a proto

$$\lim_{n \rightarrow \infty} |W_n|^{-\frac{1}{2}} |\partial W_n| = \lim_{n \rightarrow \infty} 2d n^{d-1-\frac{d}{2}} = \begin{cases} 2d & \text{pro } d = 2 \\ \infty & \text{pro } d \geq 3. \end{cases}$$

Je tedy zřejmé, že z hlediska asymptotické normality odhadu je výše definované stacionární empirické pole zcela nevhodné.

ii) **Nevychýlené empirické pole.** Budeme tedy přímo požadovat, aby platilo

$$\mathbb{E} \int f d\widehat{P} = \mathbb{E} f$$

alespoň pro nějakou dostatečně bohatou třídu funkcí  $f$ . Nechtě  $\mathcal{L}_A$  je množina lokálních funkcí, které závisí pouze na souřadnicích z konečné množiny  $A \subset T$ ,  $|A| < \infty$ . Potom pro  $f \in \mathcal{L}_A$  položíme

$$\int f d\widehat{P} = |W^A|^{-1} \sum_{t \in W^A} f \circ \tau_t(x_W),$$

kde  $W^A = \{t \in W; t + A \subset W\}$ . Je zřejmé, že empirický integrál  $\int f d\widehat{P}$  je nevychýleně definován, pokud  $W^A \neq \emptyset$ . Máme např.

$$\int X_j d\widehat{P} = |W \cap (W - j)|^{-1} \sum_{t \in W \cap (W - j)} x_t$$

a podobně

$$\int X_j X_k d\widehat{P} = |W \cap (W - j) \cap (W - k)|^{-1} \sum_{t \in W \cap (W - j) \cap (W - k)} x_{t+j} x_{t+k}.$$

Dále vidíme, že pro  $f \in \mathcal{L}_A$  a  $s \in T$  platí např.

$$\mathbb{E} \left| \int f d\widehat{P} - \int f \circ \tau_s d\widehat{P} \right| \leq K(A, s) \cdot \|f\|_1 |W^A|^{-1} |\partial W^A|,$$

kde  $K(A, s)$  je konstanta závislejší na  $A$  a  $s$ . Pokud bude opět splněna výše uvažovaná podmínka

$$|W_n|^{-1} |\partial W_n| \rightarrow 0$$

pro růst pozorovací oblasti, bude empirické pole alespoň asymptoticky stacionární. Navíc zde pozorovací oblasti  $W_n$  nemusejí být striktně pravoúhlé, ani nemusíme do  $W_n$  žádný pravoúhelník vnořovat, jak to požaduje postup i). Můžeme však přednosti obou definic spojit, pokud budeme pro  $f \in \mathcal{L}_A$  uvažovat

$$\int f d\widehat{P} = |\widetilde{W}^A|^{-1} \sum_{t \in \widetilde{W}^A} f \circ \tau_t,$$

kde nyní

$$\widetilde{W}^A = \{t \in T; t + A \subset W\}.$$

Takto definované pole splňuje pro každé  $f \in \mathcal{L}_A$ , kde  $\widetilde{W}^A \neq \emptyset$ , jak

$$\mathbb{E} \int f d\widehat{P} = \mathbb{E} f$$

tak i

$$\int f d\widehat{P} = \int f \circ \tau_s d\widehat{P}$$

neboť  $\widetilde{W}^{A+s} = \widetilde{W}^A - s$  pro každé  $s \in T$ .

Vzhledem k markovskému předpokladu nám stačí mít empirické pole definované pro  $f \in \mathcal{L}_A$  s nějakým konečným  $A \subset T$ .

## 5. GAUSSOVSKÁ MARKOVSKÁ POLE

Budeme nyní navíc předpokládat, že náhodné pole s výše definovanou kovarianční funkcí a střední hodnotou je gaussovské, tzn. že každé konečněrozměrné marginální rozdělení  $p_V^\theta$ ,  $V \subset T$ ,  $|V| < \infty$  je dáno

$$p_V^\theta(x_V) = (2\pi)^{-\frac{|V|}{2}} |R_{VV}^U|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (x_V - \mu_V^\theta)^\top (R_{VV}^U)^{-1} (x_V - \mu_V^\theta) \right\},$$

kde  $x_V$ ,  $\mu_V^\theta$ ,  $R_{VV}^U$  jsou příslušné vektory, resp. matice zúžené na podmnožinu  $V \subset T$ .

Pro dostatečně velká  $W \subset T$  můžeme využít přibližné vzorce, které platí asymptoticky, zejména

$$-|W|^{-1} \log |R_{WW}^U| \approx (2\pi)^{-d} \int_{[-\pi, \pi]^d} \log \left( 2 \sum_{k \in M} U_k \cos \langle k, \omega \rangle \right) d\omega$$

a také

$$(R_{WW}^U)^{-1} \approx A_{WW}^U = 2 \sum_{k \in M} U_k I_{WW}^k.$$

(Viz např. Janžura (1988 a, b)).

Po dosazení obdržíme přibližný výraz pro logaritmus věrohodnosti

$$\begin{aligned} -|W|^{-1} \log p_W^\theta(x_W) &\approx \\ &\approx \frac{1}{2} \left[ \log(2\pi) - (2\pi)^{-d} \int_{[-\pi, \pi]^d} \log \left( 2 \sum U_k \cos \langle k, \omega \rangle \right) d\omega + \right. \\ &\quad \left. + 2 \sum_{k \in M} U_k \left[ \widehat{R}(k) + (\widehat{\mu} - \mu^\theta) \right]^2 \right] = \widehat{L}(\theta), \end{aligned}$$

když opět aproximujeme  $|W|^{-1} x_W^\top I_{WW}^k x_W \approx \widehat{R}(k) + (\widehat{\mu})^2$

$$|W|^{-1} \mathbf{1}_W^\top I_{WW}^k x_W \approx \widehat{\mu}$$

pro všechna  $k \in M$ .

Stejnou aproximaci dostaneme i pro podmíněné rozdělení

$$\begin{aligned} p_{V|V^c}^\theta(x_V | x_{V^c}^c) &= (2\pi)^{-\frac{|V|}{2}} |A_{VV}^U|^{-\frac{1}{2}} \\ &\quad \cdot \exp \left\{ -\frac{1}{2} [A_{VT}^U x_T + h \mathbf{1}_V]^\top (A_{VV}^U)^{-1} [A_{VT}^U x_T + h \mathbf{1}_V] \right\}. \end{aligned}$$

Zde dobře vidíme význam parametrů  $\theta = (h, U)$ , které v podmíněném rozdělení vystupují přímo. Jde totiž o tzv. gibbovskou parametrizaci, kdy jsou parametricky zadávány právě tyto podmíněné distribuce.



**Poznámka:** Také se snadno ukáže, že pole chyb  $(\varepsilon_t)_{t \in T}$  v markovském modelu je nyní též gaussovské se střední hodnotou rovnou  $-h$  a kovarianční funkcí  $A_{TT}^U$ . Platí přitom

$$\varepsilon_t + h = 2U_0 [X_t - E[X_t|X_s, s \neq t]]$$

pro každé  $t \in T$ .

Vraťme se k problematice odhadu. Máme tedy přibližný výraz  $\widehat{L}(\theta)$  pro (minus) logaritmus věrohodnosti. Jelikož řád odchylky je dostatečně malý, získáme minimalizací  $\widehat{L}(\theta)$  asymptoticky maximálně věrohodný odhad se všemi jeho příznivými vlastnostmi, jako je asymptotická normalita a eficiency. Jestliže však položíme rovny nule parciální derivace výrazu  $\widehat{L}(\theta)$ , obdržíme přímo soustavu rovnic  $(S_I)$  a jejím řešením, tedy i řešením maximalizace  $\widehat{L}(\theta)$  je odhad „momentovou metodou“

$$\widehat{\theta}^I = (\widehat{h}^I, \widehat{U}^I).$$

Ke stejnému výsledku dospějeme i s využitím principu odhadu s minimální vzdáleností. Obecně, je-li

$$\widetilde{\theta} \in \arg \min_{\theta \in \Theta} D(\widehat{P}, P^\theta),$$

kde  $D(\cdot, \cdot)$  je nějaká rozumně definovaná „vzdálenost“ pravděpodobnostních měr,  $\widehat{P}$  je empirická distribuce a  $P^\theta$  je teoretické rozdělení z parametrické třídy  $\{P_\theta\}_{\theta \in \Theta}$ , nazveme  $\widetilde{\theta}$  odhadem s minimální vzdáleností.

Jestliže pro náhodná pole vezmeme jako vzdálenost asymptotickou  $I$ -divergenci

$$I(P, Q) = \lim_{W \nearrow T} |W|^{-1} \int \log \frac{dP_W}{dQ_W} dP,$$

dostaneme

$$I(\widehat{P}, P^\theta) = \widehat{L}(\theta) - \mathcal{H}(\widehat{P}),$$

kde  $\mathcal{H}(\widehat{P})$  je rychlost entropie empirického pole  $\widehat{P}$ , a jelikož nezávisí na neznámém parametru, můžeme ji z minimalizace vynechat.

Shrňme tedy výsledky tohoto odstavce.

**Tvrzení.** Pro gaussovská pole je odhad  $\widehat{\theta}^I$  asymptoticky maximálně věrohodný a současně je odhadem s minimální  $I$ -divergencí. Proto je také asymptoticky normální a eficientní.

**Důkaz.** Viz úvahy předcházející tvrzení a také např. Janžura (1988 a, b).  $\square$

Důvod, pro který se neuvažuje odhad exaktně maximálně věrohodný, spočívá zejména v numerických obtížích při vyčíslování výrazů, které obsahuje věrohodnostní funkce. Pripomeňme, že při analýze obrazu by mohla být např. řádově  $|V| = 10^3 \times 10^3$  a matice  $R_{VV}^U$  by pak byla řádově  $10^6 \times 10^6$ . Počítat

determinanty a inverze od takových matic by jistě bylo velmi obtížné, pokud vůbec možné.

Ani přibližný věrohodnostní výraz  $\hat{L}(\theta)$ , který vede na soustavu (S<sub>I</sub>), však není z numerického hlediska ideální. Zkusíme jej v následujícím odstavci nahradit funkcí, která je numericky jednodušší a přitom do určité míry imituje vlastnosti věrohodnosti.

## 6. MAXIMÁLNĚ PSEUDO – VĚROHODNÝ ODHAD

Obecně princip maximální pseudo-věrohodnosti spočívá v nahrazení věrohodnosti, tedy např.  $p_V(x_V)$ , výrazem

$$\prod_{V_1 \in \mathcal{V}} p_{V_1|V_2}(x_{V_1}|x_{V_2})$$

kde  $\mathcal{V}$  je systém nějakých relativně malých podmnožin. V případě markovských náhodných polí můžeme vzít  $\mathcal{V} = \{V_0 + t, t \in W\}$ , tedy maximálně pseudo-věrohodný odhad musí maximalizovat

$$\int \log p_{V_0|V_0^c}^\theta d\hat{P} = -\hat{L}_{V_0}(\theta).$$

I tento odhad můžeme interpretovat v rámci odhadů s minimální vzdáleností, neboť

$$\hat{L}_{V_0}(\theta) = I_{V_0}(\hat{P}|P^\theta) + k(\hat{P}),$$

kde

$$I_{V_0}(P^1|P^2) = \int \log \frac{dP_{V_0|V_0^c}^1}{dP_{V_0|V_0^c}^2} dP^1$$

je podmíněná  $I$ -divergence. Zde  $k(\hat{P})$  je už pouze konstanta závisající na  $\hat{P}$  a nikoli na neznámém parametru.

Podobně jako pro náhodná pole s diskrétní množinou stavů (Janžura (1997)) bychom ukázali, že tento odhad je asymptoticky normální, ale obecně nikoli eficientní, přičemž s rostoucí  $V_0$  se k eficienci blíží.

Ukážeme nyní, jak vypadá konkrétní realizace tohoto odhadu v nejjednodušším případě  $V_0 = \{0\}$ . Potom totiž

$$\begin{aligned} & p_{\{0\}|\{0\}^c}^\theta(x_0|x_{\{0\}^c}) = \\ & = \left(\frac{U_0}{\pi}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{4}U_0\left(\sum_{k \in M} U_k \frac{x_k - x_{-k}}{2} + h\right)^2\right\} \in \mathcal{L}_M \cup (-M). \end{aligned}$$

Pro pozorovací oblast  $W$  označme  $\widetilde{W} = W^{M \cup (-M)}$ . Můžeme tedy napsat

$$\begin{aligned} \hat{L}_{\{0\}}(\theta) & = -\frac{1}{2} \log \frac{U_0}{\pi} + \frac{1}{4} U_0 \left[ |\widetilde{W}|^{-1} x_T^\top A_{T\widetilde{W}}^U A_{\widetilde{W}T}^U x_T + \right. \\ & \quad \left. + |\widetilde{W}|^{-1} 2h 1_{\widetilde{W}}^\top A_{\widetilde{W}T}^U x_T + h^2 \right]. \end{aligned}$$

Minimalizovat tuto funkci znamená řešit soustavu

$$\begin{aligned} h &= - \sum_{k \in M} U_k \hat{\mu}_k \\ 2 \sum_{k \in M} \hat{C}_{jk} U_k &= \delta(j) \quad \text{pro } j \in M, \end{aligned}$$

kde

$$\begin{aligned} \hat{C}_{jk} &= |\widetilde{W}|^{-1} x_T^\top I_{T\widetilde{W}}^j I_{\widetilde{W}T}^k x_T - \hat{\mu}^j \hat{\mu}^k = \frac{R(\widehat{j-k}) + R(\widehat{j+k})}{2} \\ \hat{\mu}^j &= |\widetilde{W}|^{-1} 1_{\widetilde{W}}^\top I_{\widetilde{W}T}^j x_T \quad \text{pro } j, k \in M. \end{aligned}$$

Vidíme, že princip maximální pseudo-věrohodnosti vede k odhadu

$$\hat{\theta}^{\text{II}} = (\hat{h}^{\text{II}}, \hat{U}^{\text{II}})$$

metodou „Yule–Walkerových“ rovnic. Tento odhad je tedy pro gaussovská náhodná pole méně eficientní než  $\hat{\theta}^{\text{I}}$ , což není překvapivé, neboť vyžaduje více hodnot empirických kovariancí, jejich výpočet je nutně méně efektivní.

## 7. LITERATURA

- [1] Janžura M. (1988 a), *Asymptotic theory of parameter estimation for Gauss–Markov random fields*. Kybernetika 24, **3**, 161–176.
- [2] Janžura M. (1988 b), *Divergence of Gauss–Markov random fields with application to statistical inference*. Kybernetika 24, **6**, 401–412.
- [3] Janžura M. (1997), *Asymptotic results in parameter estimation for Gibbs random fields*. Kybernetika 33 (1997), 2, 133–159.