

Ornstein-Uhlenbeck mező előrejelzésére vonatkozó optimális mintavételi terv meghatározása

Sikolya Kinga

DE IK, Alkalmazott Matematika és Valószínűségszámítás Tanszék

Témavezető: Dr. Baran Sándor

Statisztikai modell

Stacionárius folyamat:

$$Y(s, t) = \theta + \varepsilon(s, t), \quad (s, t) \in \mathcal{X}.$$

$\mathcal{X} = \{[a_1, b_1] \times [a_2, b_2], a_1 < b_1, a_2 < b_2\}$: mintavételi tér.

$\varepsilon(s, t)$, $s, t \in \mathbb{R}$: stacionárius Ornstein-Uhlenbeck (OU) mező, azaz az alábbi kovariancia struktúrával rendelkező nulla várható értékű Gauss folyamat

$$E \varepsilon(s_1, t_1) \varepsilon(s_2, t_2) = \frac{\tilde{\sigma}^2}{4\alpha\beta} \exp(-\alpha|t_1 - t_2| - \beta|s_1 - s_2|),$$

ahol $\alpha > 0$, $\beta > 0$, $\tilde{\sigma} > 0$.

Átparaméterezve:

$$E \varepsilon(s_1, t_1) \varepsilon(s_2, t_2) = \sigma^2 \exp(-\alpha|t_1 - t_2| - \beta|s_1 - s_2|),$$

$\sigma := \tilde{\sigma}/(2\sqrt{\alpha\beta})$: segédparaméter.

Minta pontok

1) Monoton halmaz:

H : Hilbert tér valós vagy komplex skalárokkal.

$\langle x, y \rangle$, ahol $x, y \in H$: a belső szorzat valós része.

Egy $E \subset H \times H$ halmazt monotonnak nevezünk, ha minden $(x_1, y_1), (x_2, y_2) \in E$ esetén fennáll $\langle x_1 - x_2, y_1 - y_2 \rangle \geq 0$. [Minty, 1962, Minty, 1963]

Monoton halmazon vett megfigyelési pontokra vonatkozó feltétel:

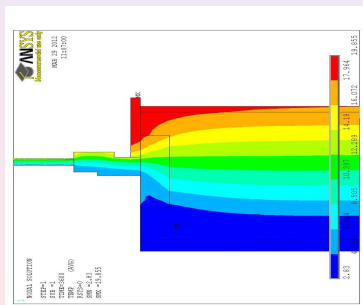
D feltétel. Az $\{(s_1, t_1), (s_2, t_2), \dots, (s_n, t_n)\} \subset \mathcal{X}$, $n \in \mathbb{N}$, megfigyelési helyek között nincs átfedés (azaz a megfigyelések között nincs ismétlődés), továbbá $0 < s_1 < s_2 < \dots < s_n$ és $0 < t_1 < t_2 < \dots < t_n$ teljesül.

2) Szabályos rács: Az $\{(s_1, t_1), (s_1, t_2), \dots, (s_n, t_m)\} \subset \mathcal{X}$, $n, m \in \mathbb{N}$ megfigyelési helyek egy téglalap rácspontjai ($n \times m$ darab).

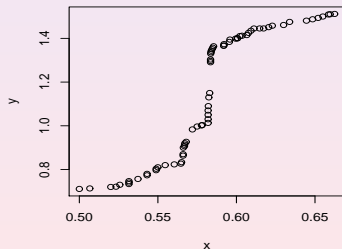
Monoton halmaz alkalmazása: stacionárius hőmérsékleti mező mérése

Cél: Egy épületben keletkező hőhíd, valamint penészedésre hajlamos helyek csökkentése.

Számítógépes szimuláció: valós hőmérsékleti eloszlás modellezése. Az adatok Minárová (2005) cikkéből származnak.



(a) Hőmérsékleti mező izotermái;



(b) Izotermák megfigyelési pontjai.

A trendparaméter becslése monoton halmaz esetén

Modell:

$$Y(s, t) = \theta + \varepsilon(s, t),$$

ahol az ε OU mező kovariancia struktúrájának α, β és σ paraméterei adottak.

Cél: A D feltételt kielégítő $\{(s_1, t_1), (s_2, t_2), \dots, (s_n, t_n)\}$ megfigyelési pontokban tekintett modell esetén meghatározni a θ trendparaméter becslésére szolgáló optimális mintavételi elrendezést.

D-optimalitási kritérium: θ paraméterre vonatkozó Fisher-féle információ-mennyiség maximalizálása.

θ paraméterre vonatkozó Fisher-féle inf. [Pázman, 2007, Xia *et al.*, 2006]:

$$M_{\theta}(n) = \mathbf{1}_n^{\top} C^{-1}(n, r) \mathbf{1}_n = 1 + \sum_{i=1}^{n-1} \frac{1 - q_i}{1 + q_i} \quad \text{ahol,}$$

$q_i := \exp(-\alpha d_i - \beta \delta_i)$ és $d_i := s_{i+1} - s_i$, $\delta_i := t_{i+1} - t_i$, $i = 1, 2, \dots, n - 1$.

$\mathbf{1}_n$, $n \in \mathbb{N}$: n hosszúságú egyesekből álló oszlopvektor.

$C(n, r)$: megfigyelések kovariancia mátrixa, ahol $r = (\alpha, \beta)^{\top}$.

A kovariancia mátrix egy Toeplitz mátrix [Kiseľák and Stehlík, 2008]:

$$C(n, r) = \begin{bmatrix} 1 & q_1 & q_1 q_2 & q_1 q_2 q_3 & \dots & \dots & \prod_{i=1}^{n-1} q_i \\ q_1 & 1 & q_2 & q_2 q_3 & \dots & \dots & \prod_{i=2}^{n-1} q_i \\ q_1 q_2 & q_2 & 1 & q_3 & \dots & \dots & \prod_{i=3}^{n-1} q_i \\ q_1 q_2 q_3 & q_2 q_3 & q_3 & 1 & \dots & \dots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & \vdots & & \ddots & \vdots \\ \prod_{i=1}^{n-1} q_i & \prod_{i=2}^{n-1} q_i & \prod_{i=3}^{n-1} q_i & \dots & \dots & q_{n-1} & 1 \end{bmatrix}$$

Tétel. [Baran és Stehlík, 2013] *A D feltételt teljesítő monoton halmaz esetén a θ trendparaméter becslésére szolgáló optimális mintavételt a következő alakú ekvidisztáns minta adja:*

$$\alpha d_i + \beta \delta_i \text{ konstans, ahol } d_i := s_{i+1} - s_i, \delta_i := t_{i+1} - t_i, i = 1, 2, \dots, n-1.$$

A trendparaméter becslése szabályos rács esetén

Modell:

$$Y(s, t) = \theta + \varepsilon(s, t),$$

ahol az ε OU mező kovariancia struktúrájának α, β és σ paraméterei adottak.

Cél: A θ paraméterbecslés szerinti D-optimális mintavétel meghatározása az $\{(s_1, t_1), (s_2, t_2), \dots, (s_n, t_m)\}$ szabályos rácpontokban megfigyelt modell esetén.

θ paraméterre vonatkozó Fisher-féle információmennyiség:

$$M_{\theta}(n, m) = \mathbf{1}_{nm}^{\top} C^{-1}(n, m, r) \mathbf{1}_{nm} = M_{\theta}^{(1)}(n) M_{\theta}^{(2)}(m) = \left(1 + \sum_{i=1}^{n-1} \frac{1-p_i}{1+p_i}\right) \left(1 + \sum_{j=1}^{m-1} \frac{1-q_j}{1+q_j}\right).$$

$p_i := \exp(-\alpha d_i)$, $q_j := \exp(-\beta \delta_j)$ ahol $d_i := s_{i+1} - s_i$ és $\delta_j := t_{j+1} - t_j$,
 $i = 1, 2, \dots, n-1$ and $j = 1, 2, \dots, m-1$.

$\mathbf{1}_k$, $k \in \mathbb{N}$: k hosszúságú egyesekből álló oszlopvektor.

$C(n, m, r)$: megfigyelések kovariancia mátrixa, ahol $r = (\alpha, \beta)^{\top}$.

A kovariancia struktúra Kronecker-szorzatként áll elő:

$$C(n, m, r) = P(n, r) \otimes Q(m, r), \quad \text{ahol}$$

$$P(n, r) := \begin{bmatrix} 1 & p_1 & p_1 p_2 & \dots & \dots & \prod_{i=1}^{n-1} p_i \\ p_1 & 1 & p_2 & \dots & \dots & \prod_{i=2}^{n-1} p_i \\ p_1 p_2 & p_2 & 1 & \dots & \dots & \prod_{i=3}^{n-1} p_i \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & p_{n-1} \\ \prod_{i=1}^{n-1} p_i & \prod_{i=2}^{n-1} p_i & \dots & \dots & p_{n-1} & 1 \end{bmatrix}$$

és $Q(m, r)$ a $P(n, r)$ -hez analóg módon van definiálva.

Tétel. [Baran, Stehlík és S, 2013b] Szabályos rács esetén a θ paraméterbecslés szerinti D -optimális mintavételt a $d_1 = d_2 = \dots = d_{n-1}$ és $\delta_1 = \delta_2 = \dots = \delta_{m-1}$ iránymenti ekvidisztáns minta adja.

IMSPE kritérium szerinti optimális mintavétel monoton halmaz esetén

Tekintsük az $Y(s, t) = \theta + \varepsilon(s, t)$ folyamatot azon $\{(s_1, t_1), \dots, (s_n, t_n)\}$ pontokban, melyek a D feltételt teljesítik.

Cél: Az $x = (x_1, x_2) \in \mathcal{X}$ pontban tekintett folyamat "krigelés" szerinti optimális mintavételének meghatározása, azaz tekintsük a $Y(x)$ legjobb lineáris torzítatlan becslését:

$$\hat{Y}(x) = \hat{\theta} + Q^\top(x)C^{-1}(n, r)(\mathbf{Y} - \mathbf{1}_n\hat{\theta}).$$

$\mathbf{Y} = (Y(s_1, t_1), \dots, Y(s_n, t_n))^\top$: megfigyelések vektora,

$\mathbf{1}_n$: n hosszúságú egyesekből álló oszlopvektor,

$C(n, r)$: \mathbf{Y} kovariancia mátrixa,

$Q(x)$: $Y(x)$ és \mathbf{Y} közötti korrelációkból álló vektor, azaz

$$Q(x) = (\varrho(x, s_1, t_1), \dots, \varrho(x, s_n, t_n))^\top, \quad \varrho(x, s_i, t_i) := \exp(-\alpha|x_1 - s_i| - \beta|x_2 - t_i|),$$

$\hat{\theta}$: θ legkisebb négyzetes becslése, azaz

$$\hat{\theta} = (\mathbf{1}_n^\top C^{-1}(n, r)\mathbf{1}_n)^{-1} \mathbf{1}_n^\top C^{-1}(n, r)\mathbf{Y}.$$

IMSPE kritérium

$\hat{Y}(x)$ átlagos négyzetes hibája (MSPE):

$$\text{MSPE}(\hat{Y}(x)) := \sigma^2 \left[1 - (\mathbf{1}, \mathcal{Q}^\top(x)) \begin{bmatrix} 0 & \mathbf{1}_n^\top \\ \mathbf{1}_n & C(n, r) \end{bmatrix}^{-1} (\mathbf{1}, \mathcal{Q}^\top(x))^\top \right].$$

IMSPE kritérium szerinti optimális mintavételi tervet az

$$\text{IMSPE}(\hat{Y}) := \sigma^{-2} \int_{\mathcal{X}} \text{MSPE}(\hat{Y}(x)) dx$$

által adott optimum segítségével határozzuk meg.

Feltehető:

Mintavételi tér: $\mathcal{X} = [0, 1]^2$, $s_1 = t_1 = 0$, $s_n = t_n = 1$.

Monoton halmaz esetén az IMSPE egzakt alakja

Tétel. [Baran, Stehlík és S, 2013a] *Ornstein-Uhlenbeck mező és monoton halmazt alkotó minta pontok esetén*

$$\text{IMSPE}(\hat{Y}) = 1 - A_n + \left(1 + \sum_{i=1}^{n-1} \frac{1 - q_i}{1 + q_i}\right)^{-1} B_n, \quad \text{ahol}$$

$$A_n = R_{1,1} + \sum_{i=1}^{n-1} \frac{R_{i,i}q_i^2 - 2R_{i+1,i}q_i + R_{i+1,i+1}}{1 - q_i^2},$$

$$B_n = 1 - 2\omega_n + R_{n,n}$$

$$+ \sum_{i=1}^{n-1} \left(\frac{R_{i,i} - 2R_{i+1,i}q_i + R_{i+1,i+1}q_i^2}{(1 + q_i)^2} + 2 \frac{(R_{n,i} - \omega_i) - (R_{n,i+1} - \omega_{i+1})q_i}{1 + q_i} \right)$$

$$+ 2 \sum_{i=2}^{n-1} \sum_{j=1}^{i-1} \frac{R_{i,j} - R_{i+1,j}q_i - R_{i,j+1}q_j + R_{i+1,j+1}q_iq_j}{(1 + q_i)(1 + q_j)},$$

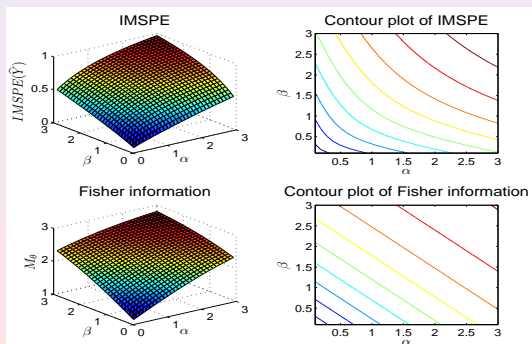
és

$$\omega_i = \frac{1}{\alpha\beta} \left[2 - e^{-\alpha(d_1 + \dots + d_{i-1})} - e^{-\alpha(d_i + \dots + d_{n-1})} \right] \left[2 - e^{-\beta(\delta_1 + \dots + \delta_{i-1})} - e^{-\beta(\delta_i + \dots + \delta_{n-1})} \right],$$

$$R_{i,j} = \left[\frac{1}{2\alpha} \left(2e^{-\alpha(d_j + \dots + d_{i-1})} - e^{-\alpha(2d_1 + \dots + 2d_{j-1} + d_j + \dots + d_{i-1})} - e^{-\alpha(d_j + \dots + d_{i-1} + 2d_i + \dots + 2d_{n-1})} \right) \right. \\ \left. + (d_j + \dots + d_{i-1})e^{-\alpha(d_j + \dots + d_{i-1})} \right] \\ \times \left[\frac{1}{2\beta} \left(2e^{-\beta(\delta_j + \dots + \delta_{i-1})} - e^{-\beta(2\delta_1 + \dots + 2\delta_{j-1} + \delta_j + \dots + \delta_{i-1})} - e^{-\beta(\delta_j + \dots + \delta_{i-1} + 2\delta_i + \dots + 2\delta_{n-1})} \right) \right. \\ \left. + (\delta_j + \dots + \delta_{i-1})e^{-\beta(\delta_j + \dots + \delta_{i-1})} \right], \quad i, j \in \mathbb{N}, \quad 1 \leq j \leq i \leq n.$$

Példa

- Hárompontos mintavétel: $n = 3$, $s_1 = t_1 = 0$, $s_2 := d$, $t_2 := \delta$, $s_3 = t_3 = 1$ ($d_1 = d$, $d_2 = 1 - d$, $\delta_1 = \delta$, $\delta_2 = 1 - \delta$).
- $d = 1/2$, $\delta = 1/2$ esetén lesz az IMSPE minimális (azaz az ekvidisztáns minta adja az optimumot).



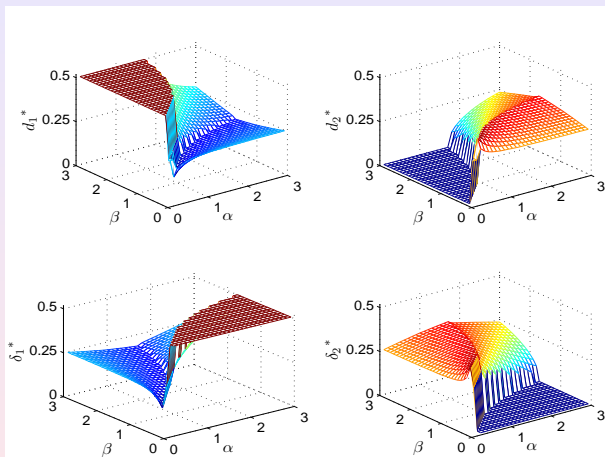
IMSPE tulajdonságai

Tétel. [Baran, Stehlík és S, 2013a] *IMSPE szimmetrikus abban az értelemben, hogy (d_i, δ_i) és (d_{n-i}, δ_{n-i}) , $i = 1, 2, \dots, n-1$, távolság párok felcserélésével az $\text{IMSPE}(\hat{Y})$ értéke változatlan marad.*

Következmény. *Az IMSPE kritérium szerinti optimális minta szimmetrikus, azaz ha $\text{IMSPE}(\hat{Y})$ -nek az optima $(d_1^*, d_2^*, \dots, d_{n-1}^*, \delta_1^*, \delta_2^*, \dots, \delta_{n-1}^*)$ -ben van, akkor $d_i^* = d_{n-i}^*$ és $\delta_i^* = \delta_{n-i}^*$, $i = 1, 2, \dots, n$.*

Példa:

- Tekintsünk egy öt pontos mintavételt: $n = 5$ és $s_1 = t_1 = 0$, $s_5 = t_5 = 1$.
- A következményből adódik, hogy ekkor az optimális minta esetén $d_1^* = d_4^*$, $d_2^* = d_3^*$ és $\delta_1^* = \delta_4^*$, $\delta_2^* = \delta_3^*$ teljesül.



$n = 5$ esetén az (α, β) korrelációs paraméterek függvényében az IMSPE optimumát adó d_1^* , d_2^* és δ_1^* , δ_2^* értékek.

IMSPE kritérium szerinti optimális mintavétel szabályos rács esetén

Tekintsük az $Y(s, t) = \theta + \varepsilon(s, t)$ folyamatot az $\{(s_1, t_1), \dots, (s_n, t_m)\}$ minta pontokban.

Az $Y(x_1, x_2)$, $(x_1, x_2) \in \mathcal{X}$, legjobb torzítatlan lineáris becslése:

$$\hat{Y}(x_1, x_2) = \hat{\theta} + R^\top(x_1, x_2)C^{-1}(n, m, r)(\mathbf{Y} - \mathbf{1}_{nm}\hat{\theta}), \quad \text{ahol}$$

$\mathbf{Y} = (Y(s_1, t_1), \dots, Y(s_n, t_m))^\top$: megfigyelések vektora,

$C(n, m, r)$: \mathbf{Y} kovariancia mátrixa,

$R(x_1, x_2)$: $Y(x_1, x_2)$ és \mathbf{Y} vektor közötti korrelációs vektor, azaz

$$R(x_1, x_2) = (\varrho(x_1, x_2, s_1, t_1), \dots, \varrho(x_1, x_2, s_i, t_j), \dots, \varrho(x_1, x_2, s_n, t_m))^\top, \quad \text{ahol}$$

$\varrho(x_1, x_2, s_i, t_j) := \varrho_1(x_1, s_i)\varrho_2(x_2, t_j)$, ahol $\varrho_1(x_1, s_i) := \exp(-\alpha|x_1 - s_i|)$ és $\varrho_2(x_2, t_j) := \exp(-\beta|x_2 - t_j|)$

$\hat{\theta}$: θ legkisebb négyzetes becslése, azaz

$$\hat{\theta} = (\mathbf{1}_{nm}^\top C^{-1}(n, m, r)\mathbf{1}_{nm})^{-1}\mathbf{1}_{nm}^\top C^{-1}(n, m, r)\mathbf{Y}.$$

Az előrejelzés átlagos négyzetes hibája (MSPE):

$$\text{MSPE}(\hat{Y}(x_1, x_2)) := \sigma^2 \left[1 - (1, R^\top(x_1, x_2)) \begin{bmatrix} 0 & \vdots & \mathbf{1}_{nm}^\top \\ \vdots & \ddots & \vdots \\ \mathbf{1}_{nm} & \vdots & C(n, m, r) \end{bmatrix}^{-1} (1, R^\top(x_1, x_2))^\top \right]$$

$$\text{IMSPE}(\hat{Y}) := \sigma^{-2} \iint_{\mathcal{X}} \text{MSPE}(\hat{Y}(x_1, x_2)) dx_1 dx_2$$

Tétel. [Baran, Stehlík és S, 2013b] *Szabályos rácspontot alkotó minta pontok esetén*

$$\begin{aligned} \text{IMSPE}(\hat{Y}) = & 1 - \left(\frac{n-1}{\alpha} - 2 \sum_{i=1}^{n-1} \frac{d_i p_i^2}{1-p_i^2} \right) \left(\frac{m-1}{\beta} - 2 \sum_{j=1}^{m-1} \frac{\delta_j q_j^2}{1-q_j^2} \right) \\ & + \left(1 + \sum_{i=1}^{n-1} \frac{1-p_i}{1+p_i} \right)^{-1} \left(1 + \sum_{j=1}^{m-1} \frac{1-q_j}{1+q_j} \right)^{-1} \left[1 - \frac{8}{\alpha\beta} \left(\sum_{i=1}^{n-1} \frac{1-p_i}{1+p_i} \right) \left(\sum_{j=1}^{m-1} \frac{1-q_j}{1+q_j} \right) \right. \\ & \left. + \left(\sum_{i=1}^{n-1} \frac{1-p_i^2 + 2\alpha d_i p_i}{\alpha(1+p_i)^2} \right) \left(\sum_{j=1}^{m-1} \frac{1-q_j^2 + 2\beta \delta_j q_j}{\beta(1+q_j)^2} \right) \right], \quad \text{ahol} \end{aligned}$$

$$p_i := \exp(-\alpha d_i), \quad q_j := \exp(-\beta \delta_j) \quad \text{ahol } d_i := s_{i+1} - s_i \text{ és } \delta_j := t_{j+1} - t_j$$

Entrópia kritérium szerinti optimális mintavétel monoton halmaz esetén

Shewry és Wynn (1987): A megfigyelések $\text{Ent}(\mathbf{Y})$ entrópiájának maximalizálásával határozta meg az optimális mintát.

Gauss esetben: n -dimenziós normál vektorként felírt $\sigma^2 C(n, r)$ kovariancia mátrixú megfigyelések:

$$\text{Ent}(\mathbf{Y}) = \frac{n}{2} (1 + \ln(2\pi\sigma^2)) + \frac{1}{2} \ln \det C(n, r).$$

Tétel. [Baran, Stehlík és S, 2013a] *Monoton halmazt alkotó minta esetén az $\text{Ent}(\mathbf{Y})$ entrópia a következő alakban áll elő*

$$\text{Ent}(\mathbf{Y}) = \frac{n}{2} (1 + \ln(2\pi\sigma^2)) + \frac{1}{2} \sum_{i=1}^{n-1} \ln(1 - q_i^2).$$

Ebben az esetben az entrópia kritérium szerinti optimális mintavételt az az ekvidisztáns minta adja, ahol $\alpha d_i + \beta \delta_i$, $i = 1, 2, \dots, n-1$, (azaz a q_i) konstans.

Megjegyzés: A $d_1 = d_2 = \dots = d_n$ és $\delta_1 = \delta_2 = \dots = \delta_n$ egyenlő távolságú monoton minta az entrópia kritérium szerint optimális.

Entrópia kritérium szerinti optimális mintavétel szabályos rács esetén

Gauss eset: a megfigyelések egy $n \cdot m$ -dimenziós $\sigma^2 C(n, m, r)$ kovariancia mátrixú normális véletlen vektort alkotnak:

$$\text{Ent}(\mathbf{Y}) = \frac{nm}{2} (1 + \ln(2\pi\sigma^2)) + \frac{1}{2} \ln \det C(n, m, r).$$

Tétel. [Baran, Stehlík és S, 2013b] *Szabályos rácsot alkotó minta esetén az $\text{Ent}(\mathbf{Y})$ entrópia a következő alakban áll elő*

$$\text{Ent}(\mathbf{Y}) = \frac{nm}{2} (1 + \ln(2\pi\sigma^2)) + \frac{m}{2} \sum_{i=1}^{n-1} \ln(1 - p_i^2) + \frac{n}{2} \sum_{j=1}^{m-1} \ln(1 - q_j^2).$$

Ebben az esetben az entrópia kritérium szerinti optimális mintavételt a $d_1 = d_2 = \dots = d_{n-1}$ és $\delta_1 = \delta_2 = \dots = \delta_{m-1}$ iránymenti ekvidisztáns minta adja.

IMSPE és az entrópia kritérium összehasonlítása

		$\alpha = 0.5, \beta = 0.8$	$\alpha = 1, \beta = 1$	$\alpha = 1, \beta = 10$
$n = 4$	optimális	0.2602	0.4008	0.9266
	egy. táv.	0.2693	0.4010	0.9326
	rel. hat. (%)	96.62	99.95	99.36
$n = 5$	optimális	0.2309	0.3699	0.8290
	egy. táv.	0.2473	0.3700	0.8409
	rel. hat. (%)	93.37	99.97	98.58
$n = 7$	optimális	0.2007	0.3423	0.7066
	egy. táv.	0.2274	0.3424	0.7288
	rel. hat. (%)	88.26	99.97	96.95
$n = 10$	optimális	0.1570	0.3262	0.6057
	egy. táv.	0.2155	0.3262	0.6390
	rel. hat. (%)	72.87	99.98	94.79

Az IMSPE értékei optimális, valamint egyenlő távolságú monoton minta esetén.

IMSPE monoton halmaz és szabályos rács esetén

		$\alpha = 0.5, \beta = 0.8$	$\alpha = 1, \beta = 1$	$\alpha = 3, \beta = 3$
$n = 4$	monoton	0.2602	0.4008	0.8487
	szab. rács	0.3580	0.5389	1.0094
	rel. hat. (%)	137.59	134.46	118.93
$n = 9$	monoton	0.1620	0.3300	0.6997
	szab. rács	0.1527	0.3018	0.7011
	rel. hat. (%)	94.26	91.45	100.20

Az IMSPE optimális értékei monoton halmazt, valamint szabályos rácsot alkotó minta esetén.

- $n = 4$: a monoton minta esetén az IMSPE optimális értékei jobbak, mint szabályos rács esetén,
- $n = 9$ és $\alpha = \beta = 3$: a relatív hatékonyság közel 100%.

θ paraméterre vonatkozó Fisher-féle információmennyiség és az entrópia kritérium monoton halmaz és szabályos rács esetén

- A metán szerkezete: 62 pontból áll [Vaghjiani és Ravishankara, 1991]
- $n = 64$ pontos mintát tekintünk
- 8×8 szabályos rács összehasonlítása az $n = 64$ pontból álló monoton halmazzal
- Mintavételi tér: $[223, 420] \times [0.84, 43.51]$

		$\alpha = 0.001, \beta = 0.01$	$\alpha = 0.1, \beta = 1$	$\alpha = 1, \beta = 10$
D-opt.	mon.	1.3118	29.8651	63.9937
	szab.rács	1.3328	57.4388	64.00
	rel.hat.(%)	98.43	51.99	99.99
entrópia	mon.	-33.0446	86.1318	90.8121
	szab.rács	-51.1507	90.7111	90.8121
	rel.hat.(%)	64.60	94.95	100

Hivatkozások



Baran, S., Stehlík, M., 2013.

Optimal designs for parameters of shifted Ornstein-Uhlenbeck sheets measured on monotonic sets. *Appl. Stoch. Model. Bus.*, submitted.



Baran, S., Sikolya, K., Stehlík, M., 2013a.

On the optimal designs for prediction of Ornstein-Uhlenbeck sheets. *Statist. Probab. Lett.* **83**, 1580–1587.



Baran, S., Sikolya, K., Stehlík, M., 2013b.

Optimal designs for the methane flux in troposphere., manuscript.



Kiseľák, J., Stehlík, M., 2008.

Equidistant D-optimal designs for parameters of Ornstein-Uhlenbeck process. *Statist. Probab. Lett.* **78**, 1388–1396.



Minárová, M., 2005.

Deformované teplotné polia a riziko vzniku hygienických problémov. Edícia vedeckých prác Slovenská technická univerzita, Vydavateľstvo STU, Bratislava (in Slovak).



Minty, G.J. 1962.

Monotone (non-linear) operators in Hilbert space. *Duke Math. J.* **29**, 341–346.



Minty, G.J. 1963.

On a monotonicity method for the solution of nonlinear equations in Banach spaces. *Proc. Natl. Acad. Sci. U.S.* **50**, 1038–1041.



Pázman, A., 2007.

Criteria for optimal design for small-sample experiments with correlated observations. *Kybernetika* **43**, 453–462.



Shewry, M.C., Wynn, H.P., 1987.

Maximum entropy sampling. *J. Appl. Stat.* **14**, 165–170.



Vaghjiani G. L. and Ravishankara A. R., 1991.

New measurement of the rate coefficient for the reaction of OH with methane. *Nature* **350**, 406–409.



Xia, G., Miranda, M.L. and Gelfand, A.E., 2006.

Approximately optimal spatial design approaches for environmental health data. *Environmetrics* **17**, 363–385.

Alkalmazott statisztika: A Down szindróma kockázatának elemzése

- ▶ a szakirodalomban ismert eljárás, az ún. diszkriminancia analízis alkalmazása Magyarország Észak-Alföldi régiójának adataira (89000 egészséges és 333 veszélyeztetett terhesség);
- ▶ a kutatásban az újdonság, hogy az α -fetoprotein (AFP) és human chorionic gonadotrophin (hCG) mellett harmadik markerként, a graviditás specifikus β -glicoproteint (SP1) használja;
- ▶ a becslés az egészséges illetve a veszélyeztetett terhességekhez tartozó markerek együttes sűrűségfüggvényeinek likelihood hányadosán alapul, mely feltételezi a markerek együttes lognormalitását;
- ▶ kiemelt figyelmet kapott ezen feltételek teljesülésének vizsgálata;
- ▶ összehasonlításként a logisztikus regressziós osztályozás eredményeit is megvizsgáltuk;
- ▶ **Publikáció:** Baran, S., Sikolya, K., Veress, L., Estimating the risk of a Down's syndrome term pregnancy using age and serum markers: Comparison of various methods. *Comm. Statist. Simulation Comput.* **42** (2013), no. 7, 1654–1672.

Gauss folyamatokkal meghajtott lineáris modellek paraméterbecslései

- ▶ Modell: $Z(s, t) := m_1 g_1(s, t) + \dots + m_p g_p(s, t) + U(s, t)$;
- ▶ $U(s, t)$: Wiener vagy Ornstein-Uhlenbeck mező (stacionárius és nullából kiinduló);
- ▶ $g_1, \dots, g_p : \mathbb{R}_+^2 \rightarrow \mathbb{R}$: adott függvények;
- ▶ $m_1, \dots, m_p \in \mathbb{R}$: ismeretlen regressziós paraméterek;
- ▶ meghatároztuk az m_1, \dots, m_p paraméterek maximum likelihood becsléseit egy G megfigyelési tartomány felett;
- ▶ a kapott elméleti eredményeket szimulációkkal is alátámasztottuk;
- ▶ **Publikációk:**
 1. Baran, S., Sikolya, K., Parameter estimation in linear regression driven by a Gaussian sheet. *Acta Sci. Math. (Szeged)* **78** (2012), no. 3-4, 689–713.
 2. Baran, S., Sikolya, K., Parameter estimation in linear regression driven by a Wiener sheet. *Ann. Math. Inform.* **39** (2012), 3–15.

Térbeli autoregresszív modellek stabilitásának vizsgálata

- ▶ Modell:

$$X_{k,\ell} = \begin{cases} \alpha X_{k-1,\ell} + \beta X_{k,\ell-1} + \varepsilon_{k,\ell}, & \text{ha } k + \ell \geq 1, \\ 0, & \text{ha } k + \ell = 0, \end{cases} \quad k, \ell \in \mathbb{Z}, k + \ell \geq 0;$$

- ▶ Stabil modell: $|\alpha| + |\beta| < 1$; instabil: $|\alpha| + |\beta| = 1$; [Basu és Reinsel, 1993]
- ▶ (α, β) paraméterek $\{X_{k,\ell} : (k, \ell) \in T_n\}$ megfigyeléseken alapuló legkisebb négyzetes becslései stabil és az instabil esetben is aszimptotikusan normálisak; [Baran, Pap és Zuijlen, 2007]
- ▶ sikerült a $\varrho := |\alpha| + |\beta|$ stabilitási paraméter bevezetésével a fent említett eredményre egy jóval egyszerűbb bizonyítást adni;
- ▶ meghatároztuk a stabilitási paraméter legkisebb négyzetes becslésének aszimptotikáját;
- ▶ **Publikáció:** Baran, S., Pap, G., Sikolya, K., Testing stability in a spatial unilateral autoregressive model. *Statistics*, submitted (arXiv:1203.4346).

Konferenciákon tartott előadások, bemutatott poszterek

1. K. Sikolya: Tudományos Diákköri Dolgozat, A Down szindróma kockázatának elemzése az anya életkora és különböző biokémiai markerek szintje alapján, XXIX. Országos Tudományos Diákköri Konferencia, 2008. április 08-10., Debrecen, I. helyezés
2. K. Sikolya, L. Veress, S. Baran: Estimating the risk of Down's syndrome term pregnancy using age and serum markers. 14th Young Statisticians Meeting, Basovizza/Tieste, Olaszország, Október 16-18, 2009. (előadás)
3. K. Sikolya, L. Veress, S. Baran: Estimating the risk of a Down's syndrome term pregnancy: comparison of various methods. 8th International Conference on Applied Informatics, Eger, Magyarország, Január, 27-30, 2010. (poszter)
4. K. Sikolya, S. Baran: Parameter estimation in linear regression driven by a Wiener sheet. CSMA 2011, Debrecen, Magyarország, Augusztus 22-24, 2011. (poszter)
5. K. Sikolya, L. Veress, S. Baran: Estimating the risk of a Down's syndrome term pregnancy using age and serum markers. CEN 2011, Zurich, Svájc, Szeptember 12-16, 2011. (előadás)
6. K. Sikolya, L. Veress, S. Baran: Estimating the risk of a Down's syndrome term pregnancy using age and serum markers. ITSM 2011, Debrecen, Hungary, Október 27, 2011. (előadás)
7. K. Sikolya, S. Baran: Parameter estimation in linear regression driven by a Gaussian sheet. GPSD 2012, Mainz, Németország, Március 5-9, 2012. (előadás)
8. K. Sikolya, S. Baran: Parameter estimation in a spatial linear regression model. CSM 2012, Szeged, Magyarország, Június 28-30, 2012. (előadás)
9. K. Sikolya, S. Baran, Gy. Pap: Testing stability in a spatial unilateral autoregressive model. 8th World Congress in Probability and Statistics, Isztambul, Törökország, Július 9-14, 2012. (előadás)