Dinamikus rendszerek paramétereinek becslése

Valószínűségszámítás, matematikai statisztika (ismétlés)

Predikciós hiba minimalizálásán alapuló paraméterbecslés

Hangos Katalin

MTA SzTAKI, Folyamatirányítási Kutató Csoport

PE Számítástudomány Alkalmazása Tanszék

VALÓSZÍNűSÉGSZÁMÍTÁS (ismétlés)

Skalár értékű valószínűségi változók

A ξ valószínűségi változó *normális vagy Gauss eloszlású*, jelölésben

$$\xi \quad \sim \quad \mathbb{N}(m, \sigma^2) \tag{1}$$

ha a változó f_ξ valószínűségi sűrűségfüggvénye:

$$f_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} \tag{2}$$

ahol m a változó *várható értéke* és σ^2 a *szórása*

Az f_ξ sűrűségfüggvényű ξ valószínűségi változó *várható értéke* és *szórásnégyzete*

$$E\{\xi\} = \int x f_\xi(x) dx \;\; , \;\; \sigma^2\{\xi\} = \int (x - E\{\xi\})^2 f_\xi(x) dx$$

Kovariancia

Két skalár értékű valószínűségi változó ξ és θ *kovarianciája*:

$$COV\{\xi,\theta\} = E\{(\xi - E\{\xi\})(\theta - E\{\theta\})\}$$

A ξ skalár értékű valószínűségi változó *varianciája* (a változó önmagával vett kovarianciája):

$$\sigma^2\{\xi\} = COV\{\xi,\xi\} = E\{(\xi - E\{\xi\})^2\}$$

Korrelációs együttható (normált kovariancia):

$$\rho\{\xi,\theta\} = \frac{E\{(\xi - E\{\xi\})(\theta - E\{\theta\})\}}{\sigma\{\xi\}\sigma\{\theta\}}$$

Vektor értékű valószinűségi változók

A ξ valószínűségi változó vektor értékű, ha

$$\xi \ : \ \xi(\omega), \quad \omega \in \Omega, \quad \xi(\omega) \in \mathbb{R}^{\mu}$$

A változó $m \in R^{\mu}$ *várható értéke egy valós vektor*.

Varianciája egy valós elemű négyzetes mátrix, az ún. *kovariancia mátrix*:

$$COV\{\xi\} = E\{(\xi - E\{\xi\})(\xi - E\{\xi\})^T\}$$

A *kovariancia mátrixok pozitív definit szimmetrikus mátrixok*, azaz

$$z^T COV\{\xi\} z \geq 0 \quad , \quad \forall z \in \mathbb{R}^{\mu}$$

Együttes valószínűségi sűrűségfüggvény

A $\xi_1, ..., \xi_n$ skalár értékű valószínűségi változók együttes valószínűségi sűrűségfüggvénye egy n változós nem-negatív függvény $f(x_1, ..., x_n)$, amelyre

$$P[a_1 \le x_1 \le b_1,\ ...,\ a_n \le x_n \le b_n] = \int_{a_1}^{b_1} ... \int_{a_n}^{b_n} f(x_1, ..., x_n) dx_1 ... dx_n$$

Marginális valószínűségi sűrűségfüggvény (a 2 változós esetben)

$$f_{\xi_1}(x_1) = \int_{-\infty}^{\infty} f_{\xi_1,\xi_2}(x_1, x_2) dx_2$$

Függetlenség (a 2 változós esetben)

$$f_{\xi_1,\xi_2}(x_1, x_2) = f_{\xi_1}(x_1) \cdot f_{\xi_2}(x_2)$$

Többdimenziós normális eloszlás

A normális vagy Gauss eloszlású m várható értékű és Σ kovariancia mátrixú valószínűségi változó

$$\xi \quad \sim \quad N(m, \Sigma)$$

egy olyan valószínűségi változókból álló vektor, amelynek minden $\xi_i, i = 1, ..., \mu$ komponense egy normális eloszlású skalár értékű valószínűségi változó.

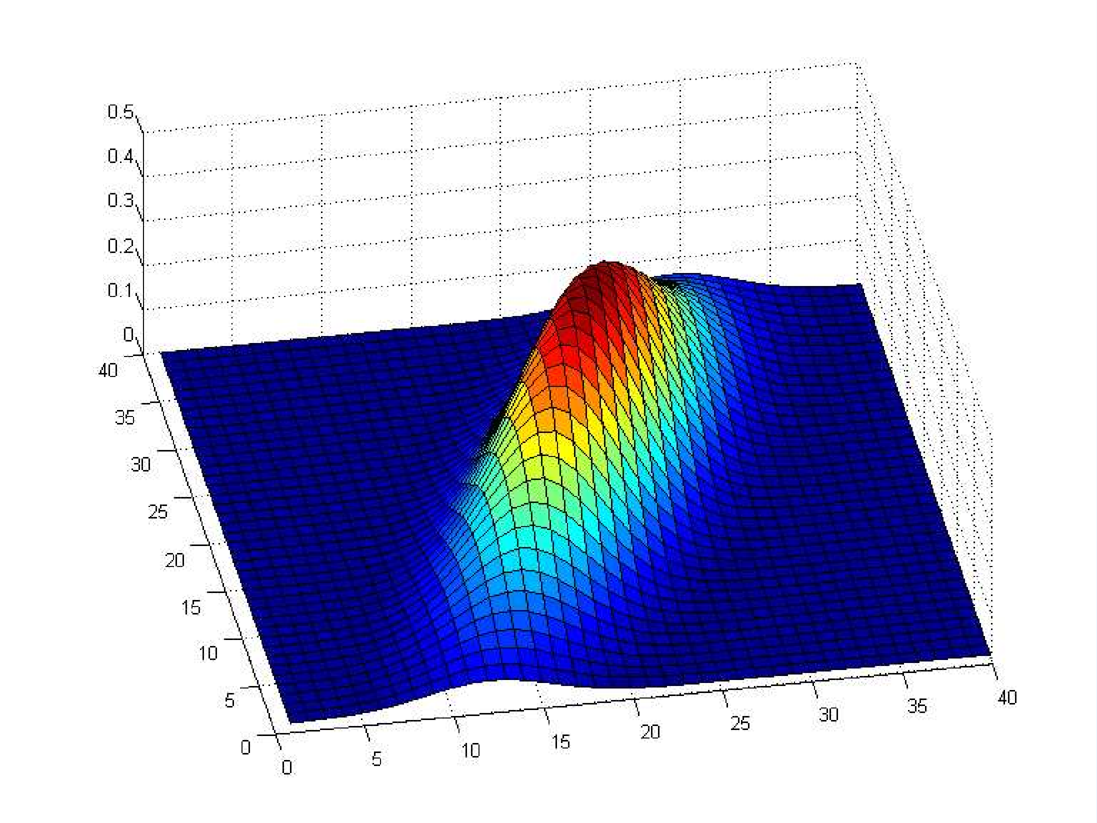
Valószínűségi sűrűségfüggvény: R a ρ_{ij} korrelációs e.h.-kból alkotott determináns

$$f(x_1, ..., x_\mu) = \frac{1}{\sqrt{2\pi}\sigma_1 ... \sigma_\mu \sqrt{R}} e^{-\frac{1}{2R}\left(\sum_{i=1}^{\mu} \sum_{j=1}^{\mu} \rho_{ij} \frac{(x_i - m_i)(x_j - m_j)}{\sigma_1 \sigma_2}\right)}$$

Két dimenziós normális eloszlás

Valószínűségi sűrűségfüggvény:

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} e^{-\frac{1}{2(1-r^2)}\left(\frac{(x_1-m_1)^2}{\sigma_1^2} - 2r\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right)}$$



Lineárisan transzformált valószínűségi változók

Egy $\xi(\omega) \in R^n$ vektor értékű valószínűségi változót a $T \in R^{n \times n}$ nemszinguláris négyzetes transzformációs mátrix alkalmazásával transzformáljuk:

$$\eta = T\xi$$

Az η valószínűségi változó statisztikái:

$$E\{\eta\} = TE\{\xi\} \quad , \quad COV\{\eta\} = TCOV\{\xi\}T^T$$

Ha a ξ valószínűségi változó $N(m_\xi, \Delta_\xi)$ normális eloszlású volt, m_ξ várható értékkel és Δ_ξ kovariancia mátrixszal, akkor η transzformált valószínűségi változó is $N(m_\eta, \Delta_\eta)$ normális eloszlású lesz, ahol

$$m_\eta = Tm_\xi \quad , \quad \Delta_\eta = T\Delta_\xi T^T$$

LINEÁRIS REGRESSZIÓ (ismétlés)

Paraméterbecslés lineáris regresszióval 1.

Feladatkitűzés

Adott:

- Egy paraméterekben lineáris determinisztikus modell

$$y^{(M)} = x^T p = \sum_{i=1}^{n} x_i p_i$$

  $p \in R^n$ az ismeretlen modell paraméterek, a determinisztikus független változók $x \in R^n$ és az $y^{(M)} \in R^\mu$ mérési hibával terhelt függő változó.

- Az m ($m \geq \mu$) mérésből álló mért értékek halmaza

$$y = \begin{bmatrix} y_1 \\ y_2 \\ .. \\ y_m \end{bmatrix} \quad , \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$$

Paraméterbecslés lineáris regresszióval 1f.

- Egy L veszteségfüggvény

$$
\begin{aligned}
L &= r^T W r = \sum_{i=1}^{m} \sum_{j=1}^{m} r_i W_{ij} r_j \\
r_i &= y_i - y_i^{(M)} = y_i - x^{(i)T} p
\end{aligned}
$$

ahol r a *reziduál vagy eltérés vektor* és W egy alkalmas *pozitív definit szimmetrikus súlymátrix*.

Feladat:
Számítsuk ki a p paraméterek egy $\hat{p}$ becslését úgy, hogy az L veszteségfüggvény minimális legyen.

L minimuma: $\frac{\partial L}{\partial p} = 0$

$$
\begin{aligned}
&(r^T W) r' = 0 \\
&(y - X^T p)^T W X^T = 0 \\
&X^T W y = X^T W X p
\end{aligned}
$$

Paraméterbecslés lineáris regresszióval 2.

Módszer: A p paraméterek legkisebb négyzetes (LS) becslése:

$$(X^T W X)\hat{p} = X^T W y \quad vagy \quad \hat{p} = (X^T W X)^{-1} X^T W y$$

Megjegyzések:

1. Feltételezzük, hogy mért függő változók értékét *additív, nulla várható értékű mérési hiba* terheli

$$y_i = y_i^{(M)} + \varepsilon_i \ \ (i = 1, \cdots, m) \ \ , \ \ E\{\varepsilon_i\} = 0$$

Ha még az egyes *mérési hibák egymástól statisztikailag függetlenek es azonos normális eloszlásúak Δ_ε kovariancia mátrixszal*, akkor a mért értékek eloszlása is normális lesz:

$$y \sim N(y^{(M)}, \Delta_\varepsilon)$$

2. A súlymátrix egy jó választása, ha az a mérési hibák ismert vagy becsült kovariancia mátrixának inverze:

$$W = \Delta_\varepsilon^{-1}$$

A regresszióval kapott paraméterek tulajdonságai

$$\hat{p} = (X^TWX)^{-1}X^TWy$$

becslési egyenletet tekinthetjük úgy is, mint az y valószínűségi változó egy lineáris transzformációját a $T = (X^TWX)^{-1}X^T$ konstans determinisztikus mátrixszal, amelynek eredménye a $\hat{p}$ valószínűségi változó.

Ennek alapján a $\hat{p}$ becslés várható értéke és $COV\{\hat{p}\} = TCOV\{y\}T^T$ szórása kiszámolható.

Ha a mérési hibák normális eloszlásúak nulla várható értékkel és Δ_ε kovariancia mátrixszal, és a súlymátrixot a Δ_ε^{-1}-nek választjuk, akkor a $\hat{p}$ becslés az alábbi tulajdonságokkal rendelkezik:

- *A becslés normális eloszlású.*
- *A becslés várható értéke $E\{\hat{p}\} = p$, tehát a becslés torzítatlan.*
- *A becslés kovariancia mátrixa*

$$COV\{\hat{p}\} = (X^TWX)^{-1}$$

- *A becslés egy maximum likelihood vagy legnagyobb valószínűségű becslés, és igy egy hatásos optimális becslés*

A PREDIKCIÓS HIBA MINIMALIZÁLÁSA

Prediktív input-output modellek: SISO eset

$$\hat{y}(k|\theta) = W_y(q^{-1},\theta)y(k) + W_u(q^{-1},\theta)u(k)$$

A fenti egyenletbeli $W_y(q^{-1},\theta)$ és $W_u(q^{-1},\theta)$ tényezők úgynevezett *lineáris szűrők*, a θ az állandó, ismeretlen, becsülni kívánt paraméterek vektora.

A diszkrét idejű lineáris időinvariáns egybemenetű-egykimenetű (DT-LTI SISO) sztochasztikus rendszerek általános alakú bemenet-kimenet modellje

$$A^*(q^{-1})y(k) = B^*(q^{-1})u(k) + C^*(q^{-1})e(k)$$

prediktív alakban:

$$\hat{y}(k|\theta) = y(k) - e(k) = (1 - H^{-1}(q^{-1},\theta))y(k) + (H^{-1}(q^{-1},\theta)G(q^{-1},\theta))u(k)$$

ahol

$$H^{-1}(q^{-1},\theta) = \frac{A^*(q^{-1})}{C^*(q^{-1})} \quad , \quad G(q^{-1},\theta) = \frac{B^*(q^{-1})}{A^*(q^{-1})}$$

ARX modellek prediktív alakja

Tekintsük a legegyszerűbb esetet, amikor az általános

$$A^*(q^{-1})y(k) = B^*(q^{-1})u(k) + C^*(q^{-1})e(k)$$

alakú input-output modellben a mozgóátlag tag nulla, azaz *a kimeneti rendszerzaj fehér*.
Ebben az esetben $C^*(q^{-1}) = 1$.
Ekkor $H^{-1}(q^{-1}, \theta) = A^*(q^{-1})$ és így

$$W_y(q^{-1}, \theta) = 1 - A^*(q^{-1}) \ \ , \ \ W_u(q^{-1}, \theta) = B^*(q^{-1})$$

A becslo elemei:

$$\theta = [-a_1 \ \ -a_2 \ \ \ldots \ \ -a_n \ \ b_0 \ b_1 \ \ \ldots \ \ b_m]^T \ \ , \ \ N > n + m$$

$$\hat{y}(k|\theta) = -a_1 y(k-1) - ... - a_n y(k-n) + b_0 u(k) + ... + b_m u(k-m)$$

Nemlineáris időinvariáns egykimenetű rendszerek

Mért érték sorozat

$$D[1,N] = D^N = \{(y(k), u(k)) \mid k = 1, \dots N\}$$

Az általános prediktív alak:

$$\hat{y}(k|\theta) = g(k, D[1, k-1]; \theta)$$

Paraméterekben lineáris rendszerek:

$$\hat{y}(k|\theta) = \theta^T g^*(k, D[1, k-1])$$

Példa: ARX modell

$$\hat{y}(k|\theta) = -a_1 y(k-1) - ... - a_n y(k-n) + b_0 u(k) + ... + b_m u(k-m)$$
$$\theta = [-a_1 \ \dots \ -a_n \ \ b_0 \ \dots \ b_m]^T$$
$$g^*(k, D[1,k-1]) = \varphi(k) = [y(k-1) \ \ \dots \ -y(k-n) \ \ u(k) \ \dots \ u(k-m)]$$

Példa: paraméterben lineáris nemlineáris eset

Paraméterekben lineáris nemlineáris ARX modell, amikor *a kimeneti rendszerzaj fehér*.

$$y(k) = a_1 y^2(k-1) + b_0 u^4(k) + e(k)$$

Az általános nemlineáris prediktív alak:

$$\theta = [a_1 \; b_0]^T \quad , \quad \hat{y}(k|\theta) = y(k) - e(k)$$

Segédváltozókkal:

$$y^2(k-1) = z(k-1) \quad , \quad u^4(k) = w(k)$$

a modell teljesen ARX alakra hozható:

$$\hat{y}(k|\theta) = a_1 z(k-1) + b_0 w(k)$$
$$\theta = [a_1 \; b_0]^T$$

A predikciós hiba

Előállítható a *predikciós hiba sorozat* a modellből és a mért érték sorozatból:

$$\varepsilon(k,\theta) = y(k) - \hat{y}(k|\theta) \quad , \quad k = 1, \ldots, N$$

A paraméterbecslés elve: Egy paraméterbecslési módszer a mért érték sorozatból előállít egy becsült paraméter vektort:

$$D^N \longrightarrow \hat{\theta}_N$$

Egy modell "jó", azaz a becsült paraméterek "jók", ha a predikciós hibák "kicsik".

A predikciós hiba nagysága

Az $\varepsilon(k,\theta)$ predikciós hiba egy jelsorozat, nagyságát *jelnormá*val mérjük.

A predikciós hiba nagysága

Előszűrés

A predikciós hiba előszűrésére akkor van szükség, ha a predikciós hibasorozat tagjai korreláltak, azaz nem tekinthetők fehérzajnak.

A norma megválasztása

$$V_N(\theta, D^N) = \frac{1}{N} \sum_{i=1}^{N} \ell(\varepsilon(k, \theta))$$

ahol $\ell(.)$ egy pozitív skalár értékű függvény.

Egykimenetű eset: legegyszerűbb és legelterjedtebb a négyzetes jelnorma

$$\ell(\varepsilon) = \frac{1}{2}\varepsilon^2$$

A predikciós hiba nagysága - 2

A norma megválasztása bonyolultabb esetekben

Súlyozott eset: nem egyenlően pontos mérések

$$V_N(\theta, D^N) = \frac{1}{N} \sum_{k=1}^{N} \beta(N,k)\ell(\varepsilon(k,\theta))$$

ahol $\beta(N,k)$ időtől függő súlytényező sorozat.

Többkimenetű eset: általánosított négyzetes jelnorma

$$\ell(\varepsilon) = \frac{1}{2}\varepsilon^T \Lambda^{-1} \varepsilon$$

A predikciós hiba minimalizálása

A paraméterbecslési módszer: $D^N \longrightarrow \hat{\theta}_N$

Az általános paraméterbecslési feladat:
Adottak

- mért értékek: $D[1,N] = D^N = \{(y(k), u(k)) \mid k = 1, \ldots N\}$
- prediktív parametrizált modell: $\hat{y}(k|\theta) = g(k, D[1,k-1]; \theta)$
  predikciós hiba sorozat (diszkrét idejű jel): $\varepsilon(k,\theta) = y(k) - \hat{y}(k|\theta)\ , \ k = 1, \ldots, N$
- norma a predikciós hibán: $V_N(\theta, D^N) = \frac{1}{N}\sum_{k=1}^{N} \ell(\varepsilon(k,\theta))$ ahol $\ell(.)$ egy pozitív skalár értékű függvény, leggyakrabban: $\ell(\varepsilon) = \frac{1}{2}\varepsilon^2$

Az ismert D^N mért érték sorozatból és a θ paraméter vektor értékéből kiszámíthatjuk a $V_N(\theta, D^N)$ norma értéket. A $k = N$ időpillanatban válasszuk meg a becsült $\hat{\theta}_N$ paraméter vektort úgy, hogy

$$\hat{\theta}_N = \hat{\theta}_N(D^N) = \arg\min_{\theta} V_N(\theta, D^N)$$

Példa: SISO ARX modellek

ALAPESET: Az általános alakú input-output modellben a mozgóátlag tag nulla, azaz *a kimeneti rendszerzaj fehér*

$$A^*(q^{-1})y(k) = B^*(q^{-1})u(k) + e(k)$$

A modell prediktív alakja:

$$\hat{y}(k|\theta) = -a_1 y(k-1) - a_2 y(k-2) \cdots - a_n y(k-n) + b_0 u(k) + \cdots + b_m u(k-m)$$

A paraméterverktor:

$$\theta = [-a_1 \ \ -a_2 \ \ \ldots \ \ -a_n \ \ b_0 \ b_1 \ \ \ldots \ \ b_m]^T$$

A predikciós hiba (**fehérzaj!**):

$$\varepsilon(k) = \hat{y}(k|\theta) - y(k) = e(k)$$