

1 Přílohy

1.1 Doplnění na čtverec

Používá se pro minimalizaci kvadratického výrazu nebo pro integraci v konvoluci dvou normálních rozdělení (tady má význam rozkladu normální sdružené hp na podmíněnou a marginální).

Skalární případ

Pro skalární veličiny x a y a konstanty a, b, c platí

$$\begin{aligned} ax^2 + 2bxy + cy^2 &= a \left[x^2 + 2x \frac{b}{a}y + \left(\frac{b}{a}y \right)^2 - \left(\frac{b}{a}y \right)^2 \right] + cy^2 = \\ a \left(x + \frac{b}{a}y \right)^2 + cy^2 - \frac{b^2}{a}y^2 &= a \left(x + \frac{b}{a}y \right)^2 + \frac{ac - b^2}{a}y^2. \end{aligned}$$

Vektorový případ

Pro veličiny x a y ve formě sloupcových vektorů a konstantní matice A, B, C odpovídajících rozměrů, A a C symetrické, platí

$$\begin{aligned} x'Ax + 2x'By + y'Cy &= x'Ax + 2x'AA^{-1}By + (A^{-1}By)'AA^{-1}By - (A^{-1}By)'AA^{-1}By + y'Cy = \\ &= \underbrace{(x + A^{-1}By)'A(x + A^{-1}By)}_{\text{kvadrát}} + \underbrace{y'(C - B'A^{-1}B)y}_{\text{zbytek}}. \end{aligned}$$

Poznámka

Ve vektorovém případě výrazu ax^2 odpovídá $x'Ax$. V tomto smyslu se dějí i úpravy (konstanta - matice) musí být uprostřed.

Výrazy lze ověřit roznásobením konce a porovnáním se začátkem.

1.2 Přírozené podmínky řízení

Přírozené podmínky řízení (Natural Conditions of Control, NCC) jsou podmínky nutné pro to, aby bylo možno konzistentně provést odhadování s modelem, který v sobě obsahuje část s řídicí veličinou, tedy modelem popsaným hp jako např. $f(y_t|u_t, d(t-1), \Theta)$ nebo $f(x_{t+1}|x_t, u_t)$, kde u_t je řízení a Θ nebo x_t jsou odhadované veličiny. Tyto podmínky lze odvodit z předpokladu, že ten kdo odhaduje, je také ten, kdo řídí. Přitom odhad i řízení je počítáno pouze z informace, která je v čase t obsažena v minulých datech $d(t-1)$. Odtud plyne, že jak odhad, tak i řízení v sobě neobsahují další informaci než tu, kterou nesou minulá data $d(t-1)$. Proto např. pro odhad parametrů $f(\Theta|u_t, d(t-1))$ platí, že veškerá dostupná informace o parametru Θ je již obsažena v datech $d(t-1)$ a v řízení u_t již další informace není. Proto je možno jej z podmínky vypustit

$$f(\Theta|u_t, d(t-1)) = f(\Theta|d(t-1)). \quad (1.1)$$

Obrácený vztah lze odvodit obdobnou úvahou, nebo jej lze pomocí Bayesova vzorce odvodit z (1.1)

$$\begin{aligned} f(u_t|d(t-1), \Theta) &= \underbrace{f(\Theta|u_t, d(t-1)) \frac{f(u_t|d(t-1))}{f(\Theta|d(t-1))}}_{\text{Bayes}} = \\ &= \underbrace{f(\Theta|d(t-1)) \frac{f(u_t|d(t-1))}{f(\Theta|d(t-1))}}_{\text{podle (1.1)}} = f(u_t|d(t-1)). \end{aligned}$$

Podobné úvahy lze místo o parametru vést i o odhadovaném stavu.

1.3 Bayesův vzorec

Odvození Bayesova vzorce

Odvození je velice jednoduché. Uvažujme tři náhodné veličiny A , B a C a sdruženou hp pro A , B podmíněnou C .

$$f(A, B|C) = \begin{cases} f(A|B, C) f(B|C) & \text{z jedné strany, nebo} \\ f(B|A, C) f(A|C) & \text{z druhé strany.} \end{cases}$$

Porovnáním obou výrazů na pravé straně dostaneme

$$f(A|B, C) f(B|C) = f(B|A, C) f(A|C).$$

Z této rovnosti pak lze vyjádřit buď $f(A|B, C)$ nebo $f(B|A, C)$ podle potřeby, a tak získáme Bayesův vzorec

$$f(B|A, C) = \frac{f(A|B, C) f(B|C)}{f(A|C)}. \quad (1.2)$$

Hlavní význam Bayesova vzorce je v tom, že přepočítává apriorní hp $f(B|C)$ z čitatele na pravé straně vzorce na aposteriorní hp $f(B|A, C)$ na levé straně. Apriorní hp popisuje náhodnou veličinu B jen v závislosti na náhodné veličině C , zatímco aposteriorní hp využívá informaci také z náhodné veličiny A , a to prostřednictvím hp $f(A|B, C)$.

Hp ve jmenovateli pravé strany výrazu (1.2) nezávisí na B a je tedy jen normalizační konstantou, kterou lze získat integrací (sumací) čitatele

$$f(A|C) = \int_{B^*} f(A|B, C) f(B|C) dB,$$

což, jak dobře víme, odpovídá vzorci pro úplnou pravděpodobnost.

Aplikace Bayesova vzorce

Pro účely odhadu neznámých parametrů modelu $f(y_t|\psi_t, \Theta)$ zvolíme

- A je výstup soustavy y_t ,
- B jsou odhadované parametry Θ a
- C jsou stará data $d(t-1)$ (případně s řízením u_t).

Dostáváme Bayesův vzorec

$$f(\Theta|d(t)) = \frac{f(y_t|\psi_t, \Theta) f(\Theta|d(t-1))}{f(y_t|d(t-1))}$$

podle (??) s tím, že

- stará data $d(t-1)$ v modelu soustavy v podmínce lze nahradit regresním vektorem ψ_t ,
- v případě řízené soustavy, když data jsou $d_t = \{y_t, u_t\}$, je třeba předpokládat **přirozené podmínky řízení** (1.1), při kterých platí

$$f(\Theta|u_t, d(t-1)) = f(\Theta|d(t-1)),$$

tedy Θ a u_t jsou podmíněně nezávislé, jestliže známe stará data $d(t-1)$.

1.4 Multinomiální rozdělení

Multinomiální rozdělení popisuje diskretní náhodnou veličinu, tj. veličinu, která může nabývat jen konečného počtu hodnot $y \in \{1, 2, \dots, n_l\}$ a jejíž jednotlivé hodnoty nastávají s pevnými pravděpodobnostmi. Speciálním případem tohoto rozdělení pro $n_l = 2$, jehož hodnoty jsou ale většinou označovány 0 a 1, je rozdělení alternativní.

Hustotu pravděpodobnosti multinomiálního rozdělení je možno vyjádřit ve formě tabulky

$$\frac{y}{f(y)} \left| \begin{array}{cccc} 1 & 2 & \cdots & n_l \\ p_1 & p_2 & \cdots & p_{n_l} \end{array} \right.,$$

kde p_i jsou pravděpodobnosti, a tak platí $p_i \geq 0$, $i = 1, 2, \dots, n_l$ a $\sum_{i=1}^{n_l} p_i = 1$.

Jiné možné vyjádření multinomiálního rozdělení je

$$f(y) = p_y, \quad y = 1, 2, \dots, n_l.$$

Model diskrétního systému je podmíněná hp, která pro každou konfiguraci hodnot veličin v podmínce popisuje modelovanou veličinu pomocí multinomiálního rozdělení

$$f(y|\psi, \Theta) = \Theta_{y|\psi}.$$

Tuto hp můžeme vyjádřit ve formě tabulky např. pro $y \in \{1, 2\}$ a $\psi = [u, v]'$, kde $u, v \in \{1, 2\}$

$$f(y|u, v)$$

$[u, v]$	$y = 1$	$y = 2$
1, 1	$\Theta_{1 11}$	$\Theta_{2 11}$
1, 2	$\Theta_{1 12}$	$\Theta_{2 12}$
2, 1	$\Theta_{1 21}$	$\Theta_{2 21}$
2, 2	$\Theta_{1 22}$	$\Theta_{2 22}$

kde $\Theta_{i|jk}$ jsou podmíněné pravděpodobnosti (proto je také jejich index rozdělen svislým indexem). Proto musí platit nezápornost $\Theta_{i|jk} \geq 0$, $\forall i, j, k$ a dále součet parametrů pro každou konfiguraci podmínky musí dát jedničku $\sum_{i=1}^2 \Theta_{i|jk} = 1$, $\forall j, k$.¹ Pro účely odhadu je výhodné formálně tento model vyjádřit v tzv. **součinném tvaru**

$$f(y|\psi, \Theta) = \prod_{i \in y^*} \prod_{\varphi \in \psi^*} \Theta_{i|\varphi}^{\delta(i|\varphi, y|\psi)}, \quad (1.3)$$

kde i je index, φ je multiindex (vektorový index), y^* , ψ^* označují množiny hodnot (čísel nebo vektorů) příslušných veličin a $\delta(i|\varphi, y|\psi)$ je Diracův impulz, tj. rovná se jedné pro $i|\varphi = y|\psi$ a jinak je nula. Přepis do součinného tvaru je skutečně formální, protože cokoli na nulu je jedna a po vynásobení zůstane jen $\Theta_{y|\psi}$.

¹Tento požadavek je po úvaze opět zřejmý. Jestliže nastalo $u = j$ a $v = k$, má y právě dvě možnosti: $y = 1$ nebo $y = 2$.

1.5 Dirichletovo rozdělení

Konjugovaným rozdělením² k multinomiálnímu je rozdělení Dirichletovo

$$f(\Theta|d(t)) = \frac{1}{B(\nu_t)} \prod_{i \in y^*} \prod_{\varphi \in \psi^*} \Theta_{i|\varphi}^{\nu_{i|\varphi;t}}, \quad (1.4)$$

kde

ν_t je statistika rozdělení se stejnou strukturou jako má parametr Θ - viz tabulka v odstavci 1.4 a další značení také odpovídá značení zavedenému v tomto odstavci,

$B(\nu)$ je zobecněná beta funkce

$$B(\nu) = \prod_{\varphi \in \psi^*} \frac{\prod_{i \in y^*} \Gamma(\nu_{i|\varphi})}{\Gamma(\sum_{i \in y^*} \nu_{i|\varphi})}, \quad (1.5)$$

kde $\Gamma(\cdot)$ je gama funkce definovaná vztahem

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt, \quad (1.6)$$

pro kterou platí

$$\Gamma(x+1) = x\Gamma(x), \quad x \in R^+. \quad (1.7)$$

1.6 Normální rozdělení

Uvažujme normální regresní model s regresní vektorem ψ_t , koeficienty θ a rozptylem šumu r , kde značíme $\Theta = \{\theta, r\}$. Jeho rovnice je

$$y_t = \psi_t' \theta + e_t, \quad e_t \sim N(0, r).$$

Podmíněná hp tohoto modelu má tvar

$$f(y_t|\psi_t, \Theta) = \frac{1}{\sqrt{2\pi}} r^{-0.5} \exp \left\{ -\frac{1}{2r} (y_t - \psi_t' \theta)^2 \right\}. \quad (1.8)$$

Střední hodnota modelu je

$$E[y_t|\psi_t, \Theta] = \psi_t' \theta,$$

rozptyl je

$$D[y_t|\psi_t, \Theta] = r.$$

Pro účely odhadování je výhodné exponent modelu (1.8) ještě upravit. Budeme postupovat následovně:

²Konjugované rozdělení je takové rozdělení apriorní hp parametrů, které s daným rozdělením, použitým pro model soustavy, produkuje v Bayesově vzorci aposteriorní rozdělení, které si zachovává stejný tvar. Tedy nedochází k tomu, že se tvar aposteriorní hp v každém čase odhadování stává stále složitější, až je nakonec aposteriorní hp pro výpočty nepoužitelná.

- exponent rozdělíme tak, abychom dostali součin dvou členů, z nichž jeden bude obsahovat jen data a druhý jen parametry

$$y_t - \psi'_t \theta = - [-1 \theta'] \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} = - [y_t \psi_t] \begin{bmatrix} -1 \\ \theta \end{bmatrix}$$

(minus je vytknuto formálně a pod kvadrátem se ztratí).

- kvadrát napíšeme jako součin a dosadíme předchozí výrazy

$$\begin{aligned} (y_t - \psi'_t \theta)^2 &= (y_t - \psi'_t \theta) (y_t - \psi'_t \theta) = \\ &= [-1 \theta'] \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} [y_t \psi_t] \begin{bmatrix} -1 \\ \theta \end{bmatrix} = [-1 \theta'] D_t \begin{bmatrix} -1 \\ \theta \end{bmatrix}, \end{aligned}$$

kde $D_t = \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} [y_t \psi_t]$ je tzv. datová matice.

Model (1.8) je s uvedenou úpravou možno zapsat takto

$$f(y_t | \psi_t, \Theta) = \frac{1}{\sqrt{2\pi}} r^{-0.5} \exp \left\{ -\frac{1}{2r} [-1 \theta'] D_t \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}. \quad (1.9)$$

1.7 Inverzní Gauss-Wishartovo (GiW) rozdělení

Toto rozdělení vzniká jako rozdělení součinu normálních rozdělení (tj. jako rozdělení likelihoodu pro normální model). Má tvar

$$f(\Theta | d(t)) \propto r^{-0.5\kappa_t} \exp \left\{ -\frac{1}{2r} [-1 \theta'] V_t \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}, \quad (1.10)$$

kde κ_t a V_t jsou statistiky rozdělení (κ_t se někdy nazývá počítadlo, protože uchovává počet dosud zpracovaných datových vektorů a matice V_t se nazývá rozšířená informační matice).

Matice V_t je symetrická a pozitivně definitní a často se rozkládá

- na submatice

$$V_t = \begin{bmatrix} V_y & V_{y\psi}' \\ V_{y\psi} & V_\psi \end{bmatrix}, \quad (1.11)$$

kde V_y je číslo, $V_{y\psi}$ je sloupcový vektor a V_ψ je čtvercová matice stupně o jeden menší než je V_t ,

- na faktory

$$V_t = L' D L, \quad (1.12)$$

kde L je dolní trojúhelníková matice s jedničkami na diagonále a D je diagonální matice s nezápornými prvky na diagonále. Tento rozklad se nazývá LD-rozklad a pro symetrickou pozitivně definitní matici je jednoznačný. Matice L a D se potom rozkládají na submatice

$$L = \begin{bmatrix} 1 & 0 \\ L_{y\psi} & L_{\psi} \end{bmatrix}, \quad D = \begin{bmatrix} D_y & 0 \\ 0 & D_{\psi} \end{bmatrix},$$

kde $L_{y\psi}$ je sloupcový vektor, L_{ψ} dolní trojúhelníková matice s jedničkami na diagonále, D_y je nezáporné číslo a D_{ψ} je diagonální matice s nezápornými prvky na diagonále.

Uvedené rozklady se dále využijí pro vyjádření potřebných charakteristik rozdělení.

1.8 Bodový odhad podle kvadratického kritéria

Nejprve ukážeme obecně, že bodový odhad optimální podle kvadratického kritéria je roven podmíněné střední hodnotě. Budeme odvozovat odhad $\hat{\Theta}_t$ parametru Θ s a posteriori hp $f(\Theta|d(t))$, který je optimální podle kritéria

$$\min_{\hat{\Theta}_t} E \left[\left(\Theta - \hat{\Theta}_t \right)^2 | d(t) \right]. \quad (1.13)$$

Kritérium umocníme, aplikujeme střední hodnotu a doplníme na čtverec v proměnné $\hat{\Theta}_t$. Dostáváme

$$\begin{aligned} & \min_{\hat{\Theta}_t} E \left[\Theta^2 - 2\hat{\Theta}_t\Theta + \hat{\Theta}_t^2 | d(t) \right] = \\ & = \min_{\hat{\Theta}_t} \left\{ E \left[\Theta^2 | d(t) \right] - 2\hat{\Theta}_t E \left[\Theta | d(t) \right] + \hat{\Theta}_t^2 \right\} = *1* \end{aligned}$$

využili jsme skutečnost, že $\hat{\Theta}_t$ je číslo

$$*1* = \min_{\hat{\Theta}_t} \left\{ E \left[\Theta^2 | d(t) \right] - E \left[\Theta | d(t) \right]^2 + E \left[\Theta | d(t) \right]^2 - 2\hat{\Theta}_t E \left[\Theta | d(t) \right] + \hat{\Theta}_t^2 \right\} = *2*$$

použili jsme výpočetní vzorec pro rozptyl $D[\Theta] = E[\Theta^2] - E[\Theta]^2$

$$*2* = \min_{\hat{\Theta}_t} \left\{ D[\Theta | d(t)] + \left(\hat{\Theta}_t - E[\Theta | d(t)] \right)^2 \right\} = D[\Theta | d(t)]$$

což dá optimální odhad

$$\hat{\Theta}_t = E[\Theta | d(t)].$$

1.9 Bodové odhady parametrů spojitého modelu

Ukážeme odvození MAP (Maximum A Posteriori Probability) bodového odhadu který maximalizuje aposteriorní hp parametrů.

Hledáme tedy maximum aposteriorní hp (1.10)

$$\begin{aligned} f(\Theta|d(t)) &\propto r^{-0.5\kappa} \exp \left\{ -\frac{1}{2r} [-1 \ \theta'] V \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\} = \\ &= r^{-0.5\kappa} \exp \left\{ -\frac{1}{2r} (V_y - 2\theta' V_{y\psi} + \theta' V_{\psi\psi} \theta) \right\}, \end{aligned}$$

kde jsme využili dělení matice V podle (1.11).

Nejdříve budeme hledat maximum podle θ , tj. derivovat podle θ a hledat řešení pro derivaci rovnu nule.³

$$\frac{\partial f(\{\theta, r\} | d(t), r)}{\partial \theta} \propto r^{-0.5\kappa} \exp \left\{ -\frac{1}{2r} [-1 \ \theta'] V \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\} \left(\frac{-1}{2r} \right) (-2V_{y\psi} + 2V_{\psi\psi}\theta) = 0.$$

Odtud pochází vztah

$$\hat{\theta} = V_{\psi}^{-1} V_{y\psi}. \quad (1.14)$$

Výsledek dosadíme do aposteriorní hp. V exponentu obdržíme zbytek po minimalizaci Λ

$$\begin{aligned} \Lambda &= V_y - 2\hat{\theta}' V_{y\psi} + \hat{\theta}' V_{\psi\psi} \hat{\theta} = \\ &= V_y - 2V_{y\psi}' V_{\psi}^{-1} V_{y\psi} + V_{y\psi}' V_{\psi}^{-1} V_{\psi\psi} V_{\psi}^{-1} V_{y\psi}, \end{aligned}$$

a tedy

$$\Lambda = V_y - V_{y\psi}' V_{\psi}^{-1} V_{y\psi}. \quad (1.15)$$

Aposteriorní hp s dosazeným bodovým odhadem (1.14) je

$$f(r|d(t)) \propto r^{-0.5\kappa} \exp \left\{ -\frac{\Lambda}{2r} \right\}.$$

Derivujeme a položíme rovno nule

$$-\kappa \frac{1}{2r} + \Lambda \frac{1}{r^2} = 0,$$

a tedy dostaneme

$$\hat{r} = \frac{\Lambda}{\kappa}. \quad (1.16)$$

$\hat{\theta}$ a \hat{r} jsou bodové odhady, které tady hledáme.

³Derivujeme vektory podle vektorů. Správnost lze ověřit derivováním podle složek a zpětným sestavením do vektorového tvaru.

1.10 Bodové odhady parametrů diskretního modelu

Bodové odhady parametru Θ multinomiálního rozdělení (1.4) dostaneme pouhou normalizací statistiky ν_t tak, aby součty jejích prvků v řádcích jejího maticového vyjádření (jako pro Θ v odstavci (1.4)) byly rovny jedné, tj.

$$\hat{\Theta}_{y|\psi;t} = \frac{\nu_{y|\psi;t}}{\sum_{i \in y^*} \nu_{i|\psi;t}}, \quad \forall y \in y^* \text{ a } \psi \in \psi^*. \quad (1.17)$$

Tento bodový odhad jsme určili jako podmíněnou střední hodnotu parametru s rozdělením podle aposteriorní hp (1.4) - pro přehlednost vynecháme časový index t

$$\begin{aligned} \hat{\Theta}_{y|\psi} &= E[\Theta_{y|\psi}|d(t)] = \int_0^\infty \Theta_{y|\psi} f(\Theta|d(t)) d\Theta = \\ &= \frac{1}{B(\nu)} \int_0^\infty \Theta_{y|\psi} \prod_{i \in y^*} \prod_{\varphi \in \psi^*} \Theta_{i|\varphi}^{\nu_{i|\varphi}} d\Theta = *1*, \end{aligned}$$

kde jsme dosadili za aposteriorní hp z (1.4) a beta funkce B je dána v (1.5). Parametr $\Theta_{y|\varphi;t}$ formálně vyjádříme v součinném tvaru (1.3)

$$\Theta_{y|\psi} = \prod_{i \in y^*} \prod_{\varphi \in \psi^*} \Theta_{i|\varphi}^{\delta(i|\varphi, y|\psi)}$$

a dosadíme. Pokračujeme v úpravě

$$\begin{aligned} *1* &= \frac{1}{B(\nu_t)} \int_0^\infty \prod_{i \in y^*} \prod_{\varphi \in \psi^*} \Theta_{i|\varphi}^{\nu_{i|\varphi} + \delta(i|\varphi, y|\psi)} d\Theta = \\ &= \frac{1}{\prod_{\varphi \in \psi^*} B(\nu_\varphi)} \prod_{\varphi \in \psi^*} \int_0^\infty \prod_{i \in y^*} \Theta_{i|\varphi}^{\nu_{i|\varphi} + \delta(i|\varphi, y|\psi)} d\Theta_{y|\psi} = *2*, \end{aligned}$$

kde

$$B(\nu_\varphi) = \frac{\prod_{i \in y^*} \Gamma(\nu_{i|\varphi})}{\Gamma(\sum_{i \in y^*} \nu_{i|\varphi})} \text{ podle (1.5)}$$

jsme využili předpoklad o nezávislosti parametrů mezi různými konfiguracemi regresního vektoru ψ , tj nezávislosti řádků v tabulkovém vyjádření parametru v odstavci 1.4.

Dále si uvědomíme, že pro jednotlivé komponenty platí

$$\int_0^\infty \prod_{i \in y^*} \Theta_{i|\varphi}^{\nu_{i|\varphi} + \delta(i|\varphi, y|\psi)} d\Theta_{y|\psi} = \begin{cases} B(\nu_\varphi) & \text{pro } \delta = 0, \\ B(\nu_\psi + 1) & \text{pro } \delta = 1. \end{cases}$$

Potom se všechny členy s $\delta = 0$ zkrátí se stejným členem ve funkci B a zůstane jen člen s $\delta = 1$ a odpovídajícím normalizačním členem ve jmenovateli. Pokračujeme ve výpočtu

$$*2* = \frac{B(\nu_\psi + \delta(i, y))}{B(\nu_\psi)} = \frac{\frac{\prod_{i \in y^*} \Gamma(\nu_{i|\psi} + \delta(i, y))}{\Gamma(\sum_{i \in y^*} \nu_{i|\psi} + 1)}}{\frac{\prod_{i \in y^*} \Gamma(\nu_{i|\psi})}{\Gamma(\sum_{i \in y^*} \nu_{i|\psi})}} = *3*.$$

Opět $\delta(i, y)$ je nula všude kromě případu, kdy $y = i$. Všechny ostatní členy se zkrátí. Dostaneme

$$*3* = \frac{\frac{\Gamma(\nu_{y|\psi} + 1)}{\Gamma(\sum_{i \in y^*} \nu_{i|\psi} + 1)}}{\frac{\Gamma(\nu_{y|\psi})}{\Gamma(\sum_{i \in y^*} \nu_{i|\psi})}} = \frac{\nu_{y|\psi}}{\sum_{i \in y^*} \nu_{i|\psi}}.$$

V první úpravě předchozího výrazu jsme využili vlastnosti gama funkce (1.7). Tím jsme dokázali vztah (1.17).

1.11 Logistická regrese

Derivace věrohodnostní funkce

Derivace logaritmu věrohodnostní funkce $\ln L$ s modelem (??) podle Θ je

$$\frac{\partial}{\partial \Theta} \ln L(\Theta) = \sum_{\tau=1}^t \left[y_\tau \psi_\tau - \frac{\exp(z_\tau)}{1 + \exp(z_\tau)} \psi_\tau \right] = \sum_{\tau=1}^t (y_\tau - p_\tau) \psi_\tau,$$

kde podle (??) je $z_\tau = \psi_\tau \Theta$ a tedy $dz_\tau/d\Theta = \psi_\tau$. Dále jsme označili

$$p_\tau = \frac{\exp(z_\tau)}{1 + \exp(z_\tau)} = P(y_t = 1 | \psi_\tau, \Theta).$$

Druhá derivace $\ln L$ podle Θ je

$$\frac{\partial^2}{\partial \Theta^2} \ln L(\Theta) = \frac{\partial}{\partial \Theta} \sum_{\tau=1}^t (y_\tau - p_\tau) \psi_\tau = \sum_{\tau=1}^t \frac{\partial}{\partial \Theta} p_\tau \psi_\tau = \sum_{\tau=1}^t p_\tau (1 - p_\tau) \psi'_\tau \psi_\tau,$$

protože

$$\begin{aligned} \frac{\partial}{\partial \Theta} p_\tau &= \frac{\partial}{\partial \Theta} \frac{\exp(z_\tau)}{1 + \exp(z_\tau)} = \frac{\exp(z_\tau) \psi'_\tau (1 + \exp(z_\tau)) - \exp(z_\tau) \exp(z_\tau) \psi'_\tau}{(1 + \exp(z_\tau))^2} = \\ &= \frac{\exp(z_\tau) \psi'_\tau}{(1 + \exp(z_\tau))^2} = \left(\frac{\exp(z_\tau)}{1 + \exp(z_\tau)} \frac{1}{1 + \exp(z_\tau)} \right) \psi'_\tau = p_\tau (1 - p_\tau) \psi'_\tau. \end{aligned}$$

Pro hledání maxima logaritmu věrohodnostní funkce $\ln L$ je výhodné použít Newtonovu metodu⁴

⁴Výhodné je to zejména pro to, že se nám podařilo analyticky spočítat jak první, tak i druhou derivaci maximalizované funkce $\ln L$.

Newtonův algoritmus

Pomocí Newtonova algoritmu je možné numericky hledat extrém⁵ nelineárních funkcí. Označme takovou funkci $g(x)$, kde $x = [x_1, x_2 \dots x_n]'$. Dále označme gradient této funkce

$$g'(x) = \begin{bmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \\ \dots \\ \frac{\partial g}{\partial x_n} \end{bmatrix}$$

a Hessovu matici

$$g''(x) = \begin{bmatrix} \frac{\partial^2 g}{\partial x_1^2} & \frac{\partial^2 g}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 g}{\partial x_1 \partial x_n} \\ \frac{\partial^2 g}{\partial x_2 \partial x_1} & \frac{\partial^2 g}{\partial x_2^2} & \dots & \frac{\partial^2 g}{\partial x_2 \partial x_n} \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 g}{\partial x_n \partial x_1} & \frac{\partial^2 g}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 g}{\partial x_n^2} \end{bmatrix}.$$

Algoritmus začíná hledáním extrému ve zvoleném bodě $x^{(0)}$ a v dalších bodech $x^{(1)}, x^{(2)}, \dots$ se hledá následujícím způsobem:

provedeme Taylorův rozvoj funkce g v bodě $x^{(i)}$ a vezmeme jeho první tři členy

$$g(x) \doteq g(x^{(i)}) + g'(x^{(i)})(x - x^{(i)}) + \frac{1}{2}g''(x^{(i)})(x - x^{(i)})^2.$$

Následující bod hledání $x^{(i+1)}$ položíme do maxima (minima) rozvoje - do bodu, kde je derivace nulová:

$$g'(x^{(i)}) + g''(x^{(i)})(x^{(i+1)} - x^{(i)}) = 0$$

Z toho plyne:

$$x^{(i+1)} = x^{(i)} - \frac{g'(x^{(i)})}{g''(x^{(i)})}.$$

Iterace provádíme tak dlouho, dokud dva po sobě následující body nejsou dostatečně blízko.

1.12 Rozšířený Kalmanův filtr

Konstrukce rozšířeného KF

Máme nelineární stavový model

$$\begin{aligned} x_{t+1} &= g(x_t, u_t) + w_t, \\ y_t &= h(x_t, u_t) + v_t. \end{aligned}$$

⁵Pozor! Jde samozřejmě o lokální extrémy.

Poslední bodový odhad stavu x_t označíme \hat{x}_t .

Linearizaci funkcí g a h z modelu provedeme rozvojem do Taylorovy řady a zanedbáním členů druhého a vyšších řádů. Dostaneme

$$\begin{aligned} g(x_t, u_t) &\doteq g(\hat{x}_t, u_t) + g'(\hat{x}_t, u_t)(x_t - \hat{x}_t) = \\ &= g'(\hat{x}_t, u_t)x_t + [g(\hat{x}_t, u_t) - g'(\hat{x}_t, u_t)\hat{x}_t], \end{aligned} \quad (1.18)$$

$$\begin{aligned} h(x_t, u_t) &\doteq h(\hat{x}_t, u_t) + h'(\hat{x}_t, u_t)(x_t - \hat{x}_t) = \\ &= h'(\hat{x}_t, u_t)x_t + [h(\hat{x}_t, u_t) - h'(\hat{x}_t, u_t)\hat{x}_t]. \end{aligned} \quad (1.19)$$

Dosazením do modelu dostaneme jeho linearizovanou verzi

$$\begin{aligned} x_{t+1} &\doteq g'(\hat{x}_t, u_t)x_t + [g(\hat{x}_t, u_t) - g'(\hat{x}_t, u_t)\hat{x}_t] + w_t, \\ y_t &\doteq h'(\hat{x}_t, u_t)x_t + [h(\hat{x}_t, u_t) - h'(\hat{x}_t, u_t)\hat{x}_t] + v_t \end{aligned}$$

a s označením

$$Q_x = [g(\hat{x}_t, u_t) - g'(\hat{x}_t, u_t)\hat{x}_t], \quad Q_y = [h(\hat{x}_t, u_t) - h'(\hat{x}_t, u_t)\hat{x}_t],$$

dostaneme lineární tvar modelu

$$x_{t+1} = g'(\hat{x}_t, u_t)x_t + Q_x + w_t, \quad (1.20)$$

$$y_t = h'(\hat{x}_t, u_t)x_t + Q_y + v_t. \quad (1.21)$$

Pro tento model již lze použít obyčejný Kalmanův filtr. Je ale potřeba si uvědomit, že bodový odhad \hat{x}_t bude v každé rovnici jiný. V daném kroku KF se nejprve provádí filtrace, které odpovídá druhá (výstupní) rovnice. Pro ni bude posledním odhadem $x_{t|t-1}$, který pochází z minulého kroku z predikce. Pro první (stavovou) rovnici bude posledním odhadem $\hat{x}_t = x_{t|t}$, který jsme obdrželi v současném kroku KF z filtrace. Odtud je patrné, že proceduru realizující KF je pro nelineární filtraci třeba rozdělit do dvou částí - do filtrace a predikce - a obě části realizovat samostatně. Po filtraci, která končí uprostřed algoritmu, je totiž třeba přepočítat funkce g' a Q_x ze stavové rovnice pro poslední odhad stavu $x_{t|t}$.

Realizace KF

Obecně jsme konstatovali, že metoda rozšířeného KF převede nelineární stavový model na lineární tak, že nelineární funkce na pravých stranách modelů rozvine do nultého a prvního členu Taylorova rozvoje. Výsledkem je linearizovaný tvar modelu (1.20), (1.21). Pro tento model je již možno použít proceduru standardního KF, ovšem pro model s konstantou. Ve skutečnosti je ale možno počítat takto:

1. Střední hodnoty odhadu stavu a predikce výstupu je možno počítat z původního nerozšířeného modelu. Důvod je tento: z rovnic (1.18) a (1.19) je vidět, že derivace g' a h' jsou násobeny přírůstkem stavu $(\xi_t - \hat{\xi}_t)$, kde $\hat{\xi}_t$ je bod, v němž provádíme Taylorův rozvoj,

což je poslední bodový odhad stavu. Stejný bodový odhad je ale dosazen za ξ_t a pomocí něho se počítá příští bodový odhad v algoritmu KF. Přírůstek tedy bude $(\hat{\xi}_t - \hat{\xi}_t) = 0$. Proto se při výpočtu střední hodnoty původního (nerozšířeného) stavu uplatní jen absolutní člen rozvoje, což je pravá strana původního stavového modelu. Část rozšířeného stavu, která obsahuje neznámé parametry, zůstává při výpočtu střední hodnoty stejná.

2. Konstanta Q_x nebo Q_y se při výpočtu kovariancí neuplatní. Důvod ukážeme ve zjednodušeném případě na výpočtu rozptylu z náhodné veličiny X plus konstanta a

$$\begin{aligned} D[X + a] &= E[(X + a - E[X + a])^2] = E[(X + a - E[X] - a)^2] = \\ &= E[(X - E[X])^2] = D[X]. \end{aligned}$$

Rozptyl náhodné veličiny $X + a$ a náhodné veličiny X je stejný. Podobně to v obecném případě platí i pro kovariance.

Realizace rozšířeného KF je tedy následující:

1. Z minulého kroku máme “poslední” odhad stavu $\hat{\xi}_t = \xi_{t|t-1}$.
2. Vezmeme tu část odhadu, která odpovídá původnímu stavu, a podle původní výstupní rovnice

$$y_t = Ax_t$$

spočteme predikci výstupu \hat{y}_t .

3. Kovariance počítáme podle matic rozšířené výstupní rovnice

$$y_t = \tilde{A}^d \xi_t$$

s vynechanou konstantou Q_y .

4. Přepočteme (filtrace) $\xi_{t|t-1} \rightarrow \xi_{t|t}$
Pozn.: K vývoji odhadovaných parametrů dochází jen v této části algoritmu.
5. Za poslední odhad stavu dosadíme $\hat{\xi}_t = \xi_{t|t}$.
6. Kovariance přepočteme s maticemi rozšířeného stavového modelu bez konstanty

$$\xi_{t+1} = \tilde{M}^d \xi_t.$$

7. Přepočet (predikce) stavu provedeme podle původní stavové rovnice

$$x_{t+1} = Mx_t,$$

přičemž část stavu ξ_t odpovídající neznámým parametrům zůstává nezměněna.