

Model

1 Spojitý model

Veličiny v dopravním systému jsou náhodné posloupnosti indexované diskretním časem t . V každém časovém okamžiku to jsou náhodné veličiny, po změření dostaneme realizace náhodné veličiny. Tyto náhodné veličiny mohou být buď spojité, nebo diskretní. V této kapitole se budeme zabývat případem, kdy výstupem systému je spojitá náhodná veličina indexovaná diskretním časem (náhodná posloupnost se spojitými hodnotami).

1.1 Princip stochastického modelu

Model je obrazem systému. Je to matematický popis závislosti modelované veličiny na jiných (vhodně vybraných) veličinách. Ve většině praktických případů funkci závislosti přímo neznáme, a tak vztahy mezi veličinami popisujeme za pomoci parametrů. Hodnoty parametrů určujeme pomocí odhadu z měřených dat. Tato neznalost parametru a poruchy ve veličinách (ať už vznikající při samotném generování nebo v procesu měření) způsobují, že je takový model prakticky vždy pod vlivem neurčitosti.

Příklad [stochastický model]

Princip modelování lze demonstrovat na situaci, kdy byla náhle zablokována silnice a přijíždějící auta se stávají do kolony. Modelovanou veličinou y_t je délka kolony narůstající v diskretním čase t . Tato délka závisí na intenzitě proudu I_t a na průměrné délce automobilu (včetně mezery mezi automobily) θ . Pro vývoj kolony v čase můžeme psát

$$y_t = y_{t-1} + \theta I_t, \quad y_0 = 0,$$

kde y_0 je počáteční délka kolony a 0 je okamžik vzniku blokace. Tento model lze uvažovat jako deterministický.

Jestliže jsme naměřili s periodou vzorkování 90 sec

t	1	2	3	4	5	6	7	8	9	10
I_t	8	6	5	9	8	9	12	5	7	4

a uvažujeme-li délku vozidla 8 m, pak vývoj délky kolony v metrech bude

t	1	2	3	4	5	6	7	8	9	10
y_t	64	112	152	224	288	360	456	496	552	584

Je ale zřejmé, že tento deterministický model nemůže platit přesně. Počítali jsme s délkou auta 8 m, což je jen hrubý odhad, a rovněž jsme zaokrouhlovali počet automobilů v každé periodě měření intenzity I_t . Proto realističtější model je model stochastický, který můžeme vyjádřit následující rovnicí s přičteným šumem reprezentujícím neurčitosti a poruchy v systému

$$y_t = y_{t-1} + \theta I_t + e_t.$$

V tomto stochastickém modelu lze uvažovat i počáteční podmínku y_0 za náhodnou. Pokud jsme “v průměru” odhadovali dobře, bude mít šum přibližně nulovou střední hodnotu. V případě, že systém příliš nemění své stochastické vlastnosti, bude rozptyl šumu konstantní. Pokud dále “vysvětlující veličiny” modelu nesou o modelované veličině dostatek informací, budou jednotlivé složky náhodné posloupnosti navzájem nezávislé. Takové náhodné posloupnosti říkáme **bílý šum** (white noise).

Model, ke kterému jsme dospěli v minulém příkladě, obsahuje deterministickou (tady lineární) funkci počítající modelovanou veličinu y_t v závislosti na nezávislé veličině I_t , jakémsi parametru θ , který vyjadřuje konkrétní míru souvislosti mezi y_t a I_t . Stochastická část modelu je šum e_t , který (podobně jako v regresní analýze) “dorovnává” modelovanou veličinu i s poruchami proti jejím ideálním hodnotám (predikci).

Na tuto rovnici lze také pohlížet jako na transformaci náhodné veličiny e_t s rozdělením $f(e_t)$ na náhodnou veličinu y_t s podmíněným rozdělením $f(y_t|y_{t-1}, I_t, \theta)$. Abychom transformační rovnici mohli použít, musí být veličiny y_{t-1} , I_t , θ známé, tj. musí se vyskytnout v podmínce rozdělení modelované veličiny. Transformace potom představuje pouhé posunutí z nulové střední hodnoty šumu, na střední hodnotu modelované veličiny y_t , která je

$$E[y_t|y_{t-1}, I_t, \theta] = E[y_{t-1} + \theta I_t + e_t|y_{t-1}, I_t, \theta] = y_{t-1} + \theta I_t.$$

Rozptyl y_t je stejný jako rozptyl e_t . (Obojí si zkuste dokázat).

1.2 Lineární regresní model s normálním šumem

Nejčastěji používaným spojitým modelem je lineární regresní model s normálním rozdělením náhodné složky - šumu. Tento model má obecně rovnici

$$y_t = \psi_t' \Theta + e_t, \tag{1.1}$$

kde y_t je modelovaná veličina,

ψ_t je regresní vektor (vektor veličin, které vysvětlují y_t),

Θ je parametr modelu (zahrnuje regresní koeficienty θ a rozptyl šumu r),

e_t je šum s normálním rozdělením s nulovou střední hodnotou a konstantním rozptylem r .

Statický regresní model obsahuje v regresním vektoru jen vstupní veličiny (většinou měřené v současném časovém okamžiku), nikoli zpožděné hodnoty modelované veličiny. Jeho rovnici tedy můžeme psát ve tvaru

$$y_t = c_1 v_{1,t} + c_2 v_{2,t} + \dots + c_m v_{m,t} + k + e_t,$$

kde $\psi_t' = [v_{1,t}, v_{2,t}, \dots, v_{m,t}, 1]$ je regresní vektor a $\theta = [c_1, c_2, \dots, c_m, k]$ je vektor regresních koeficientů. Některé z veličin regresního vektoru mohou představovat řízení.

Typický dynamický regresní model popisující vstup a výstup systému je dán regresním vektorem ve tvaru

$$\psi_t' = [u_t, y_{t-1}, u_{t-1}, \dots, y_{t-n}, u_{t-n}, 1]$$

a jemu odpovídajícím vektorem parametrů

$$\theta' = [b_0, a_1, b_1, \dots, a_n, b_n, k].$$

Dosadíme-li do (1.1), dostaneme **model ve tvaru rovnice**

$$y_t = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + \dots + a_n y_{t-n} + b_n u_{t-n} + k + e_t. \quad (1.2)$$

Tato rovnice pak definuje podmíněnou hp modelované veličiny y_t takto:

Šum e_t má podle definice rozdělení

$$f(e_t) = \frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} e_t^2 \right\}.$$

Podle transformační rovnice (1.2), která má Jakobián roven jedné, platí

$$e_t = y_t - \psi'_t \theta$$

a po dosazení do rozdělení šumu dostaneme

$$f(y_t | \psi_t, \Theta) = \frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} (y_t - \psi'_t \theta)^2 \right\}, \quad (1.3)$$

což je hledaná podmíněná hp modelované veličiny - **model systému ve tvaru hustoty**.

Řádem regresního vektoru nazveme největší zpoždění modelované veličiny v regresním vektoru.

Statický model je model nultého řádu.

V modelu 1. řádu závisí modelovaná veličina na své minulé hodnotě. Zpoždění jiných veličin (řízení, externího vstupu) řád modelu neovlivní.

1.3 Diferenciální a diferenční rovnice ve spojitém modelu

Diskretizace diferenciální rovnice

Reálný čas plyne spojitě a většina dějů v reálném světě probíhá rovněž spojitě. Je proto dobře si uvědomit, že na pozadí našeho systému stojí spojitý proces a náš systém vzniká diskretizací tohoto procesu. Celou situaci budeme demonstrovat na velmi jednoduché (idealizované) soustavě popsané diferenciální rovnicí prvního řádu bez pravé strany - jedná se tedy o problém dozrívání počátečních podmínek. Spojité proměnné zde budeme značit velkými písmeny s argumentem času τ v závorce - např. $Y(\tau)$. Spojitý systém je tedy popsán rovnicí (Y' zde značí derivaci)

$$Y'(\tau) + aY(\tau) = 0, \quad Y(0) = y_0.$$

Naším úkolem je vyjádřit tuto diferenciální rovnici pomocí diferenční rovnice v diskrétním čase t s periodou vzorkování T ($\tau = tT$) tak, aby vzorky generované diferenční rovnicí ležely na funkci, která je řešením diferenciální rovnice a daných počátečních podmínek.

Postup odvození diferenční rovnice (diskretizace) je následující. Řešení diferenciální rovnice odhadneme ve známém exponenciálním tvaru a po dosazení počátečních podmínek dostaneme

$$Y(\tau) = y_0 \exp \{-a\tau\}.$$

Do řešení dosadíme diskretní čas $\tau = tT$. Diskretizované řešení napíšeme pro čas $t + 1$

$$Y((t + 1)T) = y_0 \exp\{-a(t + 1)T\}.$$

Na levé straně je vzorek Y v čase $t + 1$, který značíme y_{t+1} . Na pravé straně upravíme exponenciálu

$$\exp\{-a(t + 1)T\} = \exp\{-atT\} \exp\{-aT\}.$$

První člen na pravé straně předchozí rovnosti je vzorek Y v diskretním čase t , druhý člen je konstanta. Dosadíme do diskretizovaného řešení a dostaneme

$$y_{t+1} = y_0 \exp\{-aT\} y_t \quad \text{nebo} \quad y_t = y_0 \exp\{-aT\} y_{t-1}$$

a tedy

$$y_t = Ay_{t-1},$$

kde $A = y_0 \exp\{-aT\}$. To je diferenční rovnice, která splňuje naše požadavky.

Podobným způsobem lze diskretizovat i diferenciální rovnice vyšších řádů, i když postup je poněkud zdlouhavější.

Typy diferenciálních a diferenčních rovnic

Pro řešení diferenciální rovnice s konstantními koeficienty bez pravé strany je rozhodující řešení její charakteristické rovnice. Pro rovnice druhého řádu

$$y'' + a_1 y' + a_0 y = 0$$

je charakteristická rovnice

$$\lambda^2 + a_1 \lambda + a_0 = 0.$$

Mohou nastat tři typy řešení charakteristické rovnice:

1. *dvě reálná řešení* λ_1 a λ_2 - diferenciální rovnice má řešení ve tvaru lineární kombinace dvou exponenciál

$$y = \alpha_1 \exp\{\lambda_1 \tau\} + \alpha_2 \exp\{\lambda_2 \tau\}$$

2. *jeden dvojnásobný kořen* λ - řešení je v tvaru

$$y = (\alpha_1 \tau + \alpha_2) \exp\{\lambda \tau\}$$

3. *dva komplexně sdružené kořeny* $\lambda + j\omega$ a $\lambda - j\omega$ - řešení diferenciální rovnice je

$$y = \exp\{\lambda \tau\} [\alpha_1 \sin(\omega \tau) + \alpha_2 \cos(\omega \tau)].$$

Pro diferenční rovnice druhého řádu s konstantními koeficienty bez pravé strany platí podobná analýza. Pro rovnici

$$y_{t+2} + a_1 y_{t+1} + a_0 y_t = 0$$

je opět rozhodující charakteristická rovnice

$$z^2 + a_1 z + a_0 = 0,$$

kde z je operátor jednokrokového předstihu $y_{t+1} = zy_t$.

Tato rovnice má podobné typy řešení (i je imaginární jednotka):

1. *dva reálné kořeny* λ_1 a λ_2 - řešení diferenční rovnice je

$$y_t = \beta_1 \lambda_1^t + \beta_2 \lambda_2^t$$

2. *jeden dvojnásobný kořen* λ - řešení je

$$y_t = \beta_1 \lambda^t + \beta_2 t \lambda^t$$

3. *dva komplexně sdružené kořeny* $\lambda \pm i\omega$ - řešení je

$$y_t = (\lambda^2 + \omega^2)^{t/2} (\beta_1 \sin(\omega t) + \beta_2 \cos(\omega t))$$

1.4 Regresní model ve stavovém tvaru

Stav systému x_t je taková veličina (vektor veličin), který v sobě obsahuje informaci o celém dosavadním vývoji systému. Jestliže známe starý stav a aktuální řídicí veličiny, jsme schopni konstruovat pravděpodobnostní popis nového stavu. Model stavu má tedy vždy 1. řád. V lineárním případě je dán rovnicí

$$x_t = Mx_{t-1} + Nu_t + w_t,$$

kde M , N jsou matice odpovídajících rozměrů, w_t je bílý šum.

V řadě případů je výhodnější pracovat s modelem 1. řádu, i když mnoho-rozměrným, než s regresním modelem s řádem vyšším než jedna. Konstrukcí stavového modelu je celá řada. Takováto procedura je známa např. z přepočtu diferenciální (nebo diferenční) rovnice n -tého řádu na n rovnic, prvního řádu. Zde uvedeme tu nejjednodušší, která přímo vychází ze struktury regresního modelu.

Uvažujme obecně regresní model

$$y_t = b_0 u_t + a_1 y_{t-1} + \dots + a_n y_{t-n} + b_n u_{t-n} + k + e_t.$$

Definujeme vektor (stav) $x_t = [y_t, u_t, y_{t-1}, u_{t-1}, \dots, y_{t-n}, u_{t-n}, 1]'$ a pro něj s pomocí regresního modelu a identických rovnic sestavíme model v následujícím tvaru

$$x_t = Mx_{t-1} + Nu_t + \epsilon_t \tag{1.4}$$

se strukturou (z důvodu přehlednosti ji uvedeme pro regresní model řádu 2)

$$\underbrace{\begin{bmatrix} y_t \\ u_t \\ y_{t-1} \\ u_{t-1} \\ 1 \end{bmatrix}}_{x_t} = \underbrace{\begin{bmatrix} a_1 & b_1 & a_2 & b_2 & k \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_A \underbrace{\begin{bmatrix} y_{t-1} \\ u_{t-1} \\ y_{t-2} \\ u_{t-2} \\ 1 \end{bmatrix}}_{x_{t-1}} + \underbrace{\begin{bmatrix} b_0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_B u_t + \underbrace{\begin{bmatrix} e_t \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}}_{w_t},$$

kde první rovnice představuje regresní model a zbytek jsou identity.

Význam tohoto převodu ukážeme na příkladě.

Příklad [regresní model]

Pro regresní model 2. řádu se zanedbatelným šumem (tedy šumem s tak malým rozptylem, že šum lze zanedbat) určete hodnotu y_5 pro počáteční podmínky y_{-1}, y_0 a dané řízení u_{-1}, u_0, \dots, u_5 .

Teoreticky je úloha jednoduchá. Pro předpověď y_1 dosadíme do regresního vektoru regresního modelu a hodnotu výstupu prostě vypočteme při současném zanedbání šumu

$$y_1 = b_0 u_1 + a_1 y_0 + b_1 u_0 + a_2 y_{-1} + b_2 u_{-1} + k.$$

Při výpočtu y_2 použijeme známé hodnoty a také vypočtenou hodnotu výstupu y_1

$$\begin{aligned} y_2 &= b_0 u_2 + a_1 y_1 + b_1 u_1 + a_2 y_0 + b_2 u_0 + k = \\ &= b_0 u_2 + a_1 (b_0 u_1 + a_1 y_0 + b_1 u_0 + a_2 y_{-1} + b_2 u_{-1} + k) + b_1 u_1 + a_0 y_0 + b_0 u_0 + k = \\ &= b_0 u_2 + (a_1 b_0 + b_1) u_1 + (a_1 b_1 + b_0) u_0 + a_1 b_2 u_{-1} + (a_1^2 + a_0) y_0 + a_1 a_2 y_{-1} + (a_1 + 1) k. \end{aligned}$$

Tímhle způsobem lze pokračovat až do času $t = 5$. Není sice příliš složité postřehnout, jakým způsobem se vyvíjejí koeficienty u jednotlivých veličin, nicméně tato úloha je daleko snazší, použijeme-li stavový model

$$x_t = Mx_{t-1} + Nu_t,$$

kde jsme podle předpokladu zadání zanedbali šum. Platí

$$x_1 = Mx_0 + Nu_1,$$

kde $x_0 = [y_0, u_0, y_{-1}, u_{-1}, 1]'$. Dále

$$x_2 = Mx_1 + Nu_2 = M(Mx_0 + Nu_1) + Nu_2 = M^2 x_0 + MNu_1 + Nu_2,$$

$$x_3 = Mx_2 + Nu_3 = M(M^2 x_0 + MNu_1 + Nu_2) + Nu_3 = M^3 x_0 + M^2 Nu_1 + MNu_2 + Nu_3.$$

Ani nemusíme pokračovat a můžeme psát obecný vzorec:

$$x_t = M^t x_0 + \sum_{i=0}^{t-1} M^i Nu_{t-i}.$$

Požadované y_5 je prvním prvkem vektoru x_5 , tedy

$$x_5 = M^5 x_0 + \sum_{i=0}^4 M^i Nu_{5-i}, \quad \text{a} \quad y_5 = x_{1,5}.$$

2 Diskrétní a logistický model

2.1 Diskrétní model

Pokud mají všechny veličiny vstupující do modelu konečný počet hodnot, hovoříme o diskrétním modelu. Seřadíme-li tyto veličiny do vektoru (tzv. rozšířený regresní vektor $\Psi_t = [y_t, \psi_t]'$), pak realizací náhodných veličin v tomto vektoru dostaneme určitou konfiguraci jejich hodnot. Diskrétní systém má konečný počet takových konfigurací, které nazýváme módy. Příkladem je odbočení auta v T-křižovatce doprava, nebo doleva (2 módy) nebo kolona nebyla a není, nebyla a je, byla a není, byla a je (4 módy). Popis diskrétního systému proto můžeme provést velmi obecně, a to tak, že každé konfiguraci hodnot veličin v rozšířeném regresním vektoru přiřadíme její pravděpodobnost, tedy

$$f(y_t | \psi_t, \Theta) = \Theta_{y_t | \psi_t}, \quad (2.1)$$

kde multi-index (rozšířený regresní vektor s diskrétními hodnotami) $y_t|\psi_t = [y_t, \psi_{1;t}, \psi_{2;t}, \dots, \psi_{n_\psi;t}]$ je vektor indexů (hodnot diskrétních veličin) a znaménko “|” je použito jen formálně a ukazuje, které veličiny jsou modelovány a které se nachází v podmínce modelu.

Podobně jako u regresního modelu nazveme **řádem modelu** největší zpoždění modelované veličiny v regresním vektoru.

Příklad [diskrétní model]

Model budeme demonstrovat na jednoduchém, ale přesto dostatečně obecném příkladě tzv. “řízení koruny s pamětí”. Jedná se o systém s dvouhodnotovým výstupem s hodnotami 1 a 2 (např. rub a líc), který je závislý na minulém výstupu (např. po dopadu se koruna zašpiní a tím se rozváže pravděpodobnostní poměr jednotlivých stran) a na dvouhodnotovém řízení, rovněž s hodnotami 1 a 2 (např. způsob, jak se položí mince na ruku při hodu). Potom platí

$$y_t \in \{1, 2\}, \quad \psi_t = [u_t, y_{t-1}] \in \{1, 2\} \times \{1, 2\}, \quad \Psi_t = [y_t | u_t, y_{t-1}]$$

a model je možno zapsat ve formě tabulky

$f(y_t u_t, y_{t-1})$	y_t	
	1	2
$[u_t, y_{t-1}]$		
1, 1	0.3	0.7
1, 2	0.8	0.2
2, 1	0.1	0.9
2, 2	0.2	0.8

kde $\Theta_{1|11} = 0.3$ a např. $\Theta_{2|21} = 0.9$.

Obecně pro parametry diskrétního modelu musí platit

$$\sum_{i \in y^*} \Theta_{i|jk} = 1, \quad \forall j, k.$$

Využití diskrétního modelu je možné zejména v situacích, kdy:

1. Modelovaná veličina je určitou klasifikací nějakého jevu - např. nehoda nastala nebo ne-nastala nebo nastala nehoda bez zranění, se zraněním, s úmrtím.
2. Nezávislé veličiny (v regresním vektoru) označují okolnosti, které doprovázejí naměřenou hodnotu modelované veličiny. Tyto veličiny mohou být nominální, například veličina denní doba s hodnotami: svítání, den, soumrak, noc.

Je potřeba si uvědomit, že čím více veličin obsahuje regresní vektor a čím více různých hodnot tyto veličiny mohou nabýt, tím větší je počet stavů takového systému. Abychom se o každém stavu něco dozvěděli, potřebujeme obrovský počet dat. Při tvorbě diskrétního modelu se proto snažíme do regresního vektoru vybrat jen podstatné veličiny (vzhledem k vysvětlení hodnot modelované veličiny) a u každé veličiny definovat co nejmenší počet hodnot. Pomoci může např. spojování hodnot: pro denní dobu místo jitra, den, soumrak, noc lze definovat světlo, šero, tma (pokud tomu nebrání konkrétní zadání úlohy).

2.2 Logistický model

Logistický model použijeme v případě, kdy modelovaná veličina je diskrétní (dvouhodnotová) $y_t \in \{0, 1\}$ a předpokládáme, že její hodnota stochasticky závisí na veličinách

$$\psi_t = [1, x_{1;t}, x_{2;t}, \dots, x_{n;t}]',$$

kteřé mohou být jak diskrétní, tak i spojité. Jednička v regresním vektoru představuje opět respektování konstanty modelu. V případě, kdy veličiny z vektoru ψ_t jsou všechny diskrétní, jedná se o diskrétní model, kterým jsme se zabývali v kapitole 2.1.

Logistický model má tvar

$$f(y_t | \psi_t, \Theta) = \frac{\exp(y_t z_t)}{1 + \exp(z_t)}, \quad (2.2)$$

kde je použita lineární regrese

$$z_t = \psi_t' \Theta + e_t. \quad (2.3)$$

Význam logistického modelu je v tom, že popisuje diskrétní veličinu y_t v závislosti na spojité veličině z_t která je vypočtena na základě lineární regrese z veličin, které mohou být jak diskrétní, tak i spojité.

Příklad [logistický model]

Modelujeme závažnost nehody $y_t \in \{0, 1\}$ v závislosti na rychlosti automobilu $v_{1;t} \in R^+$ a osvětlení $v_{2;t} \in \{1, 2, 3\}$, kde 1 znamená světlo, 2 označuje šero a 3 představuje tmu. Lineární regrese (2.3) má rovnici

$$z_t = \theta_0 + \theta_1 v_{1;t} + \theta_2 v_{2;t} + e_t$$

a model (2.2) bude tvořen pravděpodobnostmi

$$\begin{aligned} P(y_t = 1 | \psi_t, \Theta) &= \frac{\exp(z_t)}{1 + \exp(z_t)}, \\ P(y_t = 0 | \psi_t, \Theta) &= \frac{1}{1 + \exp(z_t)}, \end{aligned}$$

kde první rovnici je pro $y_t = 1$ a druhá pro $y_t = 0$.

Princip fungování tohoto modelu je následující. Lineární regrese (2.3) mapuje regresní vektor ψ_t na veličinu z_t , které přirozeně nabývá hodnot z celé reálné osy. Funkce (2.2) potom transformuje tuto reálnou osu do intervalu $(0, 1)$, takže nakonec dostaneme hodnotu ve tvaru pravděpodobnosti. Hodnoty této pravděpodobnosti blízké 1 ukazují na hodnotu $y_t = 1$, hodnoty blízké 0 mluví pro $y_t = 0$.

Jiné vyjádření logistického modelu

Jiné možné vyjádření stejného logistického modelu je možno napsat pomocí funkce *logit*, která je definována

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right). \quad (2.4)$$

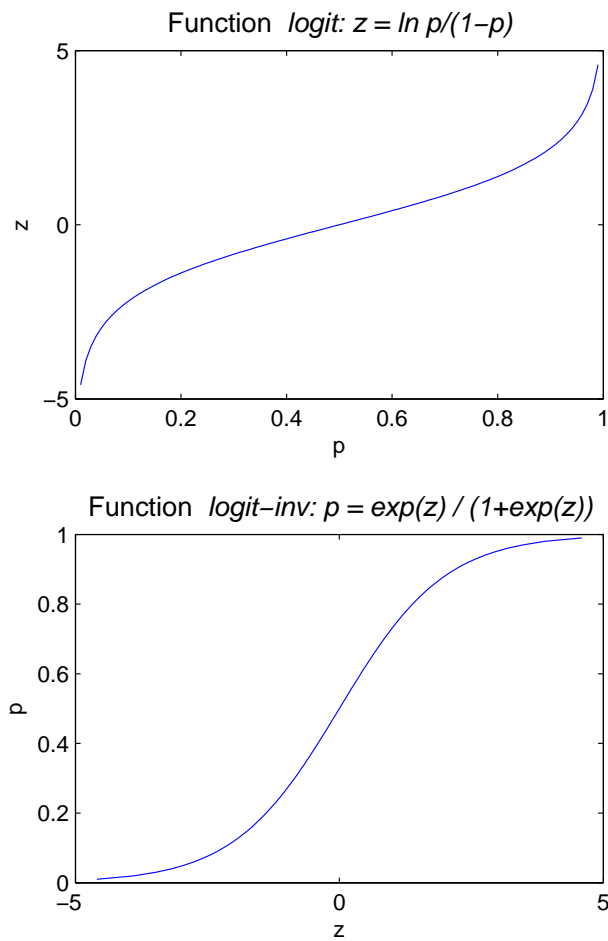
Dosadíme-li $p = P(y_t = 1 | \psi_t, \Theta)$, dostaneme logistický model ve tvaru

$$\ln \left(\frac{P(y_t = 1 | \psi_t, \Theta)}{1 - P(y_t = 1 | \psi_t, \Theta)} \right) = \psi_t \Theta + e_t.$$

Samozřejmě pro $y_t = 0$ platí

$$P(y_t = 0 | \psi_t, \Theta) = 1 - P(y_t = 1 | \psi_t, \Theta)$$

Průběh funkce *logit* je vykreslen na Obrázku 2.1.



Obrázek 2.1: Funkce $z = \text{logit}(p)$

Funkce *logit* převádí proměnnou p , která má rozsah $(0, 1)$ na veličinu z , která má hodnoty z celé reálné osy. Toho se využije při logistické regresi tak, že hodnoty $z = \psi'\Theta \in R$ se transformují na pravděpodobnosti $p \in (0, 1)$. Transformace je patrná z obrázku: velké kladné hodnoty z se zobrazí blízko $p = 1$, hodnoty z kolem nuly padnou blízko $p = 0.5$ a velké záporné hodnoty z dávají hodnoty blízko $p = 0$.

Funkci *logit_inv* inverzní k funkci *logit* dostaneme otočením horního obrázku o 45° (viz dolní obrázek). Tato funkce převádí naopak reálná čísla na pravděpodobnosti.

Odhad

3 Odhad spojitého modelu

Model je matematickým popisem vybraných veličin sledovaného procesu. Tyto veličiny popisujeme stochasticky (pomocí hustot pravděpodobnosti) v závislosti na jiných vybraných veličinách většinou jako lineární vazby pomocí diferenčních rovnic.

Při návrhu modelu řešíme dva kroky:

1. návrh struktury modelu, tj. výběr veličin a počtu kroků jejich zpoždění, na kterých modelovaná veličina závisí,
2. určení parametrů modelu, které vyjadřují konkrétní závislosti ve sledovaném procesu.

Zde se budeme zabývat druhým bodem - odhadem parametrů modelu.

3.1 Bayesovské odhadování

Parametry z hlediska bayesovství

V klasické statistice se odhadované parametry, ale i jiné odhadované veličiny, považují za náhodné veličiny. Jejich popisem je hustota pravděpodobnosti. Jestliže je pro konstrukci této hp použita jen předběžná (expertní) znalost, hovoříme o apriorním popisu. Jestliže jsou dále využita i měřená data, dostáváme aposteriorní popis.

V případě popisu parametrů se jedná o hp $f(\Theta)$, která se nazývá **apriorní hp** a která odráží prvotní znalosti o parametrech. V průběhu odhadování se měří data d_t v časech $t = 1, 2, \dots, T$, kde T je horizont intervalu odhadování. Informace z měřených dat se postupně využívá pro zpřesnění popisu parametrů a původní apriorní hp se vyvíjí na **aposteriorní hp**

$$f(\Theta) \xrightarrow{d_1} f(\Theta|d(1)) \xrightarrow{d_2} f(\Theta|d(2)) \xrightarrow{d_3} \dots \xrightarrow{d_T} f(\Theta|d(T))$$

Střední hodnota aposteriorní hp vypovídá o bodových odhadech parametrů (viz Přílohy 10.8), kovarianční matice o nepřesnosti odhadů. Přepočty hp popisující parametry se provádí na základě **Bayesova vztahu**.

Bayesův vztah

Vývoj hp parametrů, tj. postupné upřesňování hp parametrů podle informace přicházející z měřených dat d_1, d_2, \dots, d_t , se provádí podle Bayesova vzorce (viz Přílohy ??)

$$f(\Theta|d(t)) \propto f(y_t|\psi_t, \Theta) f(\Theta|d(t-1)), \quad (3.1)$$

který se počítá pro $t = 1, 2, \dots, T$ a startuje s tzv. apriorní hp $f(\Theta|d(0)) = f(\Theta)$, která odráží apriorní, expertní znalost.

Vztah (3.1) je možno zapsat také rovnou pro koncový čas T

$$f(\Theta|d(T)) = \prod_{t=1}^T f(y_t|\psi_t, \Theta) f(\Theta|d(0)) = L_T(\Theta) f(\Theta|d(0)),$$

kde

$$L_T(\Theta) = \prod_{t=1}^T f(y_t|\psi_t, \Theta) \quad (3.2)$$

je věrohodnostní funkce (likelihood).

Z uvedeného je patrné, že bayesovský odhad je vlastně klasický odhad maximální věrohodnosti korigovaný apriorní hp.

Reprodukovatelné parametrické vyjádření aposteriorní hp

Vztah (3.1) je rekurze pro funkce. Ta je prakticky nerealizovatelná, proto je třeba jednotlivé hp vyjádřit v konkrétním tvaru, který závisí na konečném počtu číselných charakteristik a tuto funkcionální rekurzi převést na rekurzi algebraickou pro charakteristiky rozdělení.¹

Navíc je třeba parametrické vyjádření volit tak, aby při postupném odhadu v čase nevznikaly nové a nové charakteristiky, tj, aby se formální tvar hp parametrů reprodukoval. Jinak se tato hp velmi rychle stane tak složitou, že ji prakticky není možno počítat v rozumném čase.

Například bude-li mít model normální rozdělení, zvolíme apriorní hp také jako normální. Násobením normálních hp vede opět na normální rozdělení, které je určené svou střední hodnotou a rozptylem. Ty je možno počítat přímo ze středních hodnot a rozptylů modelu soustavy a apriorní hp. Dostáváme tak rekurzi na číslech, nikoli na funkcích. Pokud má model tuto vlastnost, řekneme, že při odhadu dostáváme **reprodukující se aposteriorní hp**.

Obecně tento problém vyřešit nelze, ale pro případy, kterými se zde budeme zabývat, tj. pro regresní spojitý model s normálním rozdělením a pro diskrétní model s multinomiálním rozdělením, takové tzv. adjungované apriorní distribuce na parametrech existují ve formě inverzního Gauss-Wishartova a Dirichletova rozdělení. Budeme o nich podrobněji hovořit v následujících kapitolách.

¹Např. přepočítat funkci $f(x)$ na $g(x)$ lze jenom tak, že ji přepočítáme $f(x) \rightarrow g(x)$ pro každý bod $x \in R$. Jestliže ale je $f(x) = \exp\{ax\}$ a $g(x) = \exp\{bx\}$, pak stačí přepočítat $a \rightarrow b$. Cela funkce už je dána svým předpisem.

Příklad [odhadování bez reprodukce aposteriorní hp]

V tomto příkladu budeme ilustrovat situaci, kdy odhadujeme s modelem, který nevede na reprodukcí se aposteriorní hp.

Uvažujme model $f(y|a) = ay^2 - 2ay + \frac{4a+3}{6}$ pro $y \in (0, 2)$ a parametrem $a \in (-\frac{3}{4}, \frac{3}{2})$. Protože známe rozsah parametru a a nemáme o něm žádnou apriorní informaci, budeme apriorní hp volit jako rovnoměrnou, tj. $f(a) = \frac{4}{9}$ na intervalu $a \in (-\frac{3}{4}, \frac{3}{2})$.

V prvním kroku odhadu naměříme y_1 a dostaneme

$$f(a|y_1) \propto f(y_1|a) f(a) = \left(ay_1^2 - 2ay_1 + \frac{4a+3}{6} \right) \frac{4}{9}.$$

V druhém kroku

$$f(a|y_1, y_2) \propto f(y_2|a) f(a|y_1) = \left(ay_2^2 - 2ay_2 + \frac{4a+3}{6} \right) \left(ay_1^2 - 2ay_1 + \frac{4a+3}{6} \right) \frac{4}{9}$$

a tak dále. Z toho je dobře vidět, že formální tvar aposteriorní hp je stále složitější, protože je v něm třeba si pamatovat stále složitější výrazy. Po prvním kroku pracujeme s y_1^2 a y_1 . Po druhém už to bude $y_1^2 y_2^2$, $y_1^2 y_2$, $y_1 y_2^2$, $y_1 y_2$, y_1^2 , y_2^2 , y_1 , y_2 a tak dále.

Příklad [Odhadování s reprodukcí aposteriorní hp]

Zde ukážeme odhad s modelem, který vede na reprodukcí se aposteriorní hp.

Vezmeme model $f(y_t|a) = \exp\{-ay_t\}$, $y_t \geq 0$, $a > 0$. Jako apriorní hp uvažujme $f(a) = \exp\{-ay_0\}$. Potom po prvním kroku odhadu máme

$$f(a|y_1) \propto f(y_1|a) f(a) = \exp\{-ay_1\} \exp\{-ay_0\} = \exp\{-a(y_1 + y_0)\},$$

po druhém

$$f(a|y_1, y_2) \propto \exp\{-a(y_2 + y_1 + y_0)\}$$

a obecně

$$f(a|y(k)) \propto \exp\left\{-a \sum_{i=0}^k y_i\right\}.$$

Je zřejmé, že když označíme statistiku odhadu $S_k = \sum_{i=0}^k y_i$, pak přepočítání statistiky v čase k je

$$S_k = S_{k-1} + y_k.$$

Formální tvar hp parametrů se nemění a vzorec pro přepočítání statistik má stále stejný tvar.

Výsledek odhadování

Výsledkem procedury odhadu je aposteriorní hp

$$f(\Theta|d(t)),$$

kteřá dává úplný stochastický popis parametrů - tedy výčet všech možných hodnot a jejich pravděpodobnosti výskytu. Pokud je to možné (většinou z důvodů spočítatelnosti), mělo by se tam, kde máme neznámé parametry, počítat s touto celou distribucí.

Není-li možné využít v dalších výpočtech celou aposteriorní hp nebo potřebujeme-li jako odhady čísla, můžeme přejít k bodovým odhadům

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

Obě tyto varianty využití procedury odhadování lze dobře ilustrovat v dalším odstavci Odhad výstupu systému.

Odhad výstupu

Předpověď výstupu je popsána prediktivní hp $f(y_t|\psi_t, d(t-1))$. Protože parametry Θ obecně neznáme, nesmí se vyskytovat v podmínce.²

Prediktivní hp dostaneme tak, že modelujeme obě neznámé veličiny y_t i Θ pomocí sdružené hp, kterou rozložíme podle řetězového pravidla a integrujeme přes Θ , abychom dostali popis jen pro výstup y_t

$$\begin{aligned} f(y_t|\psi_t, d(t-1)) &= \int_{\Theta^*} f(y_t, \Theta|\psi_t, d(t-1)) d\Theta = \\ &= \int_{\Theta^*} f(y_t|\psi_t, \Theta) f(\Theta|d(t-1)) d\Theta, \end{aligned} \quad (3.3)$$

kde chybějící veličiny v podmínkách vypadly z důvodu nezávislosti. Předpověď výstupu v čase t (jako náhodné veličiny) je tedy dána modelem $f(y_t|\psi_t, \Theta)$ a vyjádřením neznámého parametru pomocí aposteriorní hp parametrů $f(\Theta|d(t-1))$.

Poznámka

Dobře si všimněte, jak bayesovství zachází s neznámou veličinou (tady parametrem). Každého by napadlo: mám model a potřebuji do něho parametr. Udělám odhad a ten tam dosadím. To ale není optimální. Správné je použití úplné pravděpodobnosti tak jako v (3.3). Do modelu postupně dosazují všechny možné hodnoty parametrů a počítám vážený průměr - dosazení a sčítání dělá integrál, váhy jsou dány aposteriorní hp. ◁

Bodový odhad výstupu

Bodový odhad parametru zkonstruovaný s pomocí aposteriorní hp je (viz Příloha 10.8) podmíněná střední hodnota parametru

$$\hat{\Theta}_t = E[\Theta|d(t)] = \int_{\Theta^*} \Theta f(\Theta|d(t)) d\Theta. \quad (3.4)$$

Tento bodový odhad už není úplný, ale jen částečný popis parametru (např. střední hodnota nic neříká o rozptylu). Pokud je ale aposteriorní hp dostatečně „štíhlá“, a to ona po správném odhadu je,³ je možno ji nahradit Diracovým impulzem $\delta(\Theta - \hat{\Theta}_t)$, kde $\delta(0)$ je jedna a jinde nula, tedy

$$f(\Theta|d(t-1)) \rightarrow \delta(\Theta - \hat{\Theta}_{t-1}). \quad (3.5)$$

Dosadíme za aposteriorní hp a dostaneme

$$f(y_t|\psi_t, d(t-1)) = \int_{\Theta^*} f(y_t|\psi_t, \Theta) \delta(\Theta - \hat{\Theta}_{t-1}) d\Theta = f(y_t|\psi_t, \hat{\Theta}_{t-1}). \quad (3.6)$$

Výsledek je právě ten, který bychom čekali. Jestliže máme model s neznámými parametry a spočteme bodové odhady parametrů, pak tyto bodové odhady použijeme místo neznámých parametrů. Nedostáváme optimální řešení, ale pro „štíhlou aposteriorní hp“ je to řešení často přijatelné.

²Jinak bychom tuto hp nemohli přímo použít - neměli bychom za Θ co dosadit. Parametry ale potřebujeme pro model. Proto musíme odhad Θ zabudovat do konstrukce prediktivní hp.

³Při odhadu získáváme informaci, tím klesá neurčitost a s ní i rozptyl odhadu.

3.2 Odhad parametrů regresního modelu

Odhad parametrů normálního regresního modelu na základě apriorní informace a dat $d(t)$ měřených průběžně na systému provedeme na základě Bayesova vzorce (3.1) s modelem (1.3) a vhodně zvolenou apriorní hp $f(\Theta|d(0)) = f(\Theta)$.

Model

Regresní model s regresním vektorem ψ_t , jemu odpovídajícím vektorem regresních koeficientů θ a normálním šumem s rozptylem r má tvar

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

Pro účely odhadu je výhodné tuto hp (výraz v exponentu) upravit do následujícího tvaru (viz Příloha 10.6, rovnice (10.9))

$$f(y_t|\psi_t, \Theta) = \frac{1}{\sqrt{2\pi}} r^{-0.5} \exp \left\{ -\frac{1}{2r} [-1 \theta'] D_t \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}, \quad (3.7)$$

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

Dále v Příloze 10.7 v rovnici (10.10) je uveden tvar aposteriorní hp, který odpovídá normálnímu regresnímu modelu (jedná se o tzv. konjugovanou aposteriorní hp k normálnímu regresnímu modelu, tj. takový popis parametrů, jehož tvar se při odhadu podle Bayesova vzorce reprodukuje)

$$f(\Theta|d(\tau)) \propto r^{-0.5\kappa_\tau} \exp \left\{ -\frac{1}{2r} [-1 \theta'] V_\tau \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}, \quad (3.8)$$

kde V_τ a κ_τ jsou statistiky odhadu. Pro časový okamžik t vzorec s $\tau = t$ udává aposteriorní hp, pro $\tau = t - 1$ je to apriorní hp.

Rekurze pro statistiku

Dosazením do Bayesova vzorce (3.1) dostaneme

$$\begin{aligned} & \underbrace{r^{-0.5\kappa_t} \exp \left\{ -\frac{1}{2r} [-1 \theta'] V_t \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}}_{\text{aposteriorní hp}} \propto \\ & \underbrace{r^{-0.5} \exp \left\{ -\frac{1}{2r} [-1 \theta'] D_t \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}}_{\text{model}} \underbrace{r^{-0.5\kappa_{t-1}} \exp \left\{ -\frac{1}{2r} [-1 \theta'] V_{t-1} \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}}_{\text{apriorní hp}}, \end{aligned}$$

odkud porovnáním obou stran tohoto vztahu dostaneme rovnice pro přepočítání statistik

$$V_t = V_{t-1} + D_t, \quad (3.9)$$

$$\kappa_t = \kappa_{t-1} + 1. \quad (3.10)$$

Uvedené statistiky se nazývají: V - rozšířená informační matice, κ - počítadlo vzorků a datová matice D_t . je

$$D_t = \begin{bmatrix} y_t \\ \psi_t \end{bmatrix} \begin{bmatrix} y_t & \psi_t' \end{bmatrix},$$

Algoritmus odhadu

Postup odhadu parametrů je následující:

1. Konstrukce apriorní hp $f(\Theta|d(0))$. Obecně to není jednoduchá záležitost. Jedná se o převod předpokladů nebo požadavků týkajících se parametrů Θ (jako třeba Θ má nezáporné prvky, nebo menší než nějaká horní hranice apod.) do apriorních statistik konstruované apriorní hp.

Výstupem jsou apriorní statistiky V_0 a κ_0 a z nich zkonstruovaná apriorní hp podle (3.8) s dosazenými apriorními statistikami.

2. Měření dat d_1, d_2, \dots kde $d = \{y, u\}$ a průběžný přepočítání statistik podle vztahů (3.9) a (3.10) tj.

$$V_t = V_{t-1} + D_t \text{ a } \kappa_t = \kappa_{t-1} + 1.$$

Výstupem jsou aposteriorní statistiky V_t a κ_t .

3. Konstrukce aposteriorní hp $f(\Theta|d(t))$ dosazením do vztahu (3.8) s $\tau = t$ a statistikami V_t, κ_t .
Výstupem je aposteriorní hp $f(\Theta|d(t))$.

4. Výpočet bodových odhadů $\hat{\Theta}_t$ (jsou-li potřeba). Bodové odhady jsou dány jako podmíněná střední hodnota (viz Příloha 10.8)

$$\hat{\Theta} = E[\Theta|d(t)] = \int_{\Theta^*} \Theta f(\Theta|d(t)) d\Theta.$$

V Příloze 10.9, v rovnicích (10.14) a (10.16) je ukázáno, že platí

$$\hat{\Theta}_t = V_{\psi}^{-1} V_{y\psi}, \quad \hat{r}_t = \frac{V_y - V_{y\psi}' V_{\psi}^{-1} V_{y\psi}}{\kappa_t}. \quad (3.11)$$

Matice V_{ψ} a vektor $V_{y,\psi}$ dostaneme rozdělením informační matice V_t na submatice

$$V_t = \begin{bmatrix} V_y & V_{y\psi}' \\ V_{y\psi} & V_{\psi} \end{bmatrix}, \quad (3.12)$$

kde V_y je skalár, $V_{y\psi}$ je sloupcový vektor, $V_{y\psi}'$ je řádkový vektor a V_{ψ} je čtvercová matice a

κ_t je počítadlo datových vzorků, pro které platí $\kappa_{\tau} = \kappa_{\tau-1} + 1$, $\tau = 1, 2, \dots, t$ s počáteční hodnotou κ_0 (apriorní statistika).

Výsledkem uvedeného algoritmu je:

- konstrukce aposteriorní hp podle bodu 3,
- bodové odhady parametrů (3.11).

Bodový odhad výstupu s bodovým odhadem parametru

Střední hodnotu výstupu y_t (tedy jeho optimální odhad) dostaneme přímo z modelu takto

$$\hat{y}_t = E[y_t | \psi_t, d(t-1)] = E[\psi_t' \hat{\theta}_t + e_t] = \psi_t' \hat{\theta}_t.$$

To znamená, že do modelu bez šumu⁴ dosadíme data a odhady parametrů a jednoduše spočteme výstup.

Odhadování parametrů regresního modelu a jeho výstupu budeme ilustrovat na příkladech.

Příklad [odhad regresního modelu]

Simulujte spojitý dynamický systém popsaný regresním modelem 2. řádu s regresním vektorem

$$\psi_t = [u_t, y_{t-1}, u_{t-1}, y_{t-2}, u_{t-2}, 1]',$$

jemu odpovídajícími regresními koeficienty

$$\theta = [1, 0.6, 0.5, -0.2, -0.3, 0.1]'$$

a rozptylem šumu $r = 0.01$. Odhadněte parametry $\Theta = \{\theta, r\}$ tohoto systému.

Simulace se provádí s pomocí modelu ($\sigma = \sqrt{r} = 0.1$)

$$y_t = \psi_t' \theta + \sigma e_t = u_t + 0.6y_{t-1} + 0.5u_{t-1} - 0.2y_{t-2} - 0.3u_{t-2} + 0.1 + 0.1e_t,$$

kde

$e_t \sim N_{e_t}(0, 1)$ je realizace standardního normálního šumu, ψ_t je regresní vektor ze zadání úlohy, θ jsou regresní koeficienty a σ je směrodatná odchylka šumu.

Pro odhad vytváříme rozšířený regresní vektor

$$\Psi_t = [y_t, \psi_t']' \quad (3.13)$$

pro $t = 1, 2, \dots, n_t$, se kterým přepočítáváme informační matici a počítadlo

$$\begin{aligned} V_t &= V_{t-1} + \Psi_t \Psi_t', \\ \kappa_t &= \kappa_{t-1} + 1, \end{aligned}$$

s počátečními (apriorními) hodnotami V_0 a κ_0 .

Bodové odhady určíme rozkladem informační matice podle (10.11)

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

a dále podle vzorců

$$\hat{\theta}_t = V_\psi^{-1} V_{y\psi} \quad \text{a} \quad \hat{r}_t = \frac{V_y - V_{y\psi}' V_\psi^{-1} V_{y\psi}}{\kappa_t}. \quad (3.14)$$

Podrobné řešení je uvedeno v následujícím programu

⁴Odhadem je střední hodnota. Střední hodnota šumu je nula, takže predikce se provádí jakoby s modelem bez šumu.

```

clc , clear all
// Simulation and estimation of second order
// regression model

nt=100; // number of data

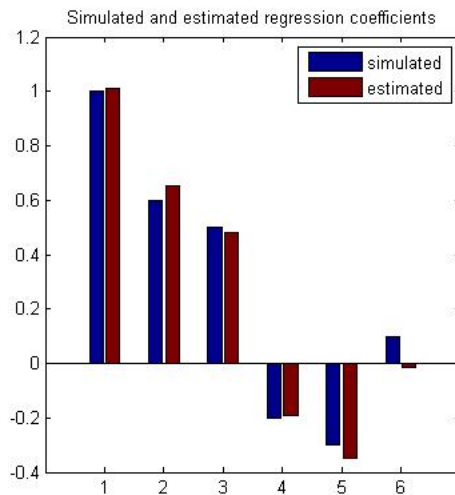
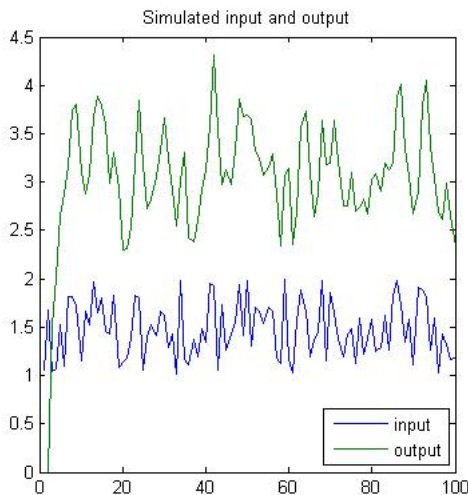
// SIMULATION
th=[1 .6 .5 -.2 -.3 .1]'; // regression coefficients
r=.01; // noise variance
s=sqrt(r); // standard deviation of noise
y=zeros(1,nt); // zero initial conditions + declar.
u=ones(1,nt)+rand(1,nt); // input declaration
// time loop for simulation
for t=3:nt
    psi=[u(t) y(t-1) u(t-1) y(t-2) u(t-2) 1]';
    y(t)=psi'*th + s*randn;
end

// ESTIMATION
V=zeros(7); // initial statistics
// time loop for estimation
for t=3:nt
    Psi=[y(t) u(t) y(t-1) u(t-1) y(t-2) u(t-2) 1]';
    V=V+Psi*Psi';
end
Vy=V(1,1);
Vyp=V(2:end,1);
Vp=V(2:end,2:end);
Eth=inv(Vp)*Vyp; // point est. of regr. coeff.
Cth=(Vy-Vyp'*inv(Vp)*Vyp)/nt; // point est. of noise var.

// print
Simulated_noise_variance=r
Estimated_noise_variance=Cth

```

Výsledky programu jsou na obrázcích



Příklad [bodový odhad regresního modelu]

Alternativní postup při odhadu parametrů normálního regresního modelu vycházející z metody nejmenších čtverců (který je ale s předchozím způsobem ekvivalentní) je následující.

Do rovnice regresního modelu (z předchozího příkladu)

$$y_t = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + a_2 y_{t-2} + b_2 u_{t-2} + k + e_t$$

postupně dosazujeme měřená data a rovnice pro $t = 1, 2, \dots, N$ píšeme pod sebe

$$y_1 = b_0 u_1 + a_1 y_0 + b_1 u_0 + a_2 y_{-1} + b_2 u_{-1} + k + e_1$$

$$y_2 = b_0 u_2 + a_1 y_1 + b_1 u_1 + a_2 y_0 + b_2 u_0 + k + e_2$$

...

$$y_N = b_0 u_N + a_1 y_{N-1} + b_1 u_{N-1} + a_2 y_{N-2} + b_2 u_{N-2} + k + e_N.$$

Vytvořenou soustavu rovnic zapíšeme maticově

$$Y = \Psi\theta + E,$$

kde Y je vektor modelované veličiny y_t , Ψ je matice regresních vektorů v řádcích, θ je vektor parametrů v pořadí odpovídajícím pořadí veličiny v regresním vektoru a E je vektor šumů.

Pro uvažovaný model bude mít soustava tvar

$$\begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix} = \begin{bmatrix} u_1 & y_0 & u_0 & y_{-1} & u_{-1} & 1 \\ u_2 & y_1 & u_1 & y_0 & u_0 & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ u_N & y_{N-1} & u_{N-1} & y_{N-2} & u_{N-2} & 1 \end{bmatrix} \begin{bmatrix} b_0 \\ a_1 \\ b_1 \\ a_2 \\ b_2 \\ k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_1 \\ \dots \\ e_N \end{bmatrix}$$

Bodový odhad parametrů je

$$\hat{\theta} = (\Psi'\Psi)^{-1} \Psi Y,$$

odhad rozptylu šumu je

$$\hat{r} = Y' (Y - \hat{\theta}\Psi) = Y' E_p,$$

kde $E_p = Y - \hat{\theta}\Psi$ je vektor chyb predikce $Y_p = \hat{\theta}\Psi$.

Poznámka

Z porovnání obou uvedených příkladů plyne

$$Y'Y = V_y, \quad Y'\Psi = V_{y\psi} \quad \text{a} \quad \Psi'\Psi = V_\psi.$$

◁

4 Odhad diskrétního a logistického modelu

4.1 Odhad parametrů diskrétního modelu

Diskrétní model popisuje systém, v němž jsou všechny veličiny diskrétní. Model je reprezentován tabulkou pravděpodobností. Model i úlohy s ním spojené jsou velice jednoduché, ale

- práce s tabulkami může být poněkud nezvyklá,
- pokud mají veličiny větší počet různých hodnot, bude tabulka modelu neúnosně velká. V tom případě je lépe přejít na logistickou regresi.

Model a jeho součinný tvar

Model (2.1) obsahuje parametry $\Theta_{y|\psi}$ (pravděpodobnosti jednotlivých konfigurací rozšířeného regresního vektoru $\Psi_t = [y_t, \psi_t]'$), které jsou v obecném případě neznámé, a tak je třeba je odhadovat z měřených dat. Pro odhad použijeme Bayesův vzorec (3.1), diskrétní model (2.1) a vhodně zvolený apriorní model parametrů $f(\Theta|d(0)) = f(\Theta)$. Dle odstavce 3.1 o analytickém tvaru hp parametrů a o reprodukovatelnosti její struktury je třeba tuto hp parametrů zvolit v analytickém tvaru, a to takovém, aby se při postupném násobení modelem soustavy jeho analytický tvar reprodukoval. Za tímto účelem přepíšeme model soustavy (2.1) formálně do tzv. **součinného tvaru**

$$f(y_t|\psi_t, \Theta) = \prod_{y|\psi \in \Psi^*} \Theta_{y|\psi}^{\delta(y|\psi, y_t|\psi_t)},$$

kde $y|\psi$ je multiindex (tj. vektorový index), který může nabývat všech možných konfigurací hodnot přípustných pro jednotlivé veličiny v něm obsažených; Ψ^* je množina všech takových konfigurací; symbol $\delta(y|\psi, y_t|\psi_t)$ je Kroneckerova funkce, která se rovná jedné, když platí $y|\psi = y_t|\psi_t$ (tj. $y = y_t$ a $\psi_1 = \psi_{1;t}$ až $\psi_{n_\psi} = \psi_{n_\psi;t}$), v ostatních případech je rovna nule.

Příklad [model falešné mince]

Pro model falešné koruny ($y = 1$ je líc, $y = 2$ je rub) lze zapsat jednotlivé pravděpodobnosti ve tvaru $f(y = 1) = p_1$ a $f(y = 2) = p_2$, kde $p_1, p_2 \geq 0$ a $p_1 + p_2 = 1$. Tento model můžeme zapsat následovně:

$$f(y) = p_1^{\delta(y,1)} p_2^{\delta(y,2)} = \prod_{i=1}^2 p_i^{\delta(y,i)}.$$

Statistika

Přepis do součinnového tvaru je čistě formální záležitost. Pro všechna $y|\psi \neq y_t|\psi_t$ dostáváme $\Theta^0 = 1$ a jen pro $y|\psi = y_t|\psi_t$ je $\Theta^1 = \Theta_{y_t|\psi_t}$. Nicméně je pro nás součinnový tvar návodem, jak volit apriorní model parametrů. Ten píšeme ve tvaru Dirichletova rozdělení - viz Přílohy 10.5

$$f(\Theta|d(0)) = \prod_{y|\psi \in \Psi^*} \Theta_{y|\psi}^{\nu_{y|\psi;0}},$$

kde $\nu_{y|\psi;0}$ je apriorní statistika (pro čas $t = 0$) pro odhad parametru Θ . Tato statistika je matice stejných rozměrů jako Θ . Podobně jako model ji můžeme reprezentovat pomocí tabulky. Počáteční statistika má tedy následující tvar

$$\nu_{y|\psi;0} \tag{4.1}$$

$[u_0, y_{-1}]$	$y_0 = 1$	$y_0 = 2$
1, 1	$\nu_{1 11}$	$\nu_{2 11}$
1, 2	$\nu_{1 12}$	$\nu_{2 12}$
2, 1	$\nu_{1 21}$	$\nu_{2 21}$
2, 2	$\nu_{1 22}$	$\nu_{2 22}$

Přepočítání statistiky

Předchozí odvozené vztahy dosadíme do Bayesova vzorce (3.1) a dostaneme pro $t = 1$

$$\begin{aligned} f(\Theta|d(1)) &\propto \underbrace{\prod_{y|\psi \in \Psi^*} \Theta_{y|\psi}^{\delta(y|\psi, y_1|\psi_1)}}_{\text{model}} \underbrace{\prod_{y|\psi \in \Psi^*} \Theta_{y|\psi}^{\nu_{y|\psi;0}}}_{\text{apriorní}} = \\ &= \underbrace{\prod_{y|\psi \in \Psi^*} \Theta_{y|\psi}^{\delta(y|\psi, y_1|\psi_1) + \nu_{y|\psi;0}}}_{\text{aposteriorní}} = \prod_{y|\psi \in \Psi^*} \Theta_{y|\psi}^{\overbrace{\nu_{y|\psi;1}}^{\text{nová statistika}}}, \end{aligned}$$

kde

$$\nu_{y|\psi;1} = \delta(y|\psi, y_1|\psi_1) + \nu_{y|\psi;0}$$

je statistika modelu parametrů $f(\Theta|d(1))$ pro čas $t = 1$.

Dále měříme data pro $t = 2, 3, \dots, N$ a přepočítáváme statistiku ν podle obecného vzorce

$$\nu_{y|\psi;t} = \delta(y|\psi, y_t|\psi_t) + \nu_{y|\psi;t-1}. \quad (4.2)$$

Význam přepočtu statistiky ne následující: Po každém změření dat přičteme jedničku do toho políčka statistiky (viz (4.1)), které odpovídá dané konfiguraci hodnot rozšířeného regresního vektoru $y_t|\psi_t$. Každé políčko tedy obsahuje počet, kolikrát příslušná konfigurace hodnot dosud nastala.

Aposteriorní hp parametrů se statistikou ν_t má tvar

$$f(\Theta|d(t)) \propto \prod_{y|\psi \in \Psi^*} \Theta_{y|\psi}^{\nu_{y|\psi;t}} \quad (4.3)$$

Bodový odhad parametrů

Bodový odhad parametru Θ je (viz Přílohy 10.10)

$$\hat{\Theta}_{y|\psi;t} = \frac{\nu_{y|\psi;t}}{\sum_{y \in y^*} \nu_{y|\psi;t}}, \quad (4.4)$$

což zcela odpovídá statistické definici pravděpodobnosti jevu označeného indexem $y|\psi$. Pro každou konfiguraci regresního vektoru ψ je v čitateli počet případů, kolikrát nastala daná hodnota y v rámci tohoto regresního vektoru (počet příznivých pokusů) a ve jmenovateli je celkový počet pokusů.

Tabulka bodových odhadů má stejný tvar jako samotný parametr nebo statistika

$$\hat{\Theta}_t \quad (4.5)$$

$[u_t, y_{t-1}]$	$y_t = 1$	$y_t = 2$
1, 1	$\hat{\Theta}_{1 11;t}$	$\hat{\Theta}_{2 11;t}$
1, 2	$\hat{\Theta}_{1 12;t}$	$\hat{\Theta}_{2 12;t}$
2, 1	$\hat{\Theta}_{1 21;t}$	$\hat{\Theta}_{2 21;t}$
2, 2	$\hat{\Theta}_{1 22;t}$	$\hat{\Theta}_{2 22;t}$

a prvky této tabulky se počítají podle vzorce (4.4).

Bodový odhad výstupu

Podobně jako v případě spojitého regresního modelu i pro diskretní model s neznámými parametry je možno neznámé parametry nahradit jejich bodovými odhady - z tabulky (viz (4.1)) vezmeme řádek odpovídající kombinaci hodnot v regresním vektoru ψ_t a v něm jako odhad \hat{y}_t výstupu y_t vezmeme tu hodnotu, která má větší pravděpodobnost.

Příklad [odhad s diskretním modelem]

V dopravní oblasti byly sledovány nehody a byly rozlišeny podle závažnosti na lehké (jen hmotná škoda) a těžké (vážné zranění nebo smrt). Nehody jako modelovanou veličinu označíme y_t (kde t nyní označuje pořadí nehody, nikoli čas) s hodnotami $y_t = 1$ - lehká nehoda a $y_t = 2$ - těžká nehoda. Předpokládáme, že na typ nehody mají hlavní vliv následující veličiny: ψ_1 - rychlost (1 normální, 2 velká), ψ_2 - počasí (1 sucho, 2 mokro) a ψ_3 - osvětlení (1 světlo, 2 tma).

Po dobu jednoho roku jsme měřili data a získali 18 následujících záznamů.

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
y_t	1	1	2	1	2	1	1	1	2	1	2	2	1	1	1	2	1	1
$\psi_{1,t}$	1	2	2	2	2	1	1	2	2	2	2	1	1	2	1	2	2	1
$\psi_{2,t}$	2	1	1	1	2	2	2	2	1	2	2	2	2	2	1	2	1	2
$\psi_{3,t}$	1	2	1	2	1	2	2	1	2	2	2	2	2	2	2	1	2	1

Model pro popis nehod podle (2.1) s využitím měřených dat bude mít tvar $f(y_t|v_t)$ s následující tabulkou

$$\Theta_{y|\psi}$$

$[v_{1;t}, v_{2;t}, v_{3;t}]$	$y_t = 1$	$y_t = 2$
[1 1 1]	$\Theta_{1 111}$	$\Theta_{2 111}$
[1 1 2]	$\Theta_{1 112}$	$\Theta_{2 112}$
[1 2 1]	$\Theta_{1 121}$	$\Theta_{2 121}$
[1 2 2]	$\Theta_{1 122}$	$\Theta_{2 122}$
[2 1 1]	$\Theta_{1 211}$	$\Theta_{2 211}$
[2 1 2]	$\Theta_{1 212}$	$\Theta_{2 212}$
[2 2 1]	$\Theta_{1 221}$	$\Theta_{2 221}$
[2 2 2]	$\Theta_{1 222}$	$\Theta_{2 222}$

Statistika odhadu podle (4.1) je reprezentována stejnou tabulkou. Aktualizace statistiky začíná s apriorní tabulkou a dále pokračuje podle vzorce (4.2), který říká: pro každé $t = 1, 2, \dots, n_t$ podle hodnot příslušných veličin y_t a $\psi_t = [v_{1;t}, v_{2;t}, v_{3;t}]'$ najděte v tabulce odpovídající políčko a k němu přičtěte jedničku.

V našem příkladě zvolíme nulovou apriorní tabulku, která odpovídá nulové apriorní informaci. S touto nulovou počáteční statistikou ν_0 dostaneme odhadovou statistiku

$$\nu_{y|\psi;0}$$

$[v_{1;t}, v_{2;t}, v_{3;t}]$	$y_t = 1$	$y_t = 2$
[1 1 1]	0	0
[1 1 2]	1	0
[1 2 1]	2	0
[1 2 2]	3	1
[2 1 1]	0	1
[2 1 2]	3	1
[2 2 1]	1	2
[2 2 2]	2	1

odkud (prostou normalizací řádků tabulky na součet jedna) dostaneme odhad parametrů modelu.

$$\hat{\Theta}_{y|\psi;n_t}$$

$[v_{1;t}, v_{2;t}, v_{3;t}]$	$y_t = 1$	$y_t = 2$
[1 1 1]	–	–
[1 1 2]	1	0
[1 2 1]	1	0
[1 2 2]	3/4	1/4
[2 1 1]	0	1
[2 1 2]	3/4	1/4
[2 2 1]	1/3	2/3
[2 2 2]	2/3	1/3

Výsledek je poměrně špatný. Parametry v prvním řádku nelze určit (zde statistika nebyla vůbec přepočtena), řádek 2,3 a 5 je deterministický (ve skutečnosti je to dáno tím, že do těchto řádků přišel jen jediný údaj) a ostatní řádky vypadají lépe, nicméně každý řádek odpovídá pokusu s hodem mincí. Umíme si představit, kolik hodů je třeba, abychom alespoň trochu objektivně posoudili pravděpodobnosti jejích stran. Podle hodnot statistiky vidíme, že do těchto řádků jsou započteny maximálně čtyři údaje. To je velmi málo a je zřejmé, že vzhledem k dimenzi tabulky statistiky máme neúnosně málo dat. Nedostatek dat v těchto případech je velmi častý a je třeba možnosti úlohy v takovém případě dobře zvážit.

Jednou z možností, jak řešit nedostatek dat, je uvažovat další informaci a tou je informace expertní. V krajním případě je dokonce možné postavit model na expertní (apriorní) informaci a změřená data použít jen k jakési korekci modelu. Tuto možnost ukážeme v další části příkladu.

Apriorní statistiku je možno sestavit takto: pro všechny řádky tabulky (tj. pro všechny možné regresní vektory)

1. určíme, jaké pravděpodobnosti bychom přiřadili hodnotám y pro konkrétní kombinaci hodnot ψ ,
2. tyto pravděpodobnosti násobíme číslem, které vyjadřuje míru důvěry k našemu přiřazení.

Např.: první regresní vektor v tabulce je $\psi = [1, 1, 1]'$, tj. rychlost = normální, počasí = sucho, světlo = dobré. V této situaci rozhodujeme o typu nehody. Můžeme říci, že jsou to ideální podmínky, a tak nehoda může být jen lehká. Volíme proto první řádek statistiky

$$\nu_{\{111;0\}} = [9, 1].$$

To odpovídá 10 údajům o nehodách, z nichž 9 bylo lehkých a 1 těžká.

Druhý regresní vektor v tabulce je $\psi = [1, 1, 2]$ - rychlost = normální, počasí = sucho, světlo = šero. Za šera jsou někdy podmínky jízdy ošemetné zvláště pro chodce, kteří jsou špatně vidět. Proto volíme

$$\nu_{\{112;0\}} = [1, 4],$$

což odpovídá pěti záznamům, jeden s lehkou a čtyři s těžkou nehodou.

Stejně budeme pokračovat i pro další regresní vektory, a tak získáme následující tabulku pro apriorní statistiku.

$$\nu_0$$

$[v_1, v_2, v_3]$	$y = 1$	$y = 2$
[1 1 1]	9	1
[1 1 2]	1	4
[1 2 1]	2	2
[1 2 2]	2	3
[2 1 1]	3	2
[2 1 2]	3	7
[2 2 1]	3	7
[2 2 2]	1	9

Odhadem z dat získáme parametry:

$$\hat{\Theta}_{y|\psi;n_t} \tag{4.6}$$

$[v_{1;t}, v_{2;t}, v_{3;t}]$	$y_t = 1$	$y_t = 2$
[1 1 1]	9/10	1/10
[1 1 2]	1/3	2/3
[1 2 1]	2/3	1/3
[1 2 2]	5/9	4/9
[2 1 1]	1/2	1/2
[2 1 2]	3/7	4/7
[2 2 1]	4/13	9/13
[2 2 2]	3/13	10/13

V řádku, kde nepřišla žádná data (např 1. řádek), zůstaly apriorní odhady. Tam, kam data přišla, je apriorní informace korigována daty.

4.2 Odhad parametrů logistického modelu

V případě, kdy modelovaná veličina je diskrétní a závisí jak na diskrétních tak i na spojitých veličinách, použijeme model logistické regrese. Tento model lze použít i v případě, kdy všechny veličiny jsou diskrétní ale mají velký počet různých hodnot, takže čistě diskrétní model by měl příliš vysokou dimenzi.

Zde se budeme zabývat odhadem modelu zavedeného podle (2.2, 2.3), tedy modelem, který má tvar

$$\text{logit}(p_t) = \psi_t' \Theta + e_t,$$

kde $p_t = P(y_t = 1|\psi_t)$ a $y_t \in \{0, 1\}$, P je pravděpodobnost a y_t je dvouhodnotový výstup. ψ_t je regresní vektor externích veličin, např. “počasí” s hodnotami sucho, mokro, námraza nebo “den v týdnu” s hodnotami všední den, víkend. Parametry modelu jsou prvky vektoru Θ . Tyto parametry se odhadují na základě změřené množiny dat $d(t) = \{y_\tau, \psi_\tau\}_{\tau=1}^t$, kde y_τ jsou skaláry a $\psi_\tau = [1, x_{1;\tau}, x_{2;\tau}, \dots, x_{n;\tau}]'$ je sloupcový vektor hodnot nezávislé proměnné. Jednička na začátku je přidána pro odhad konstanty modelu.

Lze sestavit i algoritmus pro odhad logistické regrese s více hodnotami y_t . Ten je ale složitější a my se s ním zde nebudeme zabývat.

Odhad logistického modelu

Tato úloha nemá reprodukovatelnou statistiku. Odhad se provádí jednorázově metodou ML (maximum likelihood) pro celý shromážděný datový vzorek. Za tímto účelem konstruujeme logaritmus věrohodnostní funkce

$$\ln L(\Theta) = \ln \prod_{\tau=1}^t f(y_\tau | \psi_\tau, \Theta)$$

jako součin modelů (2.2), kde platí (2.3). Po dosazení a drobných úpravách dostaneme

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

kde $z_t = \psi'_t \Theta$.

Bodové odhady leží v maximu logaritmu věrohodnostní funkce. Toto maximum lze výhodně nalézt Newtonovou metodou, protože jak gradient (první derivace), tak i Hessovu matici (druhá derivace) je možno vyjádřit v analytickém tvaru. Odvození je uvedeno v Příloze 10.11.

Předpověď výstupu

Logistickou regresi nejspíše děláme proto, abychom byli schopni pro daný regresní vektor odhadnout (pro jednotnost s předchozím výkladem říkáme předpovědět) hodnotu odpovídajícího výstupu.

Toho lze dosáhnout tak, že bodové odhady $\hat{\Theta}_t$ dosadíme do modelu (2.2), (2.3) a pro libovolný regresní vektor ψ , dostáváme odhad⁵ pravděpodobnosti hodnot y

$$f(y | \psi, \hat{\Theta}_t) = \frac{\exp(y \psi' \hat{\Theta}_t)}{1 + \exp(\psi' \hat{\Theta}_t)}.$$

Bodovou predikci \hat{y} pro regresní vektor ψ určíme jako podmíněnou střední hodnotu⁶

$$\hat{y} = E[y | \psi, d(t)] = \sum_{y=0}^1 y f(y | \psi, \hat{\Theta}_t) = \frac{\exp(\psi' \hat{\Theta}_t)}{1 + \exp(\psi' \hat{\Theta}_t)}.$$

Pokud požadujeme bodovou predikci pouze v přípustných hodnotách $\{0, 1\}$, získanou hodnotu \hat{y} zaokrouhlíme (pro $\hat{y} < 0.5$ na hodnotu 0 a pro $\hat{y} \geq 0.5$ na hodnotu 1).

Příklad [odhad logistické regrese 1]

Uvažujme diskretní veličinu y závislou na dvou veličinách regresního vektoru $\psi = [\psi_1, \psi_2]$. Tyto veličiny modelujeme jako normální náhodné veličiny $\psi_1 \sim N(-0.1, 0.25)$ a $\psi_2 \sim N(0.3, 1)$. Hodnoty y jsou přiřazeny následovně

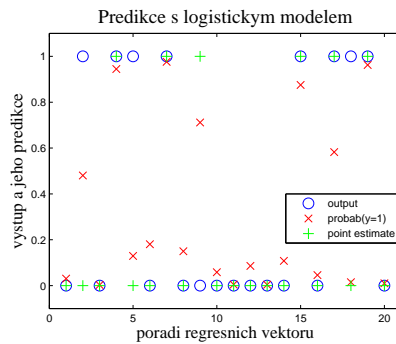
$$y = \begin{cases} 1 & \text{pro } 2\psi_1 - \psi_2 + e > 1 \\ 0 & \text{jinde} \end{cases}, \quad (4.7)$$

⁵Píšeme odhad, protože správné pravděpodobnosti bychom dostali nikoli dosazením bodových odhadů, ale násobením posteriorní hp a integrací před parametry Θ . To je ale příliš složité.

⁶Z následujícího vzorce je patrné, že platí $\hat{y} = f(y | \psi, \hat{\Theta}_t) = P(y = 1 | \psi, \hat{\Theta}_t)$.

kde $e \sim N(0, 1)$ je šum. Máme provést logistickou regresi těchto dat.

Řešení je dáno v programu T23LogRegM.m, který je možno nalézt v kapitole ?? Programy na str. ?? . Pro odhad byl použit vzorek 50 dat. Výsledek demonstrováný na dvaceti následujících (testovacích) datech je na obrázku.



Z obrázku je vidět, že bodové odhady (+) většinou správně sedí na hodnotách výstupu (o), ale pravděpodobnosti predikcí (x) vykazují určitý stupeň neurčitosti. Ta je dána použitým heuristickým generátorem dat a šumem e , který jsme do dat přičetli při určování hodnot výstupu (4.7).

Příklad [odhad s logistickým modelem 2]

Uvažujme systém s výstupem $y \in \{1, 2\}$ a dalšími veličinami, které tvoří regresní vektor $\psi = [\psi_1, \psi_2, \psi_3]$, $\psi_i \in \{1, 2\}$. Dále jsme změřili data $d(t) = d(5)$, která jsou uvedena v následující tabulce.

data	ψ	y
1	[2, 2, 2]	1
2	[1, 2, 2]	2
3	[1, 1, 1]	1
4	[1, 1, 1]	1
5	[1, 1, 1]	2

Data jsou pod vlivem šumu. Je vidět, že poslední tři datové položky jsou ve sporu. Třetí a čtvrtá říká, že regresnímu vektoru [1, 1, 1] odpovídá hodnota jedna, zatímco u páté položky je to dvojka.

Cílem příkladu je provést odhad koeficientů logistické regrese, určit predikce pro y a ukázat, jak je provedena klasifikace.

Pro odhad a predikci je opět možno použít program z m-souboru T23LogRegM.m, který je uveden na str. ?? . Odtud je možno získat výsledky:

Rovnice logistické regrese

$$\text{logit}(y) = b_0 + b_1\psi_1 + b_2\psi_2 + b_3\psi_3 + b_4\psi_4 \quad (4.8)$$

Parametry regrese

b_0	b_1	b_2	b_3
29.28	-63.22	16.49	16.76

V následující tabulce jsou uvedeny měřené hodnoty regresních vektorů ψ_t , výstupu y_t , predikované hodnoty výstupu y_t - pravděpodobnosti $P(y_t = 1)$ a bodové odhady \hat{y}_t (zaokrouhlené hodnoty pravděpodobnosti).

t	ψ	y_t	$P(y_t = 1)$	\hat{y}_t
1	[2, 2, 2]	1	0	1
2	[1, 2, 2]	2	1	2
3	[1, 1, 1]	1	0.33	1
4	[1, 1, 1]	1	0.33	1
5	[1, 1, 1]	2	0.33	1

Z tabulky je patrné, že predikce pro hodnoty regresního vektoru, při kterých se vyskytovaly sporné hodnoty výstupu, nemají pravděpodobnost nula nebo jedna. Tyto predikce říkají, že jejich hodnoty jsou nejisté, ale přiklánějí se k hodnotě nula, protože nula odpovídala dvěma případům dat, zatímco jednička pouze jednomu. Ostatní predikce jsou správné a přesné.

Predikce

5 Předpověď modelované veličiny

Jestliže sledujeme vývoj modelované veličiny v čase a popisujeme ji pomocí dynamického modelu, můžeme se v čase t zajímat o její předpověď v čase $t + k$, tedy odhadovat její hodnotu y_{t+k} . Úplný popis předpovědi můžeme vyjádřit opět pomocí podmíněné hp

$$f(y_{t+k}|y(t)). \quad (5.1)$$

Částečný popis předpovědi, tedy její bodový odhad (viz Přílohy 10.8), určuje podmíněná střední hodnota

$$E[y_{t+k}|y(t)] = \int_{y^*} y_{t+k} f(y_{t+k}|y(t)) dy_{t+k}. \quad (5.2)$$

Výpočet předpovědi se liší podle (i) počtu kroků v předpovědi (jednokroková, vícekroková), (ii) parametrů modelu (známé, neznámé), (iii) formě odhadu (bayesovský odhad, odhad s bodovým odhadem parametrů, bodový odhad výstupu), (iv) typu modelu (diskrétní, spojitý).

Data pro předpověď: Pro jednoduchost budeme v této kapitole uvažovat data tvořená pouze výstupem soustavy. Nebudeme uvažovat ani řízení, ani žádnou další externí veličinu.

Časové značení: Dále budeme uvažovat situaci, kdy jsme v čase t , tj. změřili jsme hodnotu y_t , případně jsme přepočítali statistiky odhadu parametrů a máme aposteriorní hp $f(\Theta|y(t))$, případně jsme spočítali bodové odhady parametrů a nyní se zajímáme o hodnotu výstupu k kroků dopředu.

5.1 Jednokroková předpověď

Pro $k = 1$ dostáváme jednokrokovou předpověď. O té jsme již mluvili dříve (viz odstavec 3.1) jako o odhadu výstupu. Pokud máme **model systému se známými parametry** Θ , je situace triviální. Předpověď výstupu je dána přímo modelem $f(y_{t+1}|y(t))$ nebo jeho střední hodnotou $\hat{y}_{t+1} = E[y_{t+1}|y(t)] = \int_{y^*} y_{t+1} f(y_{t+1}|y(t)) dy_{t+1}$.

Příklad [jednokroková predikce s regresním modelem]

Pro regresní model $y_t = a_1 y_{t-1} + a_2 y_{t-2} + e_t$, kde $e_t \sim N(0, r)$ je jednokroková prediktivní hp dána hustotou modelu

$$f(y_{t+1}|y(t)) = \frac{1}{\sqrt{2\pi r}} \exp\left\{-\frac{1}{2r}(y_{t+1} - a_1 y_t - a_2 y_{t-1})^2\right\}$$

a bodová předpověď je

$$E[y_{t+1}|y(t)] = a_1 y_t + a_2 y_{t-1}.$$

Obecně totiž pro model $y_t = \psi_t' \theta + e_t$ platí:

$$E [y_t | \psi_t, \Theta] = E [\psi_t' \theta + e_t] = E [\psi_t' \theta] + E [e_t] = E [\psi_t' \theta] \quad (5.3)$$

$$D [y_t | \psi_t, \Theta] = D [\psi_t' \theta + e_t] = D [e_t] = r, \quad (5.4)$$

protože $\psi_t' \theta$ je konstanta⁷.

Pokud **parametry modelu neznáme**, musíme postupovat tak, jak jsme uvedli v kapitole o odhadování ve vztahu (3.3)

$$f (y_{t+1} | y (t)) = \int_{\Theta^*} f (y_{t+1} | \psi_t, \Theta) f (\Theta | y (t)) d\Theta, \quad (5.5)$$

kde jsme doplnili a ihned zase vy-integrovali parametry Θ a rozvinuli sdruženou hp podle řetězového pravidla⁸.

Hustota pravděpodobnosti predikce, tj. prediktivní hp, je dána integrálem ze součinu hp modelu a hp popisující parametry, tj. aposteriorní hp. S aposteriorní hp jsme se již setkali při odhadu a víme, že tuto hp musíme postupně vyvíjet podle Bayesova vzorce počínaje od apriorní hp. Predikce modelované veličiny s modelem s neznámými parametry v sobě tedy obsahuje implicitně obě úlohy: odhad parametrů a předpověď výstupu.

Poznámka

Jednokroková předpověď výstupu (nebo jiné modelované veličiny) je velice důležitá. Řada úloh odhadu i řízení je založena na výpočtu chyby predikce $\hat{e}_t = y_t - \hat{y}_t$ jako rozdílu změřené veličiny a její predikce. Další akce se provádí tak, aby se tato chyba zmenšovala. ◁

5.2 Vícekroková předpověď

Pro $k > 1$ v (5.1) hovoříme o vícekové předpovědi. Při jejím odvození budeme postupovat podobně jako při konstrukci jednokrokové predikce s modelem s neznámými parametry v (5.5). Postup budeme ilustrovat na příkladě **dvou-krokové předpovědi**

$$\begin{aligned} f (y_{t+2} | y (t)) &= \int_{\Theta^*} \int_{y_{t+1}^*} f (y_{t+2}, y_{t+1}, \Theta | y (t)) dy_{t+1} d\Theta = \\ &\quad \text{(rozklad podle řetězového pravidla)} \\ &= \int_{\Theta^*} \int_{y_{t+1}^*} f (y_{t+2} | y_{t+1} y (t), \Theta) f (y_{t+1} | y (t), \Theta) f (\Theta | y (t)) dy_{t+1} d\Theta = \\ &\quad \text{(podmínky nezávislosti)} \\ &= \int_{\Theta^*} \int_{y_{t+1}^*} f (y_{t+2} | \psi_{t+2}, \Theta) f (y_{t+1} | \psi_{t+1}, \Theta) dy_{t+1} f (\Theta | y (t)) d\Theta \end{aligned} \quad (5.6)$$

⁷Protože jak ψ_t , tak i Θ jsou v podmínce.

⁸V podmínce modelu se místo celé historie uplatní jen regresní vektor ψ_t .

Označíme-li

$$f(y_{t+2}|y(t), \Theta) = \int_{y_{t+1}^*} f(y_{t+2}|\psi_{t+2}, \Theta) f(y_{t+1}|\psi_{t+1}, \Theta) dy_{t+1}$$

jako prediktor se známými parametry modelu, pak prediktivní hp (5.6) lze zapsat podobně jako v případě jednokrokové předpovědi (5.5)

$$f(y_{t+2}|y(t)) = \int_{\Theta^*} f(y_{t+2}|y(t), \Theta) f(\Theta|y(t)) d\Theta.$$

Obecná konstrukce prediktivní hp je již jednoduchým zobecněním uvedeného příkladu. Pro k -krokovou předpověď platí

$$f(y_{t+k}|y(t)) = \int_{\Theta^*} f(y_{t+k}|y(t), \Theta) f(\Theta|y(t)) d\Theta, \quad (5.7)$$

kde

$$f(y_{t+k}|y(t), \Theta) = \int_{y_{t+k-1}} \cdots \int_{y_{t+1}} f(y_{t+k}|y(t+k-1)) \cdots f(y_{t+1}|y(t)) dy_{t+k-1} \cdots dy_{t+1} \quad (5.8)$$

je k -krokový prediktor s modelem se známými parametry.

Kritickým místem při výpočtu k -krokové předpovědi s modelem s neznámými parametry je integrál v rovnici (5.7). Prediktor daný vztahem (5.8) je dán konvolucemi modelu a často je možno jej vyjádřit analyticky. Problém je ale ve výpočtu integrálu (5.7), protože aposteriorní hp je složitá funkce a její konvoluce s prediktorem je většinou analyticky nespočitatelná. Jedním ze způsobů, jak tuto potíž řešit, může být místo aposteriorní hp uvažovat jen bodové odhady parametrů (zanedbá se neurčitost v parametrech a číselné odhady se berou jako přesné). Pro bodové odhady parametrů $\hat{\Theta}_t$ je možno formálně psát hp

$$f(\Theta|d(t)) = \delta(\Theta - \hat{\Theta}_t)$$

ve tvaru Diracovy funkce: $\delta(0)$ je jedna jinak je rovna nule.

Dosadíme do (5.7) a dostaneme

$$f(y_{t+k}|y(t)) = \int_{\Theta^*} f(y_{t+k}|y(t), \Theta) \delta(\Theta - \hat{\Theta}_t) d\Theta = f(y_{t+k}|y(t), \hat{\Theta}_t). \quad (5.9)$$

To znamená, že pro výpočet předpovědi stačí vzít prediktor (5.8) a na místo neznámých parametrů dosadit jejich bodové odhady.

5.3 Vícekroková bodová předpověď

Výpočet vícekové předpovědi výstupu s modelem s neznámými parametry je úloha tak složitá, že ji v praxi nelze použít. Ukážeme zde proto na příkladech některé jednodušší případy předpovědi a její řešení.

Předpověď se spojitým modelem

K významnému zjednodušení úlohy vícekové předpovědi výstupu systému dojde, jestliže přejdeme k bodovým odhadům, a to jak pro neznámé parametry modelu (5.9), tak i pro vyjádření předpovědi ve tvaru hodnoty \hat{y}_{t+k} . Bodovou předpověď výstupu pro lineární normální regresní model budeme demonstrovat v následujícím příkladě.

Příklad [bodová předpověď]

Uvažujme tříkrokovou bodovou predikci s modelem $y_{t+1} = a_0 y_t + a_1 y_{t-1} + e_t$ s počáteční podmínkou $y_t = \xi_0$ a $y_{t-1} = \xi_{-1}$.

Dostáváme:

První krok

$$\hat{y}_{t+1} = a_0 y_t + a_1 y_{t-1} = a_0 \xi_0 + a_1 \xi_{-1},$$

Druhý krok

$$\begin{aligned}\hat{y}_{t+2} &= a_0 y_{t+1} + a_1 y_t = a_0 \hat{y}_{t+1} + a_1 \xi_0 = \\ &= a_0 (a_0 \xi_0 + a_1 \xi_{-1}) + a_1 \xi_0 = (a_0^2 + a_1) \xi_0 + a_0 a_1 \xi_{-1},\end{aligned}$$

Třetí krok

$$\begin{aligned}\hat{y}_{t+3} &= a_0 y_{t+2} + a_1 y_{t+1} = a_0 \hat{y}_{t+2} + a_1 \hat{y}_{t+1} = \\ &= a_0 ((a_0^2 + a_1) \xi_0 + a_0 a_1 \xi_{-1}) + a_1 (a_0 \xi_0 + a_1 \xi_{-1}) = \\ &= (a_0^3 + 2a_0 a_1) \xi_0 + (a_0^2 a_1 + a_1^2) \xi_{-1}.\end{aligned}$$

Předpověď s normálním modelem, kterou jsme se zabývali v předchozím příkladě, lze zobecnit tím, že budeme počítat nikoli bodový odhad předpovědi, ale celo-prediktivní hp (samozřejmě pro známé parametry nebo parametry dosažené jako bodové odhady). Využijeme toho, že konvoluce dvou normálních rozdělání, které při výpočtu předpovědi vznikají, mají opět normální rozdělání. Není proto potřeba počítat celou prediktivní hp (integrály), ale napočítávat jen střední hodnoty a rozptyly. Konečná prediktivní hp je jimi plně určena. Postup výpočtu je následující: místo střední hodnoty (což je deterministická část regresního modelu) dosazujeme rovnici modelu i se šumem. Z výsledné hodnoty y_{t+k} pak určíme hledanou střední hodnotu a rozptyl. Situaci ilustruje následující příklad.

Příklad [předpověď pro normální rozdělání]

Určete dvou-krokovou predikci s modelem $y_{t+1} = a y_t + e_{t+1}$.

$$y_{t+2} = a y_{t+1} + e_{t+2} = a (a y_t + e_{t+1}) + e_{t+2} = a^2 y_t + e_{t+2} + a e_{t+1}.$$

Protože střední hodnota konstanty je konstanta a střední hodnota šumu je nula, je

$$E[y_{t+2}|y(t)] = a^2 y_t + E[e_{t+2} + a e_{t+1}] = a^2 y_t.$$

Protože platí $D[e_{t+2}] = D[e_{t+1}] = r$ a při vytknutí konstanty z rozptylu se tato konstanta umocní na druhou, bude

$$D[y_{t+2}|y(t)] = D[e_{t+2} + a e_{t+1}] = (1 + a^2) r.$$

Hp předpovědi má tvar

$$f(y_{t+2}|y(t)) = \mathcal{N}_{y_{t+2}}(a^2 y_t, (1 + a^2) r).$$

Předpověď s diskretním modelem

Tak jako ve všech úlohách je i zde práce s diskretním modelem jednoduchá (nemusí se řešit žádné integrály a pouze se násobí a sčítají prvky tabulek). Operace s tabulkami jsou však poněkud nezvyklé a mohou působit problémy. Tuto práci budeme demonstrovat v následujícím příkladě. Tento příklad byl vybrán jako autoregrese prvního řádu, což je případ obzvláště jednoduchý. V obecnějším případě (např. pro řízený model) je práce s tabulkami mnohem náročnější.

Příklad [předpověď s diskretním modelem]

Predikci s diskretním modelem ukážeme na tom nejjednodušším příkladu, tj. s neřízeným modelem prvního řádu $f(y_t|y_{t-1})$ zadaným následující tabulkou.

$f(y_t y_{t-1})$	y_t	
	1	2
y_{t-1}		
1	$\Theta_{1 1}$	$\Theta_{2 1}$
2	$\Theta_{1 2}$	$\Theta_{2 2}$

Budeme chtít vyjádřit obecně k -krokovou predikci $f(y_k|y_0)$ s počáteční podmínkou $y_0 = 1$. V tomto jednoduchém případě lze výsledek vyjádřit analyticky, takže si můžeme dovolit uvažovat parametry modelu obecně.

Pro lepší přehlednost odvodíme předpověď pro $k = 3$ a potom jednoduše zobecníme.

Rozkladem prediktivní pf (podle (5.6), pro $k = 3$) dostaneme

$$f(y_3|y_0) = \sum_{y_2=1}^2 f(y_3|y_2) \underbrace{\sum_{y_1=1}^2 f(y_2|y_1) f(y_1|y_0)}_{f(y_2|y_0)}. \quad (5.10)$$

Nejprve si všimněme dvou-krokové předpovědi $f(y_2|y_0)$ označené svorkou. Jedná se o součet součinů prvků tabulky (modelu). Tabulku, která zadává model $f(y_t|y_{t-1})$, označíme M s prvky $m_{\tau,t}$ (všimněte si, že indexy u prvků tabulky jsou přehozené - první index jde svisle, druhý vodorovně - značení vychází ze zápisu podmíněné pravděpodobnosti). Dosadíme do předchozího výrazu pro předpověď a dostaneme

$$f(y_2|y_0) = \sum_{y_1=1}^2 m_{y_1,y_2} m_{y_0,y_1} = \sum_{y_1=1}^2 m'_{y_2,y_1} m'_{y_1,y_0} = M' M' = (M')^2,$$

kde apostrof značí transpozici a $(M')^2$ je kvadrát transponované matice. Součet součinů prvků tabulek modelu definuje součin matic.

Označíme-li tabulku pf předpovědi $f(y_2|y_0)$ jako M_2 bude platit

$$M'_2 = M' M' \text{ a tedy } M_2 = M^2.$$

Dosadíme do tří-krokové předpovědi (5.10) a s využitím vztahu $M_2 = M^2$ dostaneme

$$f(y_3|y_0) = \sum_{y_2=1}^2 f(y_3|y_2) f(y_2|y_0) = M' M'_2 = (M_2 M)' = (M^3)'. \quad (5.11)$$

Jestliže tabulku tří-krokové předpovědi označíme M'_3 bude platit

$$M_3 = M'^3.$$

Pro větší přehlednost zavedeme model číselně: $\Theta_{1|1} = 0.3$, $\Theta_{2|1} = 1 - \Theta_{1|1} = 0.7$ a $\Theta_{1|2} = 0.6$, $\Theta_{2|2} = 1 - \Theta_{1|2} = 0.4$. Tedy tabulka M' bude

$$M' = \begin{bmatrix} 0.3 & 0.7 \\ 0.6 & 0.4 \end{bmatrix}$$

a příslušná matice M (transponovaná tabulka) je

$$M = \begin{bmatrix} 0.3 & 0.6 \\ 0.7 & 0.4 \end{bmatrix}$$

Tříkroková předpověď $f(y_3|y_0)$ je reprezentována maticí

$$M_3 = M^3 = \begin{bmatrix} 0.3 & 0.6 \\ 0.7 & 0.4 \end{bmatrix} \begin{bmatrix} 0.3 & 0.6 \\ 0.7 & 0.4 \end{bmatrix} \begin{bmatrix} 0.3 & 0.6 \\ 0.7 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.51 & 0.42 \\ 0.49 & 0.58 \end{bmatrix} \begin{bmatrix} 0.3 & 0.6 \\ 0.7 & 0.4 \end{bmatrix} = \begin{bmatrix} 0.447 & 0.474 \\ 0.553 & 0.526 \end{bmatrix}.$$

Tabulka předpovědi bude dána transpozicí této matice, tedy

$f(y_3 y_0)$		
y_0	$y_3 = 1$	$y_3 = 2$
1	0.447	0.553
2	0.474	0.526

Pro obecnou k -krokovou předpověď bude platit

$$M_k = M^k$$

a tabulka této předpovědi bude transpozicí, tedy

$$f(y_k|y_0) \rightarrow (M^k)'$$

Hodnoty tabulky se po více krocích ustálí. Pro $k \geq 9$ vypadá tabulka následovně

$f(y_k y_0)$		
y_0	$y_k = 1$	$y_k = 2$
1	0.4615	0.5385
2	0.4615	0.5385

To znamená, že bez ohledu z jaké hodnoty y_0 jsme vycházeli, pro $k \geq 9$ předpovídáme $y_k = 1$ s pravděpodobností 0.4615 a $y_k = 2$ s pravděpodobností 0.5385 (což neodporuje zdravému rozumu).

Poznámka

Znovu připomeňme, že jednoduchost uvedeného příkladu je dána tím, že jsme mohli využít maticového počtu. V ostatních případech to jednoduše nelze a počítání je mnohem komplikovanější. ◁

Filtrace

6 Stav, stavový model a odhad stavu

6.1 Stav

Stav x_t je veličina systému, kterou nelze měřit a která se (na rozdíl od parametrů) v mění čase. Aby veličina mohla být stavem, musí platit, že v každém časovém okamžiku t v sobě zahrnuje veškerou informaci ze svého minulého vývoje. Pokud známe jeho hodnotu a aktuální řízení, jsme schopni spočítat distribuci jeho budoucí hodnoty.

Příklad [pojem stavu]

Uvažujme systém spojený s městskou řízenou čtyř-ramennou křižovatkou. V ramenech křižovatky měříme intenzitu provozu a zaznamenáváme poměry zelené na semaforu. Zajímáme se o délky kolon v křižovatce. Tato veličina „vektor délek kolon v křižovatce“ je v této úloze stavem. Nelze ji přímo měřit a máme-li délku kolony a víme-li, jaké řízení bude použito, můžeme počítat budoucí délku kolony. Přičteme všechna auta, co přijela do kolony před semaforem, a odečteme všechna auta, co odjela na zelenou. Různé délky aut a další zaokrouhlení spojená s periodou vzorkování zahrneme do šumu. Tento šum vnáší do výpočtů neurčitost.

6.2 Stavový model

Model stavové veličiny je reprezentován dvěma vztahy: **modelem stavu** a **modelem výstupu**. Pro oba modely se uvažuje normální rozdělení

– model dynamiky stavu

$$f(x_{t+1}|x_t, u_t) = \mathcal{N}_{x_{t+1}}(Mx_t + Nu_t, r_x), \quad (6.1)$$

– model měření výstupu

$$f(y_t|x_t, u_t) = \mathcal{N}_{y_t}(Ax_t + Bu_t, r_y), \quad (6.2)$$

kde \mathcal{N}_x a \mathcal{N}_y značí normální rozdělení.

Oba modely lze vyjádřit jako rovnice

$$x_{t+1} = Mx_t + Nu_t + w_t \quad (6.3)$$

$$y_t = Ax_t + Bu_t + v_t \quad (6.4)$$

kde $w_t \sim \mathcal{N}_w(0, r_x)$ a $v_t \sim \mathcal{N}_v(0, r_y)$.

6.3 Odhad stavu

Cílem je odhad stavu x_t pro $t = 1, 2, \dots$. Odhad stavu je podobně jako odhad parametrů dán hp stavu za podmínky změřených dat $f(x_t|d(t))$. V úloze rekurzivního odhadu parametrů (3.1) existoval jen posun času při vzorkování dat. Při odhadu stavu existují posuny dva:

– posun času v datech - tzv. **filtrace**, při které se informace z nově změřených dat vkládá prostřednictvím modelu výstupu do odhadu stavu; jde o přepočítání $f(x_t|d(t-1)) \rightarrow f(x_t|d(t))$ a

– posun času ve stavu - tzv. **predikce**, při které se stav predikuje podle modelu dynamiky stavu “naslepo”, aniž by se měřila nová data; jde o přepočítání $f(x_t|d(t)) \rightarrow f(x_{t+1}|d(t))$.

Uvedené přepočty hp jsou následující

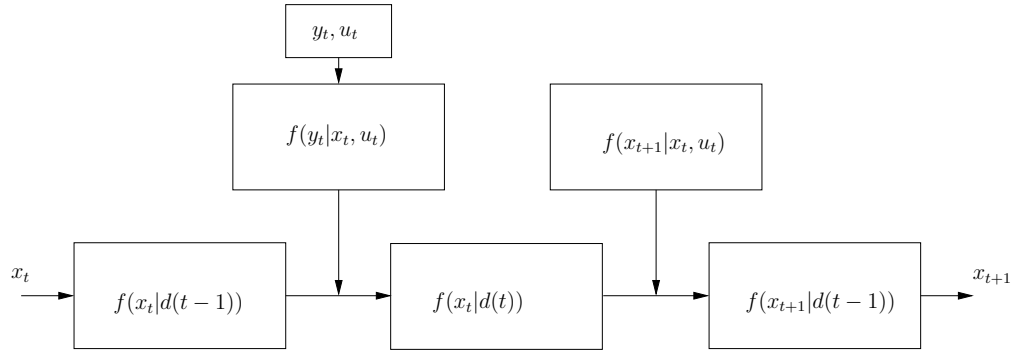
$$f(x_t|d(t)) = \frac{f(y_t|x_t, u_t) f(x_t|d(t-1))}{f(y_t|u_t, d(t-1))} \quad (6.5)$$

$$f(x_{t+1}|d(t)) = \int_{x_t^*} f(x_{t+1}|x_t, u_t) f(x_t|d(t)) dx_t, \quad (6.6)$$

kde

- $f(x_t|d(t-1)) \sim \mathcal{N}(x_{t|t-1}, R_{t|t-1})$ je popis stavu, který vstupuje do KF a který závisí na datech do času $t-1$.
 $x_{t|t-1}, R_{t|t-1}$ jsou charakteristiky tohoto popisu - střední hodnota a kovarianční matice.
- $f(x_t|d(t)) \sim \mathcal{N}(x_{t|t}, R_{t|t})$ je popis stavu po filtraci v okamžiku, kdy je odhad stavu korigován podle aktuálně změřeného výstupu y_t ; tento odhad závisí na datech $d(t) = \{y_t, d(t-1)\}$.
Charakteristiky popisu jsou $x_{t|t}, R_{t|t}$.
- $f(x_{t+1}|d(t)) \sim \mathcal{N}(x_{t+1|t}, R_{t+1|t})$ je popis stavu po predikci, kdy podle stavové rovnice modelu předpovídáme stav do příštího časového okamžiku, aniž bychom získali novou informaci z dat. Tento odhad je tedy stále závislý na datech $d(t)$ ale platí již pro příští čas $t+1$. Tento popis stavu se stává vstupním pro odhad v příští periodě KF.
Charakteristiky popisu jsou $x_{t+1|t}, R_{t+1|t}$.
- $f(y_t|x_t, u_t)$ a $f(x_{t+1}|x_t, u_t)$ jsou hp modelu stavu (6.1) a (6.2).
- $f(y_t|u_t, d(t-1)) = \int_{x_t^*} f(y_t|x_t, u_t) f(x_t|d(t-1)) dx_t$ je predikční hp výstupu (popis výstupu, nezávislý na stavu), která zde vystupuje jen jako normalizační konstanta.

Graficky lze přepočty hp znázornit takto



Poznámka

Odvození rekurze (6.5), (6.6) v případě, kdy stavový model obsahuje člen s řízením u_t , předpokládá existenci tzv. přirozených podmínek řízení. Ty říkají, že odhad stavu a řídicí veličina jsou podmíněně nezávislé. Předpokladem je platnost vztahů

$$f(x_t|u_t, d(t-1)) = f(x_t|d(t-1)) \Leftrightarrow f(u_t|x_t, d(t-1)) = f(u_t|d(t-1)).$$

Diskuze k těmto vztahům je uvedena v Příloze 10.2. ◀

6.4 Kalmanův filtr

Vztahy pro vývoj odhadu stavu podle (6.5) a (6.6) jsou (podobně jako Bayesův vzorec) vztahy pro funkce. S těmi nelze v počítači dobře pracovat. Jestliže všechny hp vyjádříme jako normální hp, lze rekurzi počítat pro jejich charakteristiky - střední hodnoty a rozptyly (kovarianční matice). Dostáváme následující číselnou rekurzi, která je dobře známa jako **Kalmanův filtr**

Filtrace:

$$\underbrace{y_p = Ax_{t|t-1} + Bu_t}_{\text{předpověď výstupu}}$$

$$\underbrace{R_p = r_v + AR_{t|t-1}A'}_{\text{kovariance výstupu}}$$

$$\underbrace{R_{r|t} = R_{t|t-1} - R_{t|t-1}A'R_p^{-1}AR_{t|t-1}}_{\text{přepočítání kovariance stavu}}$$

$$\underbrace{K = R_{t|t}A'r_y^{-1}}_{\text{Kalmanův gain}}$$

$$\underbrace{x_{t|t} = x_{t|t-1} + K(y_t - y_p)}_{\text{datová oprava stavu}}$$

Predikce:

$$x_{t|t+1} = Mx_{t|t} + Nu_t$$

předpověď stavu

$$R_{t+1|t} = r_x + MR_{t|t}M'$$

přepočítání kovariance stavu

Kalmanův filtr je realizována v programu, který je uveden v části Programy na straně ??.
Použití tohoto programu budeme ilustrovat na příkladě.

Příklad [odhad stavu]

Uvažujme stavový model

$$\begin{bmatrix} x_{1;t+1} \\ x_{2;t+1} \end{bmatrix} = \underbrace{\begin{bmatrix} 0.9 & 0.2 \\ 0 & 1 \end{bmatrix}}_M \begin{bmatrix} x_{1;t} \\ x_{2;t} \end{bmatrix} + \underbrace{\begin{bmatrix} -1.5 \\ 1 \end{bmatrix}}_N u_t + 0.6 \underbrace{\begin{bmatrix} 1 & 0.3 \\ 0 & 1 \end{bmatrix}}_{sx} \begin{bmatrix} w_{1;t} \\ w_{2;t} \end{bmatrix},$$
$$y_t = \underbrace{[0.3, 0.7]}_A \begin{bmatrix} x_{1;t} \\ x_{2;t} \end{bmatrix} + \underbrace{0.1}_{sy} e_t,$$

kde

w_t a e_t jsou šумы s nulovou střední hodnotou a konstantními kovariančními maticemi;

M , N , A jsou matice stavového modelu,

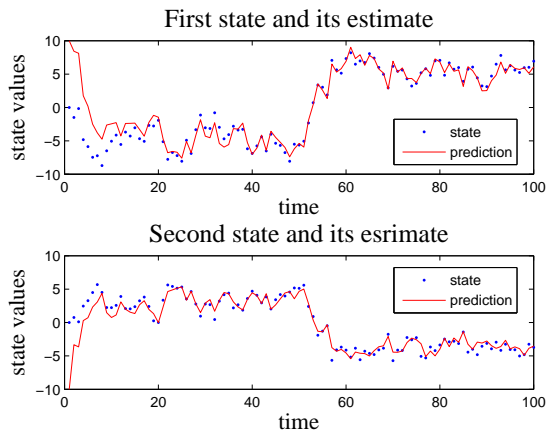
sx a sy jsou směrodatné odchylky⁹ kovariancí šumu.

- simulujte data z uvedeného stavového modelu,
- s pomocí Kalmanova filtru odhadněte simulovaný stav na daném časovém intervalu,
- simulovaný a odhadnutý stav porovnejte.

Řešení příkladu je uděláno v programu, který je uveden v části Programy, na straně ??.

Výsledky příkladu jsou na Obrázku 6.1. Tečkovaná křivka je průběh simulovaného stavu, plná čára je jeho bodový odhad. Horní graf ukazuje první složku stavu, spodní druhou.

⁹Směrodatná odchylka kovarianční matice C je definována tzv. L'DL rozkladem této matice, tj. jako horní trojúhelníková matice L' s jedničkami na diagonále, pro kterou platí $C = L'DL$, kde D je diagonální matice s nezápornými prvky na diagonále.



Obrázek 6.1: Odhad stavu pomocí Kalmanova filtru

Z grafů je patrné, že počáteční odchylka odhadu od stavu způsobená nesprávnou hodnotou počátečního odhadu rychle vymizí a odhady jsou v rámci přítomného šumu velmi přesné. Musíme mít ovšem na paměti, že stavový model má nastaveny správné parametry (parametry modelu se předpokládají známé).

◁

Poznámka k příkladu

Kalmanův filtr startuje s počátečním stavem $x_{1|0}$ a počáteční kovariancí $R_{1|0}$. Stav $x_{t|}$ a kovariance odhadu stavu $R_{t|}$ se během výpočtů přepočítávají. Matice r_x a r_y jsou kovariance šumů ze stavové a výstupní rovnice modelu. (Pozor!!! Nejsou to kovariance stavu a výstupu, ale kovariance šumů!!!). Ačkoli Kalmanův filtr pracuje se známým modelem, tyto kovariance většinou neznáme a musíme je nahradit jejich odhadem. A právě na správné nastavení těchto dvou kovariancí je další běh Kalmanova filtru velmi citlivý. Kovarianci $R_{1|0}$ lze nastavit jako diagonální s velkými čísly na diagonále (asi tak 10^3). Čím jsou tato čísla větší, tím méně důvěry vkládáme do našich počátečních odhadů stavu a tím rapidněji se tyto odhady na začátku mohou měnit. Ověřte si tuto skutečnost sami v předloženém příkladě. ◁

Další doporučené pokusy jsou:

- odhad z dat simulovaných bez a se šumem (nastaví se pomocí směrodatných odchylek r_w a r_v),
- odhad z jiné soustavy (změna parametrů - matic M, N, A),
- start odhadu s různými počátečními odhady stavu (volba $x_{1|0}$).

Klasifikace

7 Klasifikace

Klasifikace je úloha, ve které jednotlivé datové položky zařazujeme do předem daných tříd označených indexem $c = 1, 2, \dots, n_c$, kde n_c je počet tříd.

Přístupů k řešení úlohy klasifikace je celá řada. Jsou to např. algoritmy K-means, DBSCAN, COBWEB a další, shrnuté v systému Weka, volně ke stažení na

http://www.cs.waikato.ac.nz/ml/weka/index_downloading.html.

Bayesovský přístup ke klasifikaci nejčastěji využívá modely ve tvaru směsi komponent, o kterých budeme mluvit v Kapitole ???. Nicméně každý model $f(y_t|\psi_t)$ s diskretním výstupem $y_t \in \{1, 2, \dots, n_c\}$ lze využít pro klasifikaci. Hodnota předpovědi \hat{y}_t může být považována za index třídy, do které patří regresní vektor ψ_t . Model tak rozděluje prostor všech možných regresních vektorů do n_c podprostorů a každý z těchto podprostorů tvoří jednu třídu klasifikace. Využití diskretního modelu pro úlohu klasifikace ukážeme v následujícím příkladě.

Příklad [klasifikace s diskretním modelem]

Pokračujeme v příkladu 4.1.

V našem příkladě je $y_t \in \{1, 2\}$. První hodnota odhadu výstupu $\hat{y}_t = 1$ indikuje regresní vektor (vektor veličin mající vliv na vážnost nehody) z první třídy (lehké nehody) a druhá hodnota $\hat{y}_t = 2$ zařazuje okolnosti nehody do druhé (těžké nehody).

Pro výpočet odhadu výstupu máme vzorec (3.3)

$$f(y_t|\psi_t, d(t-1)) = \int_{\Theta^*} f(y_t|\psi_t, \Theta) f(\Theta|d(t-1)) d\Theta.$$

Jestliže přejdeme k bodovým odhadům parametru, tedy vezmeme aposteriorní hp jako Diracův impulz v bodovém odhadu (3.5) $f(\Theta|d(t-1)) = \delta(\Theta - \hat{\Theta}_{t-1})$, kde bodové odhady parametrů jsou prvky tabulky (4.6), platí (3.6)

$$f(y_t|\psi_t, d(t-1)) = \int_{\Theta^*} f(y_t|\psi_t, \Theta) \delta(\Theta - \hat{\Theta}_{t-1}) d\Theta = f(y_t|\psi_t, \hat{\Theta}_{t-1}).$$

Znamená to, že prediktivní hp je přímo dána bodovým odhadem výstupu s dosazeným bodovým odhadem parametrů.

V našem případě, daném tabulkou 4.6, množina všech regresních vektorů (z levé části tabulky) je rozdělena podle pravděpodobností hodnot y_t do dvou skupin: {1, 3, 4}- kde větší pravděpodobnost má $y_t = 1$ a {2, 6, 7, 8}- kde pravděpodobnější je $y_t = 0$. Regresní vektor 5 je hraniční a může být přiřazen do libovolné

skupiny. Skupiny jsou charakterizovány hodnotou předpovídaného výstupu, který jsme určili jako hodnotu s největší pravděpodobností.

Data vypovídají o tom, že lehké nehody se stávají v případech, kdy rychlost byla normální, a počasí a osvětlení byly většinou dobré. Těžké nehody nastávají většinou za velké rychlosti a při počasí a osvětlení spíše špatném.

Logistický model lze rovněž využít pro úlohu klasifikace. V odstavci 4.2 jsme ukázali, jak lze na základě odhadu logistické regrese z dat $d(t) = \{y(t), \psi(t)\}$ pro libovolně zvolený regresní vektor ψ určit jemu odpovídající predikci \hat{y} . Přitom množina $\psi(t)$ nemusí zdaleka obsahovat (v praxi také neobsahuje) všechny konfigurace hodnot regresního vektoru ψ . Odhad logistické regrese a na něm založená predikce tak provádí klasifikaci v prostoru všech konfigurací regresního vektoru ψ , tj. každý regresní vektor ψ přiřadí do jedné ze dvou skupin: první skupina regresních vektorů dává predikci 1 a druhá predikci 2.

Příklad [klasifikace s logistickou regresí]

Navazujeme na příklad 4.2.

Do odhadnuté rovnice logistické regrese (4.8) můžeme za regresní vektor postupně dosadit všechny možné hodnoty regresního vektoru a získat tak predikce i pro ty regresní vektory, které ve skutečnosti nebyly naměřeny. Tím získáme optimální odpověď na otázku: "Co bychom obdrželi, kdyby nastalo ...?". Situace je v následující tabulce

i	ψ_i	\hat{y}_i
1	[1, 1, 1]	1
2	[1, 1, 2]	2
3	[1, 2, 1]	2
4	[1, 2, 2]	2
5	[2, 1, 1]	1
6	[2, 1, 2]	1
7	[2, 2, 1]	1
8	[2, 2, 2]	1

Z této tabulky je vidět, že množina všech regresních vektorů (popisujících určitou situaci) se rozpadne na dvě množiny, vzhledem k hodnotě předpovídaného výstupu (určitého příznaku situace). První množina charakterizovaná hodnotou predikce 0 je množina regresních vektorů {1, 5, 6, 7, 8} a druhá množina obsahuje regresní vektory {2, 3, 4}. Tím jsme provedli klasifikaci regresních vektorů (situací) ψ podle hodnoty predikce výstupu (příznaku) y .

Řízení

8 Řízení se spojitým modelem

V této kapitole se stručně zmíníme o úloze optimálního řízení s popisem soustavy ve formě regresního modelu. Jestliže chceme hovořit o optimálním řízení, musíme nejprve definovat kritérium optimality. Protože pracujeme s veličinami ve formě náhodných procesů, je třeba kritérium psát jako střední hodnotu podmíněnou tím, co je nám v daném okamžiku k dispozici - apriorní znalost, tedy kritérium má tvar

$$E \left[\sum_{t=1}^N J_t | d(0) \right],$$

kde J_t je tzv. penalizační funkce, hodnotící řízení v okamžiku t . Nejčastěji se volí kvadratická penalizační funkce $J_t = y_t^2 + \omega u_t^2$, která řídí výstup na nulu a částečně pokutuje velké hodnoty řídicí veličiny.

Hodnoty tohoto kritéria na celém intervalu řízení, tj. pro $t = 1, 2, \dots, N$, nám umožní porovnávat různé způsoby řízení. Optimální řízení je takové, pro které je hodnota kritéria minimální.

Dynamické programování

Obecně lze ukázat, že optimální řídicí veličiny lze počítat rekurzivně odzadu (od konce intervalu řízení) aplikací následujících dvou kroků.

1. Výpočet střední hodnoty

$$\varphi_t = E [J_t + \varphi_{t+1}^* | u_t, d(t-1)], \quad (8.1)$$

kde J_t označuje část kritéria J odpovídající času $J_t = y_t^2 + \omega u_t^2$.

2. Minimalizace

$$\varphi_t^* = \min_{u_t} \{ \varphi_t \}, \quad (8.2)$$

přičemž u_t , pro které je dosaženo minima, je optimálním řízením v kroku t .

Rekurze startuje na konci intervalu řízení, obecně pro $t = N$ s koncovou podmínkou $\varphi_{N+1}^* = 0$.

Postup syntézy řízení ukážeme v následujícím příkladě.

Příklad

Uvažujeme následující zadání úlohy: Je dáno kvadratické kritérium optimality a model soustavy.

Kritérium

$$J = E \left[\sum_{t=1}^3 (y_t^2 + \omega u_t^2) \mid d(0) \right].$$

Model

$$y_t = ay_{t-1} + bu_t + e_t,$$

kde a, b jsou známé parametry modelu, $e_t \sim N(0, r)$ je šum se známým rozptylem r je a y_0 je známá počáteční podmínka.

Cílem je navrhnout řídicí strategii u_1, u_2, u_3 tak, aby bylo dosaženo minima kritéria J .

V tomto příkladě, kde $N = 3$, budeme postupně provádět oba zmíněné kroky. Pro $t = 3$ je $\varphi_4^* = 0$.

\Rightarrow Čas $t = 3$:

$$\begin{aligned} \varphi_3 &= E [y_3^2 + \omega u_3^2 + \varphi_4^* \mid u_3, d(2)] = E [(ay_2 + bu_3 + e_3)^2 + \omega u_3^2 \mid u_3, d(2)] = \\ &= (ay_2 + bu_3)^2 + r + \omega u_3^2 = (\omega + b^2) \left[u_3 + \frac{ab}{\omega + b^2} y_2 \right]^2 + \frac{\omega a^2}{\omega + b^2} y_2^2 + r, \end{aligned}$$

kde poslední úprava byla dosažena doplněním na čtverec v proměnné u_3 - viz ...

Minimalizace je velmi jednoduchá. Minima dosáhneme anulováním hranaté závorky (ta nemůže být menší než nula a výraz za ní je na u_3 nezávislý)

$$u_3 = -\frac{ab}{b^2 + \omega} y_2 = U_3 y_2,$$

kde $U_3 = -\frac{ab}{b^2 + \omega}$.

Zbytek po minimalizaci dostaneme, jestliže za hranatou závorku dosadíme nulu

$$\varphi_3^* = \frac{\omega a^2}{\omega + b^2} y_2^2 + r = S_3 y_2^2 + T_3,$$

kde $S_3 = \frac{\omega a^2}{\omega + b^2}$, $T_3 = r$.

\Rightarrow Čas $t = 2$:

$$\begin{aligned} \varphi_2 &= E [y_2^2 + \omega u_2^2 + \varphi_3^* \mid u_2, d(1)] = E [y_2^2 + \omega u_2^2 + S_3 y_2^2 + T_3 \mid u_2, d(1)] = \\ &\quad (\text{dosazeno za } \varphi_3^*) \\ &= E [(1 + S_3) y_2^2 + \omega u_2^2 + T_3 \mid u_2, d(1)] = \end{aligned}$$

$$\begin{aligned}
& \text{(spojeny členy s } y_2^2\text{)} \\
& = (1 + S_3) \left\{ (ay_1 + bu_2)^2 + r \right\} + \omega u_2^2 + T_3 = \\
& \quad \text{(středováno } y_2^2 \rightarrow (ay_1 + bu_2)^2 + r\text{)} \\
& = (1 + S_3) (a^2 y_1^2 + 2aby_1 u_2 + b^2 u_2^2) + (1 + S_3) r + \omega u_2^2 + T_3 = \\
& \quad \text{(umocněno)} \\
& = (1 + S_3) a^2 y_1^2 + 2(1 + S_3) aby_1 u_2 + (1 + S_3) b^2 u_2^2 + (1 + S_3) r + \omega u_2^2 + T_3 = \\
& \quad \text{(roznásobeno)} \\
& = \{ \omega + (1 + S_3) b^2 \} u_2^2 + 2(1 + S_3) aby_1 u_2 + (1 + S_3) a^2 y_1^2 + T_3 + (1 + S_3) r = \\
& \quad \text{(upraveno na kvadratický trojčlen v } u_2\text{)} \\
& = \{ \omega + (1 + S_3) b^2 \} \left[u_2^2 + 2 \frac{(1 + S_3) ab}{\omega + (1 + S_3) b^2} y_1 u_2 \right] + (1 + S_3) a^2 y_1^2 + T_3 + (1 + S_3) r = \\
& \quad \text{(vytknut koeficient } \{ \omega + (1 + S_3) b^2 \}\text{)} \\
& = \{ \omega + (1 + S_3) b^2 \} \left[u_2^2 + 2 \frac{(1 + S_3) ab}{\omega + (1 + S_3) b^2} y_1 u_2 + \left(\frac{(1 + S_3) ab}{\omega + (1 + S_3) b^2} y_1 \right)^2 \right] - \\
& \quad - \frac{(1 + S_3)^2 a^2 b^2}{\omega + (1 + S_3) b^2} y_1^2 + (1 + S_3) a^2 y_1^2 + T_3 + (1 + S_3) r = \\
& \quad \text{(připraveno na doplnění na čtverec v } u_3\text{)} \\
& = \{ \omega + (1 + S_3) b^2 \} \left[u_2^2 + \frac{(1 + S_3) ab}{\omega + (1 + S_3) b^2} y_1 \right]^2 - \\
& \quad \frac{\omega (1 + S_3) a^2}{\omega + (1 + S_3) b^2} y_1^2 + T_3 + r = S_2 y_1^2 + T_2 \\
& \quad \text{(doplněno na čtverec v } u_3\text{)}
\end{aligned}$$

Odtud plyne

$$u_2 = -\frac{(1 + S_3) ab}{\omega + (1 + S_3) b^2} y_1 = U_2 y_1,$$

kde $U_2 = -\frac{(1 + S_3) ab}{\omega + (1 + S_3) b^2}$, a

$$\varphi_2^* = S_2 y_1^2 + T_2,$$

kde $S_2 = \frac{\omega(1 + S_3)a^2}{\omega + (1 + S_3)b^2}$, $T_2 = T_3 + (1 + S_3)r$.

Zbývá dopočítat řízení pro poslední čas $t = 1$. To již ale nemusíme odvozovat. Stačí, když porovnáme výsledky z obou předchozích časů. Zjistíme, že můžeme psát obecně následující algoritmus:

Algoritmus dynamického programování pro regresní model 1. řádu

$$u_t = -\frac{(1 + S_{t+1}) ab}{\omega + (1 + S_{t+1}) b^2} y_{t-1}, \quad (8.3)$$

$$\varphi_t^* = S_t y_{t-1}^2 + T_t, \quad (8.4)$$

kde

$$S_t = \frac{\omega(1 + S_{t+1}) a^2}{\omega + (1 + S_{t+1}) b^2}, \quad T_t = T_{t+1} + (1 + S_{t+1}) r. \quad (8.5)$$

Rekurze postupuje proti času a startuje v čase N (tady $N = 3$) s hodnotou $S_{N+1} = 0$, odkud již vyjde $T_{N+1} = 1$.

Dokončíme příklad podle odvozené rekurze

\Rightarrow Čas $t = 1$:

$$u_1 = -\frac{(1 + S_2) ab}{\omega + (1 + S_2) b^2} y_0 = U_1 y_0,$$

$$\varphi_1^* = S_1 y_0^2 + T_1,$$

kde

$$S_1 = \frac{\omega(1 + S_2) a^2}{\omega + (1 + S_2) b^2}, \quad T_1 = T_2 + (1 + S_2) r$$

a S_2, T_2 byly vypočteny v kroku $t = 2$.

Tím jsme vypočítali koeficienty U_3, U_2, U_1 , ale řízení jsme zatím nemohli realizovat, protože jsme v čase $t = 0$ a hodnoty výstupu y_2 a y_1 zatím neznáme. Když jsme se dostali do času $t = 1$, závisí řízení na hodnotě y_0 , kterou už známe. Můžeme tedy začít s řízením v reálném čase. Spočteme řízení $u_1 = U_1 y_0$ a aplikujeme je. V další periodě pro $t = 2$ změříme výstup y_1 (jako odezvu na aplikované u_1) a můžeme počítat $u_2 = U_2 y_1$, které opět aplikujeme. Nakonec pro $t = 3$ změříme y_2 , vypočteme $u_3 = U_3 y_2$ a aplikujeme jej. Tím jsme splnili zadání a optimálně řídili na intervalu řízení pro $t = 1, 2, 3$.

Výsledky, které jsme odvodili, jsou obecné a platí pro libovolně dlouhý interval řízení.

Ukázka programu na toto téma je v kapitole Programy, na str. ??.

9 Řízení s diskretním modelem

Jak jsme již uvedli v Kapitole 2, obecný popis diskretního dynamického systému dostaneme tak, že každé konfiguraci hodnot jeho datového vektoru $d_t = [y_t, \psi_t]'$ přiřadíme přímo pravděpodobnost, tj.

$$f(y_t | \psi_t, \Theta) = \Theta_{y_t | \psi_t}.$$

V této kapitole budeme demonstrovat algoritmus dynamického programování s diskretním modelem. Pro větší přehlednost provedeme výklad pro konkrétní jednoduchý příklad.

Je dán model

$$f(y_t | [u_t, y_{t-1}], \Theta) = \Theta_{y_t | [u_t, y_{t-1}]},$$

kde $[u_t, y_{t-1}] = \psi_t$ a $y_t, u_t \in \{1, 2\}, \forall t$.

Kritérium pro řízení budeme definovat podobně jako u modelu pro každou konfiguraci datového vektoru zvlášť jako nezáporné číslo

$$\{J_{y_t | [u_t, y_{t-1}]} \geq 0\}_{y_t, u_t, y_{t-1} \in \{1, 2\}}.$$

Pro představu lze jak model, tak i kritérium zapsat ve formě tabulky:

Model

$$f(y_t | [u_t, y_{t-1}], \Theta)$$

u_t, y_{t-1}	$y_t = 1$	$y_t = 2$
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	$\Theta_{1 22}$	$\Theta_{2 22}$

Kritérium

$$J$$

$[u_t, y_{t-1}]$	$y_t = 1$	$y_t = 2$
1,1	$J_{1 11}$	$J_{2 11}$
1,2	$J_{1 12}$	$J_{2 12}$
2,1	$J_{1 21}$	$J_{2 21}$
2,2	$J_{1 22}$	$J_{2 22}$

Určete optimální strategii řízení u_1, u_2, u_3 , která minimalizuje zvolené kritérium řízení.

Pro řešení použijeme algoritmus (8.1) – (8.2) odvozený v minulé kapitole.

Zvolíme model a kritérium v konkrétním tvaru

$f(y_t [u_t, y_{t-1}], \Theta)$	J	
u_t, y_{t-1}	$y_t = 1$	$y_t = 2$
1,1	0.7	0.3
1,2	0.2	0.8
2,1	0.9	0.1
2,2	0.4	0.6

J	J	
u_t, y_{t-1}	$y_t = 1$	$y_t = 2$
1,1	0	1
1,2	1	0
2,1	1	2
2,2	2	1

kde je v kritériu vyjádřena skutečnost, že nechceme změny ve výstupu a pro řízení preferujeme hodnotu 1. Řízení budeme opět počítat na horizontu $N = 3$.

Pro $t = N = 3$ je podle (8.1) nutné udělat středování

$$\varphi_3(u_3, y_2) = E[J|u_3, d(2)] = \sum_{y_3=1}^2 J_{y_3|u_3 y_2} \Theta_{y_3|u_3 y_2} = J_{1|u_3 y_2} \Theta_{1|u_3 y_2} + J_{2|u_3 y_2} \Theta_{2|u_3 y_2} =$$

u_3, y_2	φ_3
1,1	$0 \times 0.7 + 1 \times 0.3 = 0.3$
1,2	$1 \times 0.2 + 0 \times 0.8 = 0.2$
2,1	$1 \times 0.9 + 2 \times 0.1 = 1.1$
2,2	$2 \times 0.4 + 1 \times 0.6 = 1.4$

a podle (8.2) minimalizace

$$\varphi_3^* = \min_{u_3} \varphi_3 = \begin{cases} 0.3 & \text{pro } y_2 = 1 \text{ (při } u_3 = 1) \\ 0.2 & \text{pro } y_2 = 2 \text{ (při } u_3 = 1) \end{cases}.$$

\mathcal{S}_3 : Optimální řízení je tedy $u_3 = 1$, a to jak pro $y_2 = 1$, tak i pro $y_2 = 2$. Tím je optimalizace v kroku 3 hotova a přecházíme na krok 2.

Pro $t = N - 1 = 2$ potřebujeme zbytek φ_3^* ve formě tabulky. Regresní vektor v kroku 2 bude $[u_2, y_1]$, φ_3^* závisí jen na y_2 (tj. bude pro všechny kombinace hodnot regresního vektoru stejné), takže tabulka bude

φ_3^*		
u_2, y_1	$y_2 = 1$	$y_2 = 2$
1,1	0.3	0.2
1,2	0.3	0.2
2,1	0.3	0.2
2,2	0.3	0.2

Prvním krokem je opět středování pře y_2

$$\begin{aligned} \varphi_2 &= E[J + \varphi_3^* | u_2, d(1)] = \sum_{y_2=1}^2 (J_{y_2 | u_2 y_1} + \varphi_3^*) \Theta_{y_2 | u_2 y_1} = \\ &= (J_{1 | u_2 y_1} + \varphi_3^*) \Theta_{1 | u_2 y_1} + (J_{2 | u_2 y_1} + \varphi_3^*) \Theta_{2 | u_2 y_1} = \end{aligned}$$

u_2, y_1	φ_2
1,1	0.8
1,2	0.7
2,1	1.6
2,2	1.9

\mathcal{S}_2 : $y_1 = 1$ je v řádku 1 a 3, v řádku 1 je menší hodnota kriteria \rightarrow pro $y_1 = 1$ volíme $u_2 = 1$. Pro $y_2 = 2$ je minimum v řádku 2, a tedy volíme také $u_2 = 1$.

$$\varphi_2^* = \begin{cases} 0.8 & \text{pro } y_2 = 1, \\ 0.7 & \text{pro } y_2 = 2. \end{cases}$$

V čase $t = 1$ je

φ_2^*		
u_1, y_0	$y_1 = 1$	$y_{21} = 2$
1,1	0.8	0.7
1,2	0.8	0.7
2,1	0.8	0.7
2,2	0.8	0.7

a

	u_1, y_0	φ_1
	1,1	1.8
=	1,2	1.7
	2,1	2.6
	2,2	2.9

\mathcal{S}_1 : odkud $u_1 = 1$ jak pro $y_0 = 1$ tak i pro $y_0 = 2$.

Strategie řízení je dána v bodech $\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3$.

Protože y_0 známe, můžeme určit u_1 , aplikovat, změřit y_1 a tak dále.

Příloha

10 Přílohy

10.1 Doplnění na čtverec

Používá se pro minimalizaci kvadratického výrazu nebo pro integraci v konvoluci dvou normálních rozdělání (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

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

Vektorový případ

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

$$x'Ax + 2x'By + y'Cy = x'Ax + 2x'AA^{-1}By + (A^{-1}By)'AA^{-1}By - (A^{-1}By)'AA^{-1}By + y'Cy =$$

$$= \underbrace{(x + A^{-1}By)' A (x + A^{-1}By)}_{\text{kvadrát}} + \underbrace{y' (C - B' A^{-1} B) y}_{\text{zbytek}}.$$

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.

10.2 Přirozené podmínky řízení

Přirozené 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 v datech $d(t-1)$ a v řízení u_t již další informace není. Proto je možno jej z podmínky vypustit

$$f(\Theta|u_t, d(t-1)) = f(\Theta|d(t-1)). \quad (10.1)$$

Obrácený vztah lze odvodit obdobnou úvahou, nebo jej lze pomocí Bayesova vzorce odvodit z (10.1)

$$\begin{aligned} f(u_t|d(t-1), \Theta) &= \underbrace{f(\Theta|u_t, d(t-1)) \frac{f(u_t|d(t-1))}{f(\Theta|d(t-1))}}_{\text{Bayes}} = \\ &= \underbrace{f(\Theta|d(t-1)) \frac{f(u_t|d(t-1))}{f(\Theta|d(t-1))}}_{\text{podle (10.1)}} = f(u_t|d(t-1)). \end{aligned}$$

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

10.3 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

$$f(B|A, C) = \frac{f(A|B, C) f(B|C)}{f(A|C)}. \quad (10.2)$$

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 (10.2) 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 (3.1) 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** (10.1), 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 $d(t-1)$.

10.4 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 $\Theta_{i|j_k}$ jsou podmíněné pravděpodobnosti (proto je také jejich index rozdělen svislínkem). Proto musí platit nezápornost $\Theta_{i|j_k} \geq 0$, $\forall i, j, k$ a dále součet parametrů pro každou konfiguraci podmínky musí dát jedničku $\sum_{i=1}^2 \Theta_{i|j_k} = 1$, $\forall j, k$.¹⁰ Pro účely odhadu je výhodné formálně tento model vyjádřit v tzv. **součinném tvaru**

$$f(y|\psi, \Theta) = \prod_{i \in y^*} \prod_{\varphi \in \psi^*} \Theta_{i|\varphi}^{\delta(i|\varphi, y|\psi)}, \quad (10.3)$$

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 $\Theta_{y|\psi}$.

10.5 Dirichletovo rozdělení

Konjugovaným rozdělením¹¹ k multinomiálnímu je rozdělení Dirichletovo

$$f(\Theta|d(t)) = \frac{1}{B(\nu_t)} \prod_{i \in y^*} \prod_{\varphi \in \psi^*} \Theta_{i|\varphi}^{\nu_{i|\varphi;t}}, \quad (10.4)$$

kde

¹⁰Tento požadavek je po úvaze opět zřejmý. Jestliže nastalo $u = j$ a $v = k$, má y právě dvě možnosti: $y = 1$ nebo $y = 2$.

¹¹Konjugované 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.

ν_t je statistika rozdělení se stejnou strukturou jako má parametr Θ - viz tabulka v odstavci 10.4 a další značení také odpovídá značení zavedenému v tomto odstavci,

$B(\nu)$ je zobecněná beta funkce

$$B(\nu) = \prod_{\varphi \in \psi^*} \frac{\prod_{i \in y^*} \Gamma(\nu_i | \varphi)}{\Gamma\left(\sum_{i \in y^*} \nu_i | \varphi\right)}, \quad (10.5)$$

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

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt, \quad (10.6)$$

pro kterou platí

$$\Gamma(x+1) = x\Gamma(x), \quad x \in \mathbb{R}^+. \quad (10.7)$$

10.6 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

$$f(y_t | \psi_t, \Theta) = \frac{1}{\sqrt{2\pi}} r^{-0.5} \exp\left\{-\frac{1}{2r} (y_t - \psi_t' \theta)^2\right\}. \quad (10.8)$$

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 (10.8) ještě upravit. Budeme postupovat následovně:

- 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 (10.8) je s uvedenou úpravou možno zapsat takto

$$f(y_t | \psi_t, \Theta) = \frac{1}{\sqrt{2\pi}} r^{-0.5} \exp \left\{ -\frac{1}{2r} [-1 \ \theta'] D_t \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}. \quad (10.9)$$

10.7 Inverzní Gauss-Wishartovo (GiW) rozdělení

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

$$f(\Theta | d(t)) \propto r^{-0.5\kappa_t} \exp \left\{ -\frac{1}{2r} [-1 \ \theta'] V_t \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}, \quad (10.10)$$

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

$$V_t = \begin{bmatrix} V_y & V_{y\psi}' \\ V_{y\psi} & V_\psi \end{bmatrix}, \quad (10.11)$$

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

$$V_t = L' D L, \quad (10.12)$$

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.

10.8 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

$$\min_{\hat{\Theta}_t} E \left[\left(\Theta - \hat{\Theta}_t \right)^2 | d(t) \right]. \quad (10.13)$$

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

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

10.9 Bodové odhady parametrů spojitého modelu

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

Hledáme tedy maximum aposteriorní hp (10.10)

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

kde jsme využili dělení matice V podle (10.11).

Nejdříve budeme hledat maximum podle θ , tj. derivovat podle θ a hledat řešení pro derivaci rovnu nule.¹²

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

Odtud pochází vztah

$$\hat{\theta} = V_\psi^{-1} V_{y\psi}. \quad (10.14)$$

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

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

a tedy

$$\Lambda = V_y - V_{y\psi}' V_\psi^{-1} V_{y\psi}. \quad (10.15)$$

¹²Derivujeme vektory podle vektorů. Správnost lze ověřit derivováním podle složek a zpětným sestavením do vektorového tvaru.

Aposteriorní hp s dosazeným bodovým odhadem (10.14) 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

$$\hat{r} = \frac{\Lambda}{\kappa}. \quad (10.16)$$

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

10.10 Bodové odhady parametrů diskrétního modelu

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

$$\hat{\Theta}_{y|\psi;t} = \frac{\nu_{y|\psi;t}}{\sum_{i \in y^*} \nu_{i|\psi;t}}, \quad \forall y \in y^* \text{ a } \psi \in \psi^*. \quad (10.17)$$

Tento bodový odhad jsme určili jako podmíněnou střední hodnotu parametru s rozdělením podle aposteriorní hp (10.4) - 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 (10.4) a beta funkce B je dána v (10.5). Parametr $\Theta_{y|\varphi;t}$ formálně vyjádříme v součinném tvaru (10.3)

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

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

$$B(\nu_\varphi) = \frac{\prod_{i \in y^*} \Gamma(\nu_{i|\varphi})}{\Gamma(\sum_{i \in y^*} \nu_{i|\varphi})} \text{ podle (10.5)}$$

jsme využili předpoklad o nezávislosti parametrů mezi různými konfiguracemi regresního vektoru ψ , tj. nezávislosti řádků v tabulkovém vyjádření parametru v odstavci 10.4.

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

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

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

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

V první úpravě předchozího výrazu jsme využili vlastnosti gama funkce (10.7). Tím jsme dokázali vztah (10.17).

10.11 Logistická regrese

Derivace věrohodnostní funkce

Derivace logaritmu věrohodnostní funkce $\ln L$ s modelem (2.2) 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 (2.3) 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

$$\frac{\partial}{\partial \Theta} p_\tau = \frac{\partial}{\partial \Theta} \frac{\exp(z_\tau)}{1 + \exp(z_\tau)} = \frac{\exp(z_\tau) \psi'_\tau (1 + \exp(z_\tau)) - \exp(z_\tau) \exp(z_\tau) \psi'_\tau}{(1 + \exp(z_\tau))^2} =$$

$$= \frac{\exp(z_\tau) \psi'_\tau}{(1 + \exp(z_\tau))^2} = \left(\frac{\exp(z_\tau)}{1 + \exp(z_\tau)} \frac{1}{1 + \exp(z_\tau)} \right) \psi'_\tau = p_\tau (1 - p_\tau) \psi'_\tau.$$

Pro hledání maxima logaritmu věrohodnostní funkce $\ln L$ je výhodné použít Newtonovu metodu¹³

Newtonův algoritmus

Pomocí Newtonova algoritmu je možné numericky hledat extrém¹⁴ 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.

10.12 Rozšířený Kalmanův filtr

Konstrukce rozšířeného KF

Máme nelineární stavový model

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

¹³Výhodné je to zejména pro to, že se nám podařilo analyticky spočítat jak první, tak i druhou derivaci maximalizované funkce $\ln L$.

¹⁴Pozor! Jde samozřejmě o lokální extrém.

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

$$\begin{aligned} g(x_t, u_t) &\doteq g(\hat{x}_t, u_t) + g'(\hat{x}_t, u_t)(x_t - \hat{x}_t) = \\ &= g'(\hat{x}_t, u_t)x_t + [g(\hat{x}_t, u_t) - g'(\hat{x}_t, u_t)\hat{x}_t], \end{aligned} \quad (10.18)$$

$$\begin{aligned} h(x_t, u_t) &\doteq h(\hat{x}_t, u_t) + h'(\hat{x}_t, u_t)(x_t - \hat{x}_t) = \\ &= h'(\hat{x}_t, u_t)x_t + [h(\hat{x}_t, u_t) - h'(\hat{x}_t, u_t)\hat{x}_t]. \end{aligned} \quad (10.19)$$

Dosažením 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

$$x_{t+1} = g'(\hat{x}_t, u_t)x_t + Q_x + w_t, \quad (10.20)$$

$$y_t = h'(\hat{x}_t, u_t)x_t + Q_y + v_t. \quad (10.21)$$

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 (10.20), (10.21). 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 (10.18) a (10.19) 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 dosažen 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

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

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.