

Ivan Nagy

**Základy bayesovského
odhadování a řízení**

Obsah

1	Matematické modely v dopravě	9
1.1	Příklad dopravní úlohy	9
1.2	Způsoby řešení dopravní úlohy	10
1.3	Typy úloh odhadování a řízení	14
2	Úvod do řízení a odhadování	17
2.1	Model systému	17
2.2	Základní úloha řízení	20
2.3	Základní úloha odhadování	23
2.4	Programy	31
3	Pravděpodobnost	37
3.1	Základní pojmy	37
3.2	Diskrétní rozdělení	40
3.3	Spojité rozdělení	42
3.4	Kombinované rozdělení	45
3.5	Operace s hustotami pravděpodobností	45
4	Model jako hustota pravděpodobnosti	53
4.1	Regresní model	54
4.1.1	Dynamický regresní model s normálním šumem	56
4.1.2	Statický regresní model s normálním šumem	57
4.1.3	Statický regresní model s rovnoměrným šumem	58
4.1.4	Statický regresní model s diskrétním šumem	59
4.1.5	Mnohorozměrný regresní model	60
4.2	Modely s diskrétními parametry	61
5	Předpovídání a odhadování	65
5.1	Úvodní poznámky a označení	65
5.2	Bodová předpověď a bodové odhady	67
5.3	Předpovídání a odhadování	69

5.4	Příklady	72
5.4.1	Regresní model obecného řádu	72
5.4.2	Statický regresní model s normálním šumem	75
5.4.3	Statický regresní model s rovnoměrným šumem	81
5.4.4	Statický regresní model s diskrétním šumem	84
5.4.5	Modely s diskrétními parametry	87
6	Řízení při známých parametrech modelu	93
6.1	Odvození v integrální formě	94
6.2	Odvození v operátorovém vyjádření	98
7	Řízení při neznámých parametrech modelu	99
7.1	Důvěřivá strategie řízení	100
7.2	Opatrná strategie řízení	102
8	Odhadování multimodelů	107
8.1	Multimodely a jejich popis	107
8.2	Odhad parametrů multimodelu	110
8.3	Příklad pro statický multimodel	114
8.4	Programový systém MixTools	116
9	Spline modely	131
9.1	Spline funkce	131
9.2	Práce se spline funkcemi	134
9.3	Spline model intenzity dopravního toku	137
9.4	Program pro odhadování spline-modelu	138
10	Příklady	143
10.1	Regresní model se střední hodnotou ve tvaru polynomu	143
10.2	Regresní model popisující exponenciálu	143
10.3	Regresní model popisující funkci sinus	144
10.4	Regresní model s rovnoměrným šumem	146
10.5	Diskrétní model	147
10.6	Spojité model s diskrétní veličinou	147
10.7	Regresní model jako filtr	151
10.8	Odhad směrových vztahů	151
10.9	Spojité řízení	156
10.10	Řízení a odhad s diskrétním modelem	157
11	Dodatky	165
11.1	Doplnění na čtverec	165

11.2 Operátor průměrování a jeho vlastnosti	165
11.3 Funkce gama a beta	166
11.4 Jednorázový výpočet hp předpovědi	167
11.5 Bodová předpověď výstupu soustavy	168
11.6 Přímý přepoččet bodových odhadů	169

Předmluva

Tato práce vznikla jako několika stránkové pojednání o základech bayesovského odhadování a řízení založeného na dynamickém programování. Byla určena pro studenty vybraných projektů katedry aplikované matematiky na Fakultě dopravní ČVUT. Postupně byla doplňována příklady a rozšířena o speciální tematiku modelů směsi distribucí a modelů, využívajících aproximaci pomocí spline funkcí. Tyto aplikace jsou zvláště vhodné pro odhadování a řízení v dopravě, neboť umožňují modelování více stavových systémů a veličin s hladkými průběhy.

Práce podává (doufejme, že) srozumitelný výklad základů bayesovského přístupu ke statistice a ukazuje jejich aplikaci na odhadování a řízení s lineárními modely. Výklad je založen na příkladech, demonstrujících konkrétní teoretické oblasti. Protože výklad má být blízký studentům, není cílem práce podat co nejobecnější teorii i s důkazy předkládaných tvrzení, ale spíše ukázat "k čemu to je" a "jak se to použije". Proto jsou výsledné algoritmy, pokud je to možné, doloženy příslušnými programy v jazyce MATLAB. Předpokládaná znalost matematiky je na úrovni standardního základního kurzu na technické vysoké škole. Speciální oblasti, které daná problematika vyžaduje, jsou podány buď přímo v textu nebo v kapitole Dodatky.

První kapitola práce je motivační. Na velmi jednoduchém dopravním příkladě ukazuje některé základní rysy modelování a odhadu tohoto modelu. Druhá kapitola ukazuje základní přístup k odhadování a řízení pro modely bez popisu pomocí hustoty pravděpodobnosti. Ukazuje se zde podstata metody nejmenších čtverců (pro odhad) a dynamického programování (pro návrh optimálního řízení). Třetí kapitola je věnována základům bayesovského pohledu na pravděpodobnost a pravděpodobnostního popisu diskrétních i spojitých náhodných veličin. Tematicky nejdůležitějšími jsou kapitoly 4 až 7. Ve čtvrté kapitole je zaveden pravděpodobnostní popis modelu vyjádřený jako podmíněná hustota pravděpodobnosti. Tento obecný popis je potom specifikován pro jednotlivé důležité varianty modelů, a to jak pro diskrétní, tak i pro spojitě modelované veličiny. Kapitola pátá pojednává o odhadu modelů z předchozí kapitoly a předpovědi modelované veličiny na základě odhadnutého modelu. Šestá a sedmá kapitola je věnována problematice řízení. Nejdříve jsou uvažovány modely se známými parametry, později se předpokládají parametry neznámé. Další část práce se zabývá speciální, avšak v dopravě velmi důležitou, problematikou. Jedná se o popis soustavy pomocí multimodelu a o modely, které se opírají o aproximaci modelované veličiny pomocí spline-funkce. V desáté kapitole je uvedena řada příkladů, které mají ilustrovat možnosti ne zcela přímočarého využití probrané teorie. V poslední kapitole jsou uvedena některá odvození, která by svou délkou narušovala plynulost výkladu v předchozích kapitolách.

I když je tato práce původně určena pro studenty a doktorandy Fakulty dopravní, může být užitečná i studentům ostatních vysokých škol nebo technickým pracovníkům, zajímajícím se o danou problematiku.

Na tomto místě bych rád poděkoval všem, kdo mi s přípravou této publikace pomohli.

Jmenovitě chci poděkovat Ing. V. Peterkovi, DrSc., který byl mým školitelem a prvním zprostředkovatelem této tematiky. Déle pak svému otci Doc. RNDr. J. Nagyovi, CSc., který byl jedním z prvních čtenářů a komentátorů. Můj největší dík však patří lektorovi Ing. M. Kárnému, DrSc., který mě mnohému naučil a svými připomínkami velmi podstatně přispěl k vylepšení této práce. Protože při svých komentářích respektoval můj styl výkladu, ačkoli sám by k němu přistoupil jinak, byl bych rád, kdyby se nedostatky této práce nespojovaly s jeho jménem.

Ivan Nagy

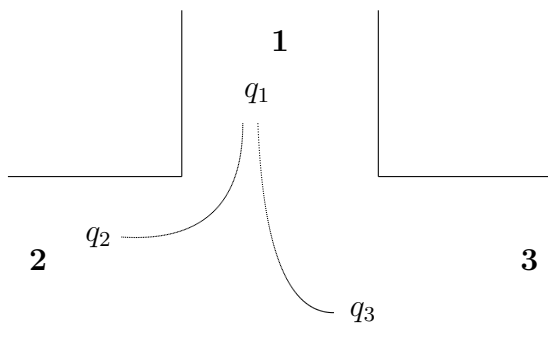
Kapitola 1

Matematické modely v dopravě

Jako motivaci pro naše pojednání o bayesovském odhadování a řízení v dopravě uvedeme jednoduchý příklad. Jeho úkolem je volně nastínit problematiku a ilustrovat základní pojmy, se kterými se dále budeme setkávat.

1.1 Příklad dopravní úlohy

Uvažujme křižovatku tvaru T, jejíž schéma je na obrázku 1.1. Automobily do ní přijíždějí



Obrázek 1.1: Schéma křižovatky

po silnici **1** a odbočují buď doprava do silnice **2**, nebo doleva do silnice **3**. Na silnicích **2** a **3** měříme **intenzitu provozu** q , tj. počet vozidel, která tudy ze silnice **1** projedou za jednu minutu. **Naším úkolem je odhadovat podíl α vozidel odbočujících doprava**, nebo jinak řečeno, určit pravděpodobnost α , že auto náhodně přijíždějící po silnici **1** odbočí doprava. Tuto pravděpodobnost patrně přesně neurčíme, ale budeme se snažit najít co nejlepší její odhad, který označíme $\hat{\alpha}$. Pravděpodobnost α považujeme za parametr této úlohy, a mluvíme proto dále o **odhadu parametrů**.

Měření provádíme každou celou minutu a časové okamžiky měření označíme přirozenými čísly $1, 2, \dots, t$. Získáme tak data (posloupnost naměřených hodnot)

$$q_{i,\tau} \text{ pro } i = 2, 3 \text{ a } \tau = 1, 2, \dots, t. \quad (1.1)$$

Nabízí se několik řešení se stejným výsledkem ale různou obecností.

1.2 Způsoby řešení dopravní úlohy

Odhad pomocí statistické definice pravděpodobnosti

Náhodným pokusem je průjezd vozidla křižovatkou. Pokus má dva možné výsledky: odbočení doprava nebo doleva. Sledovaný jev je odbočení doprava. Jeho pravděpodobnost α je při velkém počtu pokusů dána podílem počtu příznivých výsledků (počtu odbočení doprava) a počtu pokusů (celkového počtu vozidel, která křižovatkou projela). Protože počet automobilů projíždějících křižovatkou měříme průběžně v termínech intenzit, můžeme odhad poměru vozidel odbočujících doprava vyjádřit jako následující zlomek

$$\hat{\alpha} = \frac{\sum_{\tau=1}^t q_{2,\tau}}{\sum_{\tau=1}^t (q_{2,\tau} + q_{3,\tau})}. \quad (1.2)$$

Budeme-li tento výpočet opakovat pro dostatečně velký počet časových okamžiků, budou se hodnoty $\hat{\alpha}$ lišit navzájem jen nepatrně a budou zřejmě blízké skutečné hodnotě α .

♣ Vzorec (1.2) určuje dobrý odhad skutečného poměru vozidel odbočujících na křižovatce doprava.

Rekurzivní odhad pomocí statistické definice pravděpodobnosti

Abychom si nemuseli pamatovat celé posloupnosti starých dat, zavedeme statistiky (částečné součty intenzit)

$$\kappa_i(\tau) = \kappa_i(\tau - 1) + q_{i,\tau}, \quad i = 2, 3, \quad \tau = 1, 2, \dots, t, \quad (1.3)$$

s počáteční hodnotou $\kappa_i(0) = 0$, $i = 2, 3$. Pro takto zavedené statistiky platí

$$\kappa_i(t) = \sum_{\tau=1}^t q_{i,\tau} \quad (1.4)$$

a pro odhad parametru α dostaneme

$$\hat{\alpha} = \frac{\kappa_2(t)}{\kappa_2(t) + \kappa_3(t)}. \quad (1.5)$$

♣ Vzorec (1.5) představuje stejný odhad jako (1.2), ale zapsaný pomocí průběžně počítaných statistik.

P o z n á m k a: *Odhady je možno z průběžně počítaných statistik počítat také průběžně.*

P o z n á m k a: *Máme-li určitý (tzv. apriorní) odhad hodnoty parametru α , můžeme jej do statistik zabudovat tak, že nezvolíme jejich počáteční hodnoty nulové, ale takové, aby vyjadřovaly naše představy o tomto parametru:*

- poměr $\frac{\kappa_2(0)}{\kappa_2(0) + \kappa_3(0)}$ vyjadřuje náš počáteční odhad hodnoty parametru α ,

- velikost $\kappa_2(0) + \kappa_3(0)$ vyjadřuje váhu tohoto počátečního odhadu vzhledem k měřeným datům.

Cvičení: Na následujícím příkladě ověřte vliv počátečních hodnot statistiky. Provedli jsme jedno měření a obdrželi data $q_{2,1} = 2$, $q_{3,1} = 8$. Počáteční hodnoty volíme

1. $\kappa_2(0) = 1$, $\kappa_3(0) = 1$;
2. $\kappa_2(0) = 100$, $\kappa_3(0) = 100$.

Ověřte, že velké hodnoty počátečních podmínek mají tendenci se více prosadit proti prvním naměřeným datům.

Odhad pomocí deterministického modelu

Z proudu aut $q_{1,\tau} = q_{2,\tau} + q_{3,\tau}$, přicházejícího po silnici **1**, odbočí podíl α aut doprava a vytvoří proud q_2 . Platí tedy:

$$q_{2,\tau} = \alpha q_{1,\tau}, \quad \tau = 1, 2, \dots, t. \quad (1.6)$$

Je zřejmé, že takový model je idealizace. Ve skutečnosti se bude poměr aut odbočujících doprava v každém okamžiku měření vždy trochu lišit a lze hovořit pouze o jeho "ideální" hodnotě. Tento nedostatek deterministického modelu lze řešit zavedením poruchové veličiny, která popisuje okamžité odchylky od průměrné ("ideální") hodnoty. Můžeme postupovat dvojnásobem.

Deterministický model s poruchou v parametru

Uvažujeme poruchu, která působí na parametr α a způsobuje, že tento parametr je proměnný, tj. kolísá kolem své "ideální" hodnoty. Pro jednoduchost budeme v tomto případě uvažovat konstantní příjezd vozidel, tedy $q_{1,\tau} = q_1$ pro všechna τ . Model má potom tvar

$$q_{2,\tau} = \alpha_\tau q_1. \quad (1.7)$$

Odhadování parametrů provedeme takto:

Pro jednotlivé časové okamžiky můžeme psát rovnosti

$$\begin{aligned} q_{2,1} &= \alpha_1 q_1, \\ q_{2,2} &= \alpha_2 q_1, \\ \dots &\dots, \\ q_{2,t} &= \alpha_t q_1. \end{aligned} \quad (1.8)$$

Sečteme levé i pravé strany těchto rovností a dostaneme

$$\sum_{\tau=1}^t q_{2,\tau} = \sum_{\tau=1}^t \alpha_\tau q_1. \quad (1.9)$$

Pomocí (1.4) a rovnosti $t q_1 = \kappa_1(t) = \kappa_2(t) + \kappa_3(t)$ můžeme rovnost (1.9) psát ve tvaru

$$\kappa_2(t) = \sum_{\tau=1}^t q_{2,\tau} = \sum_{\tau=1}^t \frac{1}{t} \alpha_\tau t q_1 = \frac{1}{t} \sum_{\tau=1}^t \alpha_\tau (\kappa_2(t) + \kappa_3(t)).$$

Odtud dosazením do (1.5) dostáváme

$$\hat{\alpha} = \frac{\kappa_2(t)}{\kappa_2(t) + \kappa_3(t)} = \frac{1}{t} \sum_{\tau=1}^t \alpha_{\tau}. \quad (1.10)$$

Vidíme, že:

- ♣ Odhad (1.5) udává průměrnou hodnotu časově proměnného parametru α_{τ} .

Poznámka: *Popsaný postup vede k optimálnímu odhadu jen v nejjednodušších případech. Zaručuje totiž jen nulový průměr odchylek v rovnicích, nikoli nejmenší odchylky. Použité kritérium musí zaručit, aby se sčítané chyby nemohly navzájem kompenzovat. Proto se jako kritérium používá součet čtverců nebo absolutních hodnot odchylek. Součet čtverců má proti součtu absolutních hodnot výhodu, že jej lze derivovat. To je dobré pro hledání minima.*

Poznámka: *Přístup, používající proměnné parametry modelu, není obvyklý. Práce s takovým modelem je značně obtížná, neboť vede na řešení diferenčních rovnic s proměnnými koeficienty.*

Deterministický model s poruchou v datech

Uvažujeme poruchu v datech. Model (1.6) má potom tvar

$$q_{2,\tau} = \alpha q_{1,\tau} + e_{\tau}, \quad (1.11)$$

kde

e_{τ} je porucha modelující skutečnost, že některá auta místo odbočení doprava, jak by v "ideálním" případě měla, odbočí doleva,

α je parametr, který je v tomto případě konstantní.

Odhad parametru může být následující: Pro jednotlivé časové okamžiky dostáváme rovnice

$$\begin{aligned} q_{2,1} &= \alpha q_{1,1} + e_1, \\ q_{2,2} &= \alpha q_{1,2} + e_2, \\ \dots &\dots, \\ q_{2,t} &= \alpha q_{1,t} + e_t. \end{aligned} \quad (1.12)$$

Opět sečteme levé a pravé strany rovnic, využijeme (1.4) a dostáváme

$$\kappa_2(t) = \alpha[\kappa_2(t) + \kappa_3(t)] + \sum_{\tau=1}^t e_{\tau}.$$

V případě, že platí $\sum_{\tau=1}^t e_{\tau} = 0$, dostáváme vzorec (1.10), tj.

$$\hat{\alpha} = \frac{\kappa_2(t)}{\kappa_2(t) + \kappa_3(t)}. \quad (1.13)$$

Poznámka: *Vidíme, že když budou mít rovnice (1.12) v průměru nulovou chybu, bude vzorec pro odhad $\hat{\alpha}$ dán vztahem (1.13), stejně jako dříve. Tato úvaha je namístě, jestliže*

předpokládáme, že skutečně nějaká porucha e_τ , $\tau = 1, 2, \dots, t$, existuje a vstupuje do zmíněných rovnic.

Běžnější situace však je, když máme pouze naměřená data $q_{2,\tau}$ a $q_{3,\tau}$ pro $\tau = 1, 2, \dots, t$ a porucha e_τ je pouze fiktivní. Potom lze předchozí předpoklad o poruše formulovat jako požadavek: Určete odhad parametru α v rovnicích (1.12) tak, aby pro naměřená data měla porucha v průměru nulovou chybu. Tento požadavek je určitým kritériem, podle něhož neznámý parametr určíme. Výsledkem je již odvozený vzorec (1.13).

♣ Vzorec (1.13) je opět stejný jako v (1.5) a dává odhad, pro nějž mají rovnice (1.12) v průměru nulovou poruchu.

Cvičení: Ověřte, že stejný výsledek dostaneme i při použití daleko kvalitnějšího kritéria, kterým je minimum součtu čtverců poruch v jednotlivých rovnicích, tedy

$$\min_{\alpha} \sum_{i=1}^t e_i^2.$$

Návod: v rovnicích vyjádřete odchylky a zapište kritérium – součet čtverců odchylek. Jeho minimum vypočtete tak, že derivaci kritéria podle α položíte rovnu nule. Odtud vyjádřete α .

Řešení pomocí pravděpodobnostního modelu

Pravděpodobnost, že náhodně vybrané auto odbočí doprava, je α , doleva $1 - \alpha$. V časovém okamžiku τ přijede za jednu minutu $q_{1,\tau} = q_{2,\tau} + q_{3,\tau}$ aut. Pravděpodobnost, že jich $q_{2,\tau}$ odbočí doprava a $q_{1,\tau} - q_{2,\tau} = q_{3,\tau}$ doleva, je dána binomickým rozdělením

$$P(q_{2,\tau}, q_{3,\tau} | \alpha) = \binom{q_{2,\tau} + q_{3,\tau}}{q_{2,\tau}} \alpha^{q_{2,\tau}} (1 - \alpha)^{q_{3,\tau}}. \quad (1.14)$$

Tento vztah představuje pravděpodobnostní model naší křižovatky v časovém okamžiku τ .

Věrohodnostní funkce $L_t(\alpha)$: Obsahuje v sobě informaci o hledaném parametru získanou z naměřených dat – viz (5.5). Je dána součinem modelů (1.14) v jednotlivých časových okamžicích, tj.

$$L_t(\alpha) = \prod_{\tau=1}^t P(q_{2,\tau}, q_{3,\tau} | \alpha) = \prod_{\tau=1}^t \binom{q_{2,\tau} + q_{3,\tau}}{q_{2,\tau}} \alpha^{\sum_{\tau=1}^t q_{2,\tau}} (1 - \alpha)^{\sum_{\tau=1}^t q_{3,\tau}}.$$

Označíme-li $k = \prod_{\tau=1}^t \binom{q_{2,\tau} + q_{3,\tau}}{q_{2,\tau}}$ a využijeme-li označení (1.4), můžeme věrohodnostní funkci (likelihood) zapsat ve tvaru

$$L_t(\alpha) = k \alpha^{\kappa_2(t)} (1 - \alpha)^{\kappa_3(t)}, \quad (1.15)$$

kde časový vývoj statistik $\kappa_i(t)$, $i = 2, 3$, je dán vztahem (1.3).

Maximálně věrohodný odhad parametru α : Za odhad se volí nejpravděpodobnější hodnota parametru α , tj. maximalizuje se věrohodnostní funkce. Hledá se tedy α , pro které je splněna rovnost

$$\frac{dL_t(\alpha)}{d\alpha} = 0,$$

kde L_t je věrohodnostní funkce z (1.15). Tato podmínka je splněna pro

$$\hat{\alpha} = \frac{\kappa_2(t)}{\kappa_2(t) + \kappa_3(t)}. \quad (1.16)$$

♣ To je opět stejný odhad jako v (1.5).

Cvičení: Ověřte odhad anulováním derivace věrohodnostní funkce a výpočtem α .

Středně kvadratický odhad parametru α : Zvolíme-li za kritérium střední hodnotu kvadratického výrazu $(\alpha - \hat{\alpha})^2$, pomocí kterého měříme kvalitu odhadu $\hat{\alpha}$ parametru α , dostaneme optimální odhad (jak později ukážeme) jako střední hodnotu α podmíněnou všemi naměřenými daty až do času t (označíme je $q(t)$). Její hustota pravděpodobnosti $f(\alpha|q(t))$ je dána normalizovanou věrohodnostní funkcí (tj. věrohodnostní funkcí dělenou jejím integrálem přes celý definiční obor), pro kterou z výpočetních důvodů nyní zvolíme počáteční podmínku $\kappa_i(0) = -1$. Dostaneme

$$f(\alpha|q(t)) = \frac{\alpha^{\kappa_2(t)-1}(1-\alpha)^{\kappa_3(t)-1}}{\int_0^1 \alpha^{\kappa_2(t)-1}(1-\alpha)^{\kappa_3(t)-1} d\alpha}.$$

Poznámka: Počáteční podmínka $\kappa_i(0) = -1$ je prozatím zavedena formálně z výpočetních důvodů. Již teď je ale možno konstatovat, že možnost zavedení počátečních podmínek (tzv. apriorní informace) je velikou předností bayesovského pravděpodobnostního přístupu – viz druhou poznámku ke vzorci (1.3).

Pro odhad α podle (11.3) a (11.5) platí

$$\begin{aligned} \hat{\alpha} = E[\alpha|q(t)] &= \int_0^1 \alpha f(\alpha|q(t)) d\alpha = \frac{\int_0^1 \alpha \cdot \alpha^{\kappa_2(t)-1}(1-\alpha)^{\kappa_3(t)-1} d\alpha}{\int_0^1 \alpha^{\kappa_2(t)-1}(1-\alpha)^{\kappa_3(t)-1} d\alpha} = \\ &= \frac{B[\kappa_2(t) + 1, \kappa_3(t)]}{B[\kappa_2(t), \kappa_3(t)]} = \frac{\kappa_2(t)}{\kappa_2(t) + \kappa_3(t)}. \end{aligned} \quad (1.17)$$

♣ Odhad je opět stejný jako v (1.5), ale s počáteční podmínkou $\kappa_i(0) = -1$.

Cvičení: Pokuste se odvodit tvrzení, že minimalizací střední hodnoty kvadratického kritéria dostaneme optimální odhad parametru α jako střední hodnotu.

Návod: Kritérium umocněte a proveďte středování. Uvědomte si, že náhodná veličina je α , zatímco odhad $\hat{\alpha}$ je konstanta. (Úprava je podobná odvození výpočetního vzorce pro rozptyl.)

1.3 Typy úloh odhadování a řízení

V předchozích odstavcích jsme nastínili problematiku modelování a odhadu dopravního systému. Viděli jsme, že na systému měříme určité veličiny, tzv. **výstupy systému**, a v některých případech můžeme jiné veličiny sami zadávat, tzv. **vstupy systému**. Systém je popsán **modelem**. To je funkce, která umožňuje určit výstup (nebo jeho statistické charakteristiky) pomocí **parametrů modelu a starších dat**, změřených na systému). **Úkolem**

odhadování je odhadnout hodnoty parametrů (nebo jejich charakteristiky) ze změřených dat. *Úkolem řízení* je určit vstupní veličiny systému na určitý počet kroků dopředu, tzv. **horizont řízení**, tak, aby výstupní veličina měla požadovanou hodnotu.

Pro lepší orientaci nyní uvedeme základní úlohy, kterými se budeme dále zabývat.

- **Řízení pro známé parametry modelu systému**
V tomto ne příliš realistickém případě uvažujeme model, jehož parametry jsou známé. Naším úkolem je určit hodnoty vstupních veličin tak, aby výstupní veličina byla co nejbližší zadané, nejčastěji konstantní, hodnotě.
- **Odhad parametrů systému**
Zde se snažíme z naměřených dat určit parametry modelu systému tak, aby tento model co nejlépe popisoval výstup systému.
- **Předpověď výstupu pro neznámé parametry modelu**
Úkolem je nalézt takový popis systému, který by předpovídal budoucí hodnotu výstupu jen na základě starších dat, nikoliv parametrů. Takový popis lze získat spojením modelu systému s odhadem jeho parametrů.
- **Řízení výstupu pro neznámé parametry modelu**
Řídící veličina, podobně jako předpověď výstupu, nesmí záviset na neznámých parametrech modelu systému. Proto musíme syntézu řízení rovněž spojit s odhadem. Podle způsobu, jakým to provedeme, rozlišujeme důvěřivou a opatrnou strategii řízení.
 - *Důvěřivá strategie řízení*, kdy za neznámé parametry dosazujeme jejich bodové odhady.
 - *Opatrná strategie řízení*, kdy při syntéze uvažujeme popis parametrů pomocí jejich hustot pravděpodobnosti.

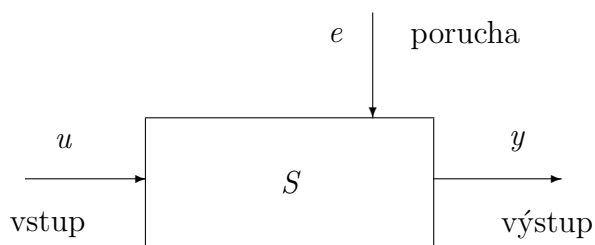
Kapitola 2

Úvod do řízení a odhadování

V této kapitole načrtneme první pohled na úlohy řízení a odhadu. Pro začátek nebudeme model systému vyjadřovat pomocí hustoty pravděpodobnosti. Budeme uvažovat deterministický popis systému, ve kterém je náhodnost modelována nějakou poruchou. Algoritmy, které při takto zjednodušeném přístupu odvodíme, budou přesto dobře ukazovat základní myšlenky algoritmů bayesovského odhadování a řízení, kterými se budeme zabývat později. Jako základní doplňkovou literaturu lze doporučit např. [1] pro řízení a [2] pro odhad metodou nejmenších čtverců.

2.1 Model systému

- ♣ Model systému je představován matematickým vztahem, který určuje rozdělení pravděpodobnosti výstupu systému na základě určitých konstant (parametrů modelu), vstupní veličiny, starších dat a rozdělení pravděpodobnosti poruchy (šumu modelu).



Obrázek 2.1: Blokové schéma modelu systému

Veličiny, které v modelu vystupují, je možno znázornit pomocí blokového schématu na obrázku 2.1. Do systému S vstupuje signál u – vstup a vystupuje signál y – výstup. Vstup zadáváme my jako operátoři, zpravidla pomocí regulátoru, který jsme určili a který pracuje se staršími vstupy a výstupy – v tzv. zpětné vazbě. Do soustavy dále vstupuje porucha e . Hodnoty této poruchy nemůžeme měřit. Jen na ni klademe nějaké předpoklady, jako např., že její dlouhodobý průměr je nula, nebo že její hodnoty jsou nezávislé na minulých datech. Vstup i výstup soustavy vzorkujeme, tj. měříme jejich hodnoty v ekvidistantních časových okamžicích. Tyto okamžiky označujeme celými čísly. Měřená data jsou tedy posloupnostmi hodnot jednotlivých signálů, jejich indexy označují okamžiky měření. V časovém okamžiku τ tedy uvažujeme tři konečné posloupnosti y_τ , u_τ , e_τ pro $\tau = 1, 2, \dots, t, \dots, N$. Údaje označené indexem τ budeme pro $\tau = t$ nazývat "současnými", pro $\tau < t$ "starými", nebo také "již naměřenými", pro $\tau > t$ "budoucími". Celkový počet časových okamžiků, pro které budeme soustavu odhadovat (a případně řídit), jsme označili N .

Regresní model řádu ν

Tento model je nejčastěji používaným popisem systému. Určuje aktuální výstup soustavy y_t jako lineární kombinaci aktuálního vstupu u_t a starších vstupů a výstupů (dat). Jako složka náhodnosti se přičítá porucha e_t . Model můžeme zapsat pomocí posloupností dat y_τ , u_τ , e_τ ve tvaru

$$y_t = b_0 u_t + \sum_{\tau=1}^{\nu} (a_\tau y_{t-\tau} + b_\tau u_{t-\tau}) + c + e_t, \quad t = 1, 2, \dots, N, \quad (2.1)$$

nebo ve vektorovém tvaru

$$y_t = \varphi_t^T \theta + e_t, \quad t = 1, 2, \dots, N, \quad (2.2)$$

kde

ν je hloubka paměti, tzv. **řád modelu**,

$a_1, a_2, \dots, a_\nu, b_0, b_1, b_2, \dots, b_\nu, c$ jsou parametry modelu (konstanty),

y_τ je vzorkovaný výstup (posloupnost čísel),

u_τ je vzorkovaný vstup (posloupnost čísel, definující po částech konstantní vstup),

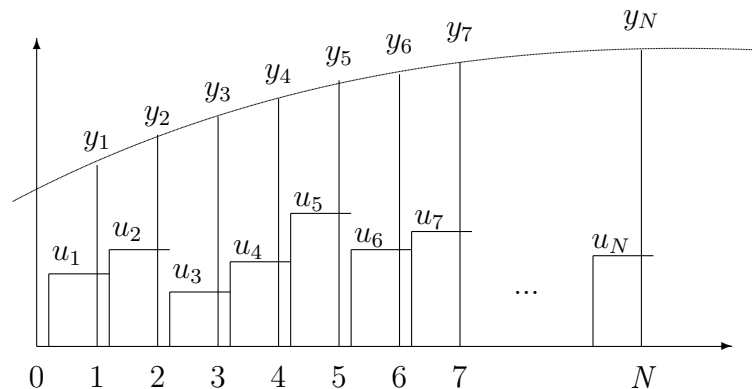
e_τ je náhodná porucha (posloupnost čísel – realizací náhodné veličiny s nulovou střední hodnotou a konstantním rozptylem σ^2),

symbol T označuje transpozici,

$\theta^T = [b_0, a_1, b_1, \dots, a_\nu, b_\nu, c]$ je **vektor regresních parametrů modelu**,

$\varphi_t^T = [u_t, y_{t-1}, u_{t-1}, \dots, y_{t-\nu}, u_{t-\nu}, 1]$ je **regresní vektor modelu**.

Příklad posloupnosti vzorkovaného vstupu a výstupu je znázorněn na obrázku 2.2.



Obrázek 2.2: Grafy výstupu a vzorkovaného vstupu

Příklad [Regresní model 1. řádu bez šumu]

Zvolíme-li v (2.1) $\nu = 1$, $b_0 = 1$, $a_1 = 0.5$, $b_1 = 0$ a poruchu $e_\tau = 0$ pro všechna τ , dostaneme následující tvar regresního modelu

$$y_t = 0.5y_{t-1} + u_t. \quad (2.3)$$

Tento model popisuje deterministický dynamický systém. Je-li $u_\tau = 0$ pro všechna τ , nazývá se model autonomní, nebo také neřízený, a popisuje dozrívání počáteční podmínky y_0 . Je-li $u_\tau \neq 0$ pro nějaké τ , pak model popisuje řízený systém. Např. pro počáteční podmínku $y_0 = 10$ a řízení: $u_1 = 1$, $u_\tau = 0$ pro $\tau > 1$ je výpočet odezvy tohoto modelu v tabulce

τ	u_τ	výpočet y	y_τ
0	—	—	10
1	1	$0.5 \times 10 + 1$	6
2	0	$0.5 \times 6 + 0$	3
3	0	$0.5 \times 3 + 0$	1.5
...

(konec příkladu)

Poznámka: Pro deterministický systém je možno při známých vstupech u_τ celou odezvu předem spočítat a řídit bez zpětné vazby. Stochastický systém je třeba řídit se zpětnou vazbou.

Příklad [Regresní model 1. řádu se šumem]

Model má tvar

$$y_t = 0.5y_{t-1} + u_t + e_t, \quad (2.4)$$

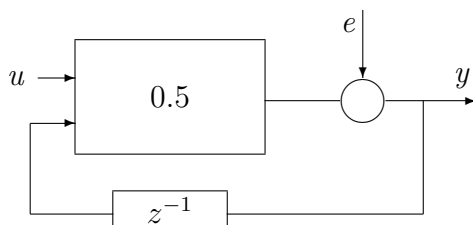
kde e_t je porucha (šum), o které předpokládáme, stejně jako v (2.2), že nijak nesouvisí se staršími daty měřeními na soustavě. Tento model je strukturou shodný s předchozím, ale navíc jsme přidali poruchu, která způsobuje narušení jinak ideálního průběhu výstupu generovaného deterministickým systémem.

Model popisuje stochastický dynamický systém. Výpočet jeho odezvy y pro daný vstupní signál u a jednu vybranou realizaci poruchy e je naznačen v následující tabulce

τ	u_τ	e_τ	výpočet y	y_τ
0	—	—	—	10
1	1	0.1	$0.5 \times 10 + 1 + 0.1$	6.1
2	0	-0.3	$0.5 \times 6.1 + 0 - 0.3$	2.75
3	0	0.2	$0.5 \times 2.75 + 0 + 0.2$	1.575
...

(konec příkladu)

Poznámka: Do každé další hodnoty výstupu se započítává i šum. Zapojení, které uvažujeme, je schematicky znázorněno na obrázku 2.3).



Obrázek 2.3: Schéma zapojení modelu se šumem

2.2 Základní úloha řízení

Budeme předpokládat, že uvažovaná soustava je popsána deterministickým regresním modelem, tj. modelem (2.1) bez šumu

$$y_t = b_0 u_t + \sum_{\tau=1}^{\nu} (a_{\tau} y_{t-\tau} + b_{\tau} u_{t-\tau}) + c,$$

a že známe potřebné počáteční podmínky

$$y_0, y_{-1}, \dots, u_0, u_{-1}, \dots$$

Cílem je sestavit plán konečných vstupů u_{τ} na horizontu řízení N tak, aby tyto vstupy minimalizovaly kritérium

$$\bar{\mathcal{J}} = \frac{1}{N} \sum_{\tau=1}^N y_{\tau}^2. \quad (2.5)$$

Úlohu budeme demonstrovat na příkladě. Při odvození syntézy, stejně jako i při výpočtech potřebných v odhadování, budeme často používat doplnění kvadratického výrazu na úplný čtverec. Tento postup je připomenut v dodatku 11.1.

Řízení s deterministickým modelem 1. řádu

Uvažujme horizont pro řízení $N = 3$, kritérium (2.5)

$$\mathcal{J} = \sum_{\tau=1}^3 y_{\tau}^2 \quad (2.6)$$

a deterministický regresním modelem 1. řádu podle (2.3), ale s obecnými koeficienty a, b

$$y_t = a y_{t-1} + b u_t. \quad (2.7)$$

Potřebnou počáteční podmínkou je y_0 .

Jednorázová minimalizace kritéria

Protože se jedná o deterministickou soustavu se známými počátečními podmínkami, lze všechny hodnoty výstupu, figurující v kritériu pro známé hodnoty vstupů, vyjádřit (předpovídat) pomocí počátečních podmínek a řízení

$$\begin{aligned} y_1 &= a y_0 + b u_1, \\ y_2 &= a y_1 + b u_2 = a(a y_0 + b u_1) + b u_2 = a^2 y_0 + a b u_1 + b u_2, \\ y_3 &= a y_2 + b u_3 = a(a^2 y_0 + a b u_1 + b u_2) + b u_3 = a^3 y_0 + a^2 b u_1 + a b u_2 + b u_3. \end{aligned}$$

Dosadíme-li tyto předpovídané hodnoty do kritéria (2.6), dostaneme \mathcal{J} jako funkci neznámých u_1, u_2, u_3 (počáteční výstup y_0 a parametry a, b předpokládáme známé). Trojice řízení, která kritérium (2.6) minimalizuje, představuje řízení optimální.

Poznámka: Jak jsme se již zmínili, pro deterministický systém je možno spočítat hodnoty optimálního řízení pro všechny časové okamžiky najednou. Není třeba čekat na postupně generované výstupy a zavádět zpětnou vazbu. Minimalizaci ale můžeme provádět také postupně.

Vzhledem k návaznosti na dynamické systémy a přirozeně vyžadovanou kauzalitu regulátoru budeme minimalizovat "odzadu" proti směru času.

Protože chceme získat řízení kauzální, kdy současný vstup závisí jen na minulých (nikoli budoucích) datech, je třeba pomocí předpovědi eliminovat budoucí výstupy, které by způsobovaly nekauzalitu akčních zásahů. Není třeba eliminovat všechny výstupy na horizontu řízení. Vypočtený akční zásah může záviset na minulých, a tedy již známých výstupech. V čase $t = 3$ generujeme vstup u_3 , kterým řídíme výstup y_3 . Tento výstup tedy pro určení u_3 nesmíme použít. V kritériu za něj dosadíme z rovnice modelu. Dostaneme

$$\mathcal{J} = y_1^2 + y_2^2 + (ay_2 + bu_3)^2.$$

Tím jsme se zbavili y_3 , které není pro výpočet u_3 k dispozici a dostali jsme kritérium jako funkci u_3 a starších dat. Jeho minimalizací podle u_3 dostaneme

$$u_3 = -\frac{a}{b}y_2,$$

a obdobně v časech $\tau = 2$ a $\tau = 1$ obdržíme u_2 a u_1 . Je tedy

$$u_\tau = -\frac{a}{b}y_{\tau-1} \quad \text{pro } \tau = 3, 2, 1.$$

Poznámka: Řízení je kauzální, jak jsme to požadovali a zajistili při jeho odvození. Pro výpočet u_τ se používá starší hodnota $y_{\tau-1}$, která je v okamžiku generování akční veličiny již známá.

Poznámka: Postup minimalizace podle klesajícího času zde není nutný. Nezávislost na směru minimalizace je dána tím, že v každém kroku jsme schopni minimalizovat příslušnou část kritéria tak, aby žádný zbytek nepřecházel do následujícího kroku sekvenční minimalizace. Např. na u_3 závisí jen poslední člen kritéria. Volbou optimálního řízení $u_3 = -(a/b)y_2$ tuto část kritéria zcela anulujeme. Ostatní členy kritéria zůstávají nedotčeny. Pouze jejich počet se postupně zmenšuje.

Poznámka: Uvedené kritérium se nazývá kritériem pro minimální rozptyl. Výpočet optimálního řízení podle tohoto kritéria je velmi jednoduchý, avšak praktická kvalita řízení může být pro některé modely špatná. Řídící zákon může generovat nekonečně velké hodnoty řídicí veličiny a řízení tak může být prakticky nerealizovatelné.

Postupná minimalizace kritéria

Abychom předešli nedostatkům, které sebou nese kritérium pro minimální rozptyl, budeme uvažovat kritérium, které penalizuje nejenom výstupy, ale také vstupy. Velikost penalizace vstupů volí prostřednictvím proměnné $\omega > 0$ uživatel. Jeho efekt je možno vyzkoušet v programu pro optimální řízení na str. 31. V našem příkladě má kritérium tvar

$$\mathcal{J} = \sum_{\tau=1}^3 (y_\tau^2 + \omega u_\tau^2). \quad (2.8)$$

Dosadíme-li z rovnice modelu (2.7) pro čas $t = 3$, lze kritérium (2.8) psát jako součet tří členů

$$\mathcal{J} = [y_1^2 + \omega u_1^2] + [y_2^2 + \omega u_2^2] + [(ay_2 + bu_3)^2 + \omega u_3^2].$$

V tomto případě se nám nepodaří volbou řízení jednotlivé členy kritéria postupně zcela anulovat. Minimalizaci začneme podle u_3 a od posledního členu, který jediný na u_3 závisí. Zbytek po jeho minimalizaci bude funkcí předchozích vstupů a výstupů a bude možné jej sloučit s ostatními členy

kritéria a zahrnout do další minimalizace. Protože v tomto případě se nám volbou optimálního řízení nepodaří jednotlivé členy kritéria zcela anulovat, a protože tyto členy jsou po minimalizaci vždy funkcemi předchozích dat, je zde nutné použít postup minimalizace "od posledního členu", který nyní provedeme.

Minimalizace 3. členu: Minimalizaci provedeme tak, že 3. člen kritéria

$$\mathcal{J}_3 = (ay_2 + bu_3)^2 + \omega u_3^2$$

doplníme na úplný čtverec v proměnné u_3 . Kvadrát můžeme anulovat volbou u_3 , zbytek po doplnění se připojí k ostatním členům.

$$\begin{aligned} \mathcal{J}_3 &\equiv (ay_2 + bu_3)^2 + \omega u_3^2 = u_3^2(b^2 + \omega) + 2aby_2u_3 + a^2y_2^2 = \\ &= (b^2 + \omega) \left[u_3^2 + 2u_3 \frac{ab}{b^2 + \omega} y_2 + \left(\frac{ab}{b^2 + \omega} \right)^2 y_2^2 - \left(\frac{ab}{b^2 + \omega} \right)^2 y_2^2 \right] + a^2y_2^2 = \\ &= (b^2 + \omega) \left[u_3 + \frac{ab}{b^2 + \omega} y_2 \right]^2 + \frac{a^2\omega}{b^2 + \omega} y_2^2. \end{aligned}$$

Označíme-li

$$r_3 = \frac{ab}{b^2 + \omega}, \quad s_3 = \frac{a^2\omega}{b^2 + \omega} = \frac{a\omega}{b} r_3,$$

pak optimálním akčním zásahem u_3 a zbytkem po minimalizaci třetího členu jsou

$$u_3 = -r_3 y_2, \quad \min_{u_3} \mathcal{J}_3 = s_3 y_2^2.$$

Kritérium po minimalizaci podle u_3 pak můžeme psát ve tvaru

$$\mathcal{J} = [y_1^2 + \omega u_1^2] + [y_2^2 + \omega u_2^2] + s_3 y_2^2 = [y_1^2 + \omega u_1^2] + [(1 + s_3)y_2^2 + \omega u_2^2].$$

Minimalizace 2. členu a zbytku: Akční zásah u_2 neovlivní prvního sčítance, takže minimalizujeme pouze druhého sčítance.

$$\begin{aligned} \mathcal{J}_2 &\equiv (1 + s_3)y_2^2 + \omega u_2^2 = (1 + s_3)(ay_1 + bu_2)^2 + \omega u_2^2 = \\ &= (1 + s_3)(a^2y_1^2 + 2aby_1u_2 + b^2u_2^2) + \omega u_2^2 = [(1 + s_3)b^2 + \omega] \left[u_2^2 + 2u_2 \frac{(1 + s_3)ab}{(1 + s_3)b^2 + \omega} y_1 + \right. \\ &\quad \left. + \left(\frac{(1 + s_3)ab}{(1 + s_3)b^2 + \omega} \right)^2 y_1^2 - \left(\frac{(1 + s_3)ab}{(1 + s_3)b^2 + \omega} \right)^2 y_1^2 \right] + (1 + s_3)a^2y_1^2 = \\ &= [(1 + s_3)b^2 + \omega] \left[u_2 + \frac{(1 + s_3)ab}{(1 + s_3)b^2 + \omega} y_1 \right]^2 + \frac{(1 + s_3)a^2\omega}{(1 + s_3)b^2 + \omega} y_1^2. \end{aligned}$$

Označíme-li

$$r_2 = \frac{(1 + s_3)ab}{(1 + s_3)b^2 + \omega}, \quad s_2 = \frac{(1 + s_3)a^2\omega}{(1 + s_3)b^2 + \omega} = \frac{a\omega}{b} r_2,$$

můžeme optimální akční zásah u_2 a zbytek po minimalizaci druhého členu psát ve tvaru

$$u_2 = -r_2 y_1, \quad \min_{u_2} \mathcal{J}_2 = s_2 y_1^2.$$

Kritérium po minimalizaci podle u_2 je

$$\mathcal{J} = [y_1^2 + \omega u_1^2] + s_2 y_1^2 = (1 + s_2)y_1^2 + \omega u_1^2.$$

Minimalizace 1. členu a zbytku:

$$\mathcal{J}_1 = (1 + s_2)y_1^2 + \omega u_1^2.$$

Pro $y_1 = ay_0 + bu_1$ obdobnými úpravami jako v předchozích krocích s označením

$$r_1 = \frac{(1 + s_2)ab}{(1 + s_2)b^2 + \omega}, \quad s_1 = \frac{a\omega}{b}r_1$$

dojdeme opět k výsledku

$$u_1 = -r_1y_0, \quad \min_{u_1} \mathcal{J}_1 = s_1y_0^2.$$

Výsledky mají formálně stejný tvar, pouze časový index klesá. Celou postupnou minimalizaci lze zapsat v následujícím algoritmu.

ALGORITMUS**ŘÍZENÍ**

Generování řídicího zákona

$$s_{N+1} = 0$$

for $t = N : -1 : 1$

$$r_t = \frac{(1 + s_{t+1})ab}{(1 + s_{t+1})b^2 + \omega}$$

$$s_t = \frac{a\omega}{b}r_t$$

end %for t

$$J_0 = s_1$$

Generování řídicí veličiny

y_0 - apriorní data

for $t = 1 : N$

$$u_t = -r_t y_{t-1}$$

end %for t

Poznámka: Kritérium "minimum variance" dostaneme, položíme-li $\omega = 0$. Potom bude

$$r_i = \frac{a}{b}, \quad u_i = -\frac{a}{b}y_{i-1} \quad \text{a} \quad s_i = 0, \quad i = 3, 2, 1,$$

což odpovídá výsledkům, které jsme již pro tento případ odvodili.

Cvičení: Uvedený algoritmus naprogramujte v jazyce MATLAB a testujte jeho funkci pro různé penalizace vstupu ω při různých počátečních podmínkách.

Návod: program je uveden v odstavci 2.4.

2.3 Základní úloha odhadování

Pro návrh řízení potřebujeme znát koeficienty modelu řízené soustavy. Ty ale ve většině praktických případů známy nejsou. Proto je musíme zjišťovat z dat měřených na soustavě. Tento proces nazýváme **odhadováním**. Základem odhadování je minimalizace kritéria

$$\mathcal{J} = \sum_{\tau=1}^N e_{\tau}^2 \tag{2.9}$$

kde e_τ , $\tau = 1, 2, \dots, N$, interpretujeme jako chyby, se kterými je splněna soustava rovnic (2.1) po dosazení naměřených dat a odhadnutých koeficientů. Chyby pro odhady minimalizující kritérium se nazývají **rezidua** a koeficienty modelu jsou pak **bodové odhady parametrů**.

Základní úlohu budeme nejprve demonstrovat na jednoduché úloze se statickým modelem. Při jejím řešení budeme používat **operátor průměrování** $\bar{x} = (\sum_{i=1}^N x_i)/N$, jehož základní vlastnosti jsou popsány v dodatku 11.2.

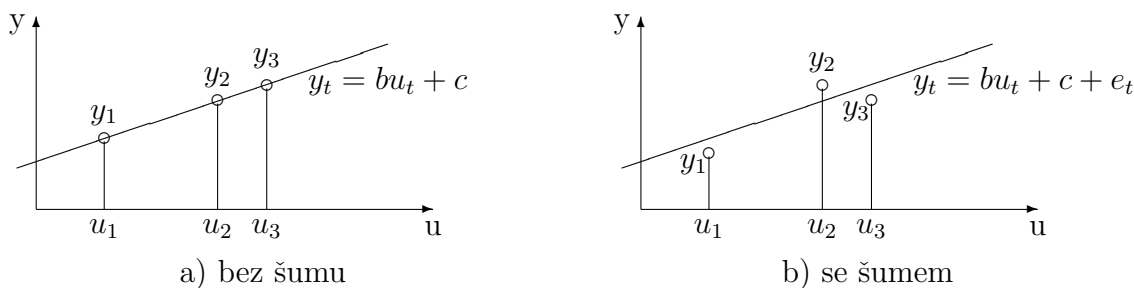
Odhad statického regresního modelu

A. Skalární vyjádření

Formulace úlohy pro statický model: Uvažujme regresní model, kde výstup y_t v daném okamžiku t závisí pouze na veličinách u_t a e_t ve stejném časovém okamžiku. Takový model se nazývá statický a je popsán rovnicemi

$$y_t = bu_t + c + e_t, \quad t = 1, 2, \dots, N. \quad (2.10)$$

Koeficienty b a c jsou regresní koeficienty modelu. Tento model pro identicky nulové e_t popisuje lineární závislost mezi vstupem u a výstupem y . Dvojice $[u_t, y_t]$ leží na přímce se směrnici b , jak ukazuje obr. 2.4 a). Bude-li $e_t \neq 0$, bude tato lineární závislost "narušena šumem", jak ukazuje obr. 2.4 b). Skutečné hodnoty výstupu y_t nebudou ležet na přímce, ale budou odchýleny vertikálně o hodnoty šumu, které jsme v rovnici označili e_t . Naším úkolem je určit odhady parametrů b , c přímky tak, aby součet čtverců těchto odchylek byl minimální.



Obrázek 2.4: Graf výstupu modelu

Výpočet regresních koeficientů: Pro model (2.10) a kritérium (2.9) máme odhadnout regresní koeficienty modelu b, c tak, abychom minimalizovali zvolené kritérium (2.9). Budeme přitom používat operátor průměrování a příslušnou symboliku (viz kapitola 11, část 11.2). Symbolem veličiny bez indexu označíme celý vektor hodnot této veličiny a pruhem značíme průměr hodnot příslušné veličiny. Pomocí této symboliky upravíme kritérium (2.9) na tvar

$$\bar{\mathcal{J}} = \frac{1}{N} \mathcal{J} = \frac{1}{N} \sum_{\tau=1}^N e_\tau^2 = \overline{e^2},$$

kde $e = [e_1, e_2, \dots, e_N]^T$. V příkladě budeme pro přehlednost uvažovat horizont $N = 3$, (tj. odhady parametrů počítáme ze tří naměřených datových dvojic $[y_\tau, u_\tau]$, $\tau = 1, 2, 3$).

Z rovnice modelu vyjádříme $e = y - bu - c$ a dosadíme do kritéria

$$\begin{aligned}\bar{\mathcal{J}} = \overline{e^2} &= \overline{(y - bu - c)^2} = \overline{y^2 + b^2u^2 + c^2 - 2byu - 2cy + 2bcu} = \\ &= \overline{y^2} + b^2\overline{u^2} + c^2 + 2b\overline{yu} - 2c\overline{y} + 2bc\overline{u}.\end{aligned}$$

Tento výraz budeme doplňovat na čtverec v proměnných b a c . Začneme proměnnou c ,

$$\begin{aligned}\bar{\mathcal{J}} &= c^2 - 2c(\overline{y} - b\overline{u}) + (\overline{y} - b\overline{u})^2 - (\overline{y} - b\overline{u})^2 + \overline{y^2} + b^2\overline{u^2} - 2b\overline{yu} = \\ &= [c - (\overline{y} - b\overline{u})]^2 + \overline{y^2} - \overline{y}^2 - 2b(\overline{yu} - \overline{y}\cdot\overline{u}) + b^2(\overline{u^2} - \overline{u}^2).\end{aligned}$$

Protože na c závisí jen první sčítanec, který je nezáporný, dostaneme minimum kritéria jeho anulováním, tedy volbou $c = \hat{c} = \overline{y} - b\overline{u}$ pro libovolné b . Zbytek kritéria po této minimalizaci bude

$$\min_c \bar{\mathcal{J}} = S_y^2 - 2bS_{yu} + b^2S_u^2,$$

kde S_y^2 , S_{yu} a S_u^2 jsou druhé výběrové momenty proměnných y a u definované vztahy

$$\begin{aligned}S_y^2 &= \overline{(y - \overline{y})^2} = \overline{y^2} - \overline{y}^2, \\ S_{yu} &= \overline{(y - \overline{y})(u - \overline{u})} = \overline{yu} - \overline{y}\overline{u}, \\ S_u^2 &= \overline{(u - \overline{u})^2} = \overline{u^2} - \overline{u}^2.\end{aligned}\tag{2.11}$$

Zbytek kritéria dále doplníme na čtverec v proměnné b a budeme minimalizovat

$$\begin{aligned}\min_b \bar{\mathcal{J}} &= S_u^2 \left[b^2 - 2b \frac{S_{yu}}{S_u^2} + \left(\frac{S_{yu}}{S_u^2} \right)^2 - \left(\frac{S_{yu}}{S_u^2} \right)^2 \right] + S_y^2 = \\ &= S_u^2 \left[b - \frac{S_{yu}}{S_u^2} \right]^2 + S_y^2 - \frac{(S_{yu})^2}{S_u^2}.\end{aligned}$$

Minimum kritéria je dosaženo pro $b = \hat{b} = \frac{S_{yu}}{S_u^2}$. Minimální hodnota kritéria je

$$\min_{b,c} \bar{\mathcal{J}} = S_y^2 - \frac{(S_{yu})^2}{S_u^2}\tag{2.12}$$

s odhady regresních koeficientů

$$\hat{b} = \frac{S_{yu}}{S_u^2}, \quad \hat{c} = \overline{y} - \hat{b}\overline{u},\tag{2.13}$$

kde druhé výběrové momenty jsou definovány v (2.11).

Cvičení: Uvedený algoritmus naprogramujte v jazyce MATLAB a vyzkoušejte jej pro různé disperse šumů a různé posloupnosti vstupů.

Návod: Nejdříve vypočtete druhé výběrové momenty, potom určete odhady regresních koeficientů. Odhadnutou přímku vykreslete a porovnejte se skutečnou. (Pro kontrolu je program uveden v odstavci 2.4).

Uvedené odvození je názorné, ale nehodí se pro modely vyšších řádů. Nyní uvedeme pro stejný model vhodnější způsob odvození pomocí maticového počtu.

B. Maticové vyjádření

Statický model z předchozího příkladu vyjádříme podobně jako v (2.2) pomocí vektorů

$$y_\tau = [u_\tau \ 1] \begin{bmatrix} b \\ c \end{bmatrix} + e_\tau = \varphi_\tau^T \theta + e_\tau.$$

Pro $\tau = 1, 2, 3$ můžeme sestavit maticový model

$$y = \Phi \theta + e, \quad (2.14)$$

kde

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, \quad \Phi = \begin{bmatrix} u_1 & 1 \\ u_2 & 1 \\ u_3 & 1 \end{bmatrix} = \begin{bmatrix} \varphi_1^T \\ \varphi_2^T \\ \varphi_3^T \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix}, \quad \theta = \begin{bmatrix} b \\ c \end{bmatrix}.$$

Kritérium je

$$\begin{aligned} \bar{J} &= \frac{1}{N} \sum_{i=1}^3 e_i^2 = \frac{1}{N} [e_1 \ e_2 \ e_3] \begin{bmatrix} e_1 \\ e_2 \\ e_3 \end{bmatrix} = \frac{1}{N} e^T e = \frac{1}{N} [y - \Phi \theta]^T [y - \Phi \theta] = \\ &= \frac{1}{N} [y^T y - y^T \Phi \theta - \theta^T \Phi^T y + \theta^T \Phi^T \Phi \theta]. \end{aligned}$$

Doplníme na úplný čtverec, jak jsme uvedli v odstavci 11.1. Abychom učinili úpravu přehlednější, použijeme symboliku z odstavce 11.1, a tedy označíme

$$\Phi^T \Phi = A, \quad \Phi^T y = w, \quad y^T y = c.$$

Pak, využijeme-li rovnost $(A^{-1}w)^T A(A^{-1}w) = w^T A^{-1}w$, je

$$\begin{aligned} \bar{J} &= \frac{1}{N} [y^T y - y^T \Phi \theta - \theta^T \Phi^T y + \theta^T \Phi^T \Phi \theta = \theta^T A \theta - \theta^T w - w^T \theta + c] = \\ &= \frac{1}{N} [\theta^T A \theta - \theta^T A A^{-1} w - w^T A A^{-1} \theta + (A^{-1} w)^T A (A^{-1} w) - w^T A^{-1} w + c] = \\ &= \frac{1}{N} [(\theta - A^{-1} w)^T A (\theta - A^{-1} w) + c - w^T A^{-1} w]. \end{aligned}$$

Minima kritéria dosáhneme volbou

$$\theta = \hat{\theta} = A^{-1} w = (\Phi^T \Phi)^{-1} \Phi^T y. \quad (2.15)$$

Minimální hodnota kritéria je

$$\min_{\theta} \bar{J} = \frac{1}{N} [c - w^T A^{-1} w] = \frac{1}{N} [y^T y - y^T \Phi (\Phi^T \Phi)^{-1} \Phi^T y]. \quad (2.16)$$

Struktura matic: Podívejme se nyní blíže na strukturu jednotlivých matic obsažených v předchozích výsledcích a porovnejme je s výsledky předchozího příkladu pro skalární odvození.

Základní prvky

$$\frac{1}{3} y^T y = \bar{y}^2, \quad \frac{1}{3} \Phi^T y = \bar{\varphi} y = \begin{bmatrix} \bar{y} u \\ \bar{y} \end{bmatrix},$$

$$\frac{1}{3}\Phi^T\Phi = \overline{\varphi\varphi^T} = \begin{bmatrix} \overline{u^2} & \overline{u} \\ \overline{u} & 1 \end{bmatrix}, \quad \left(\frac{1}{3}\Phi^T\Phi\right)^{-1} = \frac{1}{S_u^2} \begin{bmatrix} 1 & -\overline{u} \\ -\overline{u} & \overline{u^2} \end{bmatrix}.$$

Odhady parametrů

$$\begin{aligned} \hat{\theta} &= \frac{1}{S_u^2} \begin{bmatrix} 1 & -\overline{u} \\ -\overline{u} & \overline{u^2} \end{bmatrix} \begin{bmatrix} \overline{yu} \\ \overline{y} \end{bmatrix} = \frac{1}{S_u^2} \begin{bmatrix} \overline{yu} - \overline{y}\overline{u} \\ -\overline{u}\overline{yu} + \overline{y}\overline{u^2} & -\overline{y}\overline{u^2} + \overline{y}\overline{u}^2 \end{bmatrix} = \\ &= \frac{1}{S_u^2} \begin{bmatrix} S_{yu} \\ \overline{y}S_u^2 - \overline{u}S_{yu} \end{bmatrix} = \begin{bmatrix} \frac{S_{yu}}{S_u^2} \\ \overline{y} - \left(\frac{S_{yu}}{S_u^2}\right)\overline{u} \end{bmatrix} = \begin{bmatrix} \hat{b} \\ \hat{c} \end{bmatrix}. \end{aligned}$$

Zbytek kritéria

$$\min_{\theta} \bar{J} = \overline{y^2} - \begin{bmatrix} \overline{yu} & \overline{y} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \overline{y} - \hat{b}\overline{u} \end{bmatrix} = \overline{y^2} - \hat{b}\overline{yu} + \overline{y}^2 - \hat{b}\overline{y}\overline{u} = S_u^2 - \hat{b}S_{yu}.$$

Výsledky samozřejmě souhlasí s těmi, které jsme odvodili při použití skalárního zápisu – srv. (2.13) a (2.12).

Cvičení: Algoritmus odhadování naprogramujte v jazyce MATLAB a porovnejte jeho výsledky s výsledky předchozího algoritmu, odvozeného ve skalární symbolice.

Návod: Ze simulovaných dat sestavte podle (2.14) vektor y a datovou matici Φ . Z nich podle (2.15) zkonstruujte odhad parametrů $\hat{\theta}$.

Odhad regresního modelu obecného řádu

V předchozím příkladě jsme použili maticovou symboliku odvození odhadu pro statický regresní model. Nyní použijeme tuto metodu k odvození odhadu pro dynamický model s jedním vstupem a jedním výstupem a obecným řádem ν .

Model

Uvažujeme regresní model (2.2), popsáný v okamžiku t vztahem

$$y_t = \varphi_t^T \theta + e_t, \quad t = 1, 2, \dots, N, \quad (2.17)$$

kde $\varphi_t^T = [u_t, y_{t-1}, u_{t-1}, \dots, y_{t-\nu}, u_{t-\nu}, 1]$ a $\theta^T = [b_0, a_1, b_1, \dots, a_\nu, b_\nu, c]$.

Maticový model

Pro data naměřená do daného (aktuálního) okamžiku t , tj. v okamžicích $\tau = 1, 2, \dots, t$ lze předchozí model zapsat maticově

$$Y = \Phi\theta + E, \quad (2.18)$$

kde

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{bmatrix}, \quad \Phi = \begin{bmatrix} u_1 & y_0 & u_0 & \dots & y_{-\nu+1} & u_{-\nu+1} & 1 \\ u_2 & y_1 & u_1 & \dots & y_{-\nu+2} & u_{-\nu+2} & 1 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ u_N & y_{N-1} & u_{N-1} & \dots & y_{N-\nu} & u_{N-\nu} & 1 \end{bmatrix} = \begin{bmatrix} \varphi_1^T \\ \varphi_2^T \\ \dots \\ \varphi_N^T \end{bmatrix}$$

a podobně je sestaven i vektor šumů E .

Matice Φ má N řádků (počet naměřených dat) a $2\nu + 1$ sloupců (délka regresního vektoru).

Kritérium

$$\begin{aligned} \bar{J} &= \frac{1}{N}(Y - \Phi\theta)^T(Y - \Phi\theta) = \frac{1}{N} \sum_{i=1}^N (y_i - \varphi_i^T \theta)(y_i - \varphi_i^T \theta) = & (2.19) \\ &= \frac{1}{N} \sum_{i=1}^N \left(\begin{bmatrix} y_i & \varphi_i^T \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right) \left(\begin{bmatrix} y_i & \varphi_i^T \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right) = \\ &= \frac{1}{N} \sum_{i=1}^N \left(\begin{bmatrix} -1 & \theta^T \end{bmatrix} \begin{bmatrix} y_i \\ \varphi_i \end{bmatrix} \right) \left(\begin{bmatrix} y_i & \varphi_i^T \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right) = \\ &= \begin{bmatrix} -1 & \theta^T \end{bmatrix} \left(\frac{1}{N} \sum_{i=1}^N \begin{bmatrix} y_i \\ \varphi_i \end{bmatrix} \begin{bmatrix} y_i & \varphi_i^T \end{bmatrix} \right) \begin{bmatrix} -1 \\ \theta \end{bmatrix}. \end{aligned}$$

Součet v kulaté závorce posledního výrazu, který označíme $V(t)$, se nazývá **rozšířená informační matice**

$$V(N) = \frac{1}{N} \sum_{i=1}^N \begin{bmatrix} y_i \\ \varphi_i \end{bmatrix} \begin{bmatrix} y_i & \varphi_i^T \end{bmatrix} = \begin{bmatrix} \overline{y^2} & \overline{y\varphi^T} \\ \overline{y\varphi} & \overline{\varphi\varphi^T} \end{bmatrix}. \quad (2.20)$$

Tato matice a počet zpracovaných vzorků N obsahují veškerou informaci o odhadovaných parametrech, kterou jsme získali z naměřených dat. Jejich hodnoty je možno počítat buď jednorázově, jak to ukazuje vztah (2.20), nebo rekurentně pro $t = 1, 2, \dots, N$,

$$V(t) = \frac{1}{\kappa(t)} \left\{ (\kappa(t) - 1) V(t-1) + \begin{bmatrix} y_t \\ \varphi_t \end{bmatrix} \begin{bmatrix} y_t & \varphi_t^T \end{bmatrix} \right\} \quad (2.21)$$

$$\kappa(t) = \kappa(t-1) + 1, \quad (2.22)$$

kde nově zavedenou veličinu κ nazveme počítadlem vzorků. Rekurzivní výpočet startuje s hodnotami $V(0) = \mathbf{0}$, $\kappa(0) = 0$ ($\mathbf{0}$ značí matici nul), nebo s apriorní informací vyjádřenou počáteční hodnotou $V(0) = V_0$, $\kappa(0) = \kappa_0$, kde matice V_0 je symetrická a pozitivně definitní a κ_0 vyjadřuje důvěru k naší apriorní informaci tak, jako kdyby byla získána z počtu κ_0 vzorků. Tedy, čím je κ_0 větší, tím větší je naše důvěra v apriorní informaci.

Rozšířená informační matice má blokovou strukturu a pro její jednotlivé bloky zavedeme značení

$$V(t) = \begin{bmatrix} \overline{y^2} & \overline{y\varphi^T} \\ \overline{y\varphi} & \overline{\varphi\varphi^T} \end{bmatrix} \equiv \begin{bmatrix} V_y(t) & V_{y\varphi}^T(t) \\ V_{y\varphi}(t) & V_\varphi(t) \end{bmatrix}. \quad (2.23)$$

Tak jako jsme doplněním na čtverec získali vztahy (2.15) a (2.16), lze postupovat i v tomto případě. Z (2.19) plyne (pro přehlednost vynecháme časový index, který může být buď průběžný

t nebo konečný N)

$$\frac{1}{\kappa} [-1, \theta^T] \begin{bmatrix} V_y & V_{y\varphi}^T \\ V_{y\varphi} & V_\varphi \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix} = V_y - 2V_{y\varphi}^T \theta + \theta^T V_\varphi \theta.$$

Tento výraz doplníme na čtverec v proměnné θ (viz 11.1) a dostaneme

$$\begin{aligned} \frac{1}{\kappa} [\theta^T V_\varphi \theta - 2\theta^T V_\varphi V_\varphi^{-1} V_{y\varphi} + (V_\varphi^{-1} V_{y\varphi})^T V_\varphi (V_\varphi^{-1} V_{y\varphi}) + V_y - V_{y\varphi}^T V_\varphi^{-1} V_{y\varphi}] = \\ = \frac{1}{\kappa} [(\theta - V_\varphi^{-1} V_{y\varphi})^T V_\varphi (\theta - V_\varphi^{-1} V_{y\varphi}) + V_y - V_{y\varphi}^T V_\varphi^{-1} V_{y\varphi}], \end{aligned}$$

kde $V_\varphi^T = V_\varphi$, neboť V_φ je symetrická matice. Anulováním první části tohoto výrazu dostáváme odhady parametrů,

$$\hat{\theta} = V_\varphi^{-1} V_{y\varphi}, \quad (2.24)$$

zbytek je minimální hodnota kritéria

$$\min_{\theta} \bar{\mathcal{J}} = \frac{1}{\kappa} [V_y - V_{y\varphi}^T V_\varphi^{-1} V_{y\varphi}]. \quad (2.25)$$

Odvozené vzorce i postup výpočtu shrneme do následujícího algoritmu.

ALGORITMUS

ODHAD

Přepočítání informační matice

Máme: $V(t-1), \kappa(t-1)$ – stará informační matice a počítadlo.

Změříme: y_t, u_t – nová data.

Sestavíme: $\varphi_t^T = [u_t, y_{t-1}, \varphi_{t-1}(1 : 2(\nu-1)), 1]$ [viz (2.17)]

– nový regresní vektor (stará data posuneme o 2 kroky doprava, poslední tři vypustíme, na první dvě pozice umístíme u_t a y_{t-1} a na poslední dáme jedničku).

Přepočteme:

$$V(t) = V(t-1) + \begin{bmatrix} y_t \\ \varphi_t \end{bmatrix} \begin{bmatrix} y_t & \varphi_t^T \end{bmatrix}, \quad \kappa(t) = \kappa(t-1) + 1 \quad [\text{viz (2.21), (2.22)}]$$

– nová informační matice a počítadlo.

Výpočet odhadů koeficientů

Rozdělíme:

$$V(t) = \begin{bmatrix} V_y(t) & V_{y\varphi}^T(t) \\ V_{y\varphi}(t) & V_\varphi(t) \end{bmatrix} \quad [\text{viz (2.23)}]. \quad (2.26)$$

Odhad parametrů:

$$\hat{\theta}_t = V_\varphi^{-1}(t)V_{y\varphi}(t) \quad [\text{viz (2.24)}]. \quad (2.27)$$

Minimální hodnota kritéria je:

$$\frac{\Lambda_t}{\kappa(t)} = \frac{1}{\kappa(t)} [V_y(t) - V_{y\varphi}^T(t)V_\varphi^{-1}(t)V_{y\varphi}(t)] = \frac{1}{\kappa(t)} [V_y(t) - \hat{\theta}_t^T V_{y\varphi}(t)] \quad [\text{viz (2.25)}]. \quad (2.28)$$

Cvičení: Algoritmus odhadování naprogramujte v jazyce MATLAB a ověřte jeho funkci pro data simulovaná různými modely (tj. různého řádu, s různě velikým šumem a různým vstupním signálem). Porovnejte simulované a odhadnuté parametry pro případ shody struktury simulovaného a odhadovaného modelu. Rozmyslete si, co se při odhadování děje při neshodě struktury (např. neshodě řádů) těchto modelů.

Návod: Program pro případné experimenty je uveden v následující části 2.4.

Poznámka: Všimněte si, že pro výpočet bodového odhadu parametrů (2.27) je třeba provést inverzi matice V_φ , která musí být pozitivně definitní. Tato vlastnost se vlivem numeriky může při postupných přepočtech ztratit.

Poznámka: Numericky lepšího řešení je možno dosáhnout pomocí metod rozkladu pozitivně definitní matice na součin dvou shodných (transponovaných) trojúhelníkových matic. Tento rozklad se použije pro informační matici V ve vztahu (2.21) a vztah se upraví tak, aby prováděl přepočet přímo pro jeden trojúhelník (který lze nazvat odmocninou informační matice). Tímto postupem se zaručí trvalá definitnost přepočítávané informační matice. Toto pojednání však překračuje rámec našeho výkladu a případného zájemce odkazujeme např. na literaturu [3] nebo podrobnější pojednání v [4].

2.4 Programy

Cílem těchto, ale i dalších, programů je ukázat praktickou realizaci uvedených algoritmů a dát tak možnost "pohrát" si s nimi. Dalším cílem je také zadat určitou formu zápisu jak pro hlavní programy, tak i pro podprogramy, pro vytvářený dalších programů.

Program pro optimální řízení soustavy 1. řádu

Tento program realizuje algoritmus řízení ze str. 23.

```
% Optimální řízení deterministické soustavy 1. řádu
% -----
clc, clear, clear all
n = 30;           % horizont řízení
a = 0.9; b=0.2;   % parametry simulované soustavy
om = 0.1;        % penalizace u
y(1)=10;         % počáteční y
r=zeros(1,n+1); s=zeros(1,n+1);

% konstrukce řídicího zákona (výpočet r, s)
for t=n:-1:1
    s1 = 1+s(t+1);
    r(t)= s1*a*b/(s1*b*b+om);
    s(t)= a*om/b*r(t);
end
w = zeros(1,n);      % žádaná hodnota výstupu (na tu řídíme)

% simulace s řízením soustavy
for t=1:n
    if t>=10, w(t) = 20; end
    if t<=20, w(t) = -15; end
    u(t+1) = -r(t)*(y(t)-w(t)); % regulátor
    y(t+1) = a*y(t)+b*u(t+1);  % soustava
end

% zobrazení výsledků
t1 = 1:(n+1);
figure(1), plot(t1,u,'o',t1,y,1:n,w,':')
title('Vstupy (o) a výstupy (-) soustavy při optimálním řízení')
```

Poznámka: Proměnné $r(t)$ a $s(t)$ jsou ve skutečnosti skalární koeficienty, které je možno průběžně přepočítávat. My jsme je v programu zavedli jako vektory, abychom uchovali jejich průběžné hodnoty pro pozdější vykreslení. Výpočty těchto koeficientů se provádějí podle algoritmu v závěru odstavce 2.2.

Poznámka: Žádaná hodnota výstupu w je hodnota, na které chceme udržovat výstupní veličinu y . K této úloze přejdeme tak, že místo kritéria (2.5), kde se penalizuje výstup y_T^2 penalizujeme odchylku výstupu od žádané hodnoty $(y_T - w)^2$. Logicky, jestliže první kritérium zaručuje blízkost výstupu k nule, zaručuje druhé kritérium blízkost výstupu k žádané hodnotě. Žádaná hodnota je ukládána do vektoru rovněž z důvodů vykreslení.

Doporučené experimenty

1. Výstup y má sledovat žádanou hodnotu w . Změna žádané hodnoty v cyklu pro t a sledování reakce výstupu je základním experimentem.
2. Penalizace vstupu zaručuje, že hodnoty vstupu se nebudou měnit k extrémním hodnotám (pro om by optimální řízení vyžadovalo nekonečné hodnoty vstupu). Toto utlumení hodnot vstupu má za následek zhoršení kvality řízení. Pro menší hodnoty om bude tedy lepší kvalita řízení s většími změnami vstupu a naopak. Ověřte např. pro $om=0.01$ a $om=0.5$
3. Hodnoty koeficientů regulátoru se postupně ustalují. Tento fakt je možno pozorovat na obrázku

```
figure(1),plot(t1(end:-1:1),r,t1(end:-1:1),s)
```

Program pro odhad statického modelu – pomocí momentů

Tento program ukazuje jednorázový algoritmus odhadu využívající jako statistiky momenty měřených signálů. Algoritmus používá pouze měřená data, nikoliv apriorní informaci. Možnost využití apriorní informace ukážeme v dalším příkladě, který pro konstrukci statistiky používá rozšířenou informační matici. Tento program používá vzorce (2.13) a (2.12).

```
% Odhad statického modelu metodou nejmenších čtverců
% (pomocí momentů signálů)
% -----
clc, clear, clear all
randn('seed',17);
n = 30; % horizont odhadování
as = 0.9; bs=0.2; % parametry simulované soustavy
ae = 1; % amplituda šumu simulované soustavy
yy=zeros(1,n); uu=zeros(1,n);

% simulace a měření dat
for t=1:n
    e = ae*randn; % šum s amplitudou sig=1
    u = t; % vstup (lineárně narůstá s časem)
    y = as*u+bs+e; % simulovaná soustava
    yy(t) = y; % data y
    uu(t) = u; % data u
end

% výpočet charakteristik dat
mu = sum(uu)/n; % střední hodnota u
my = sum(yy)/n; % střední hodnota y
du = uu-mu; % odchylka u
dy = yy-my; % odchylka y
s2u= du*du'/n; % rozptyl u
s2y= dy*dy'/n; % rozptyl y
syu= dy*du'/n; % kovariance y,u

% odhad koeficientů regresní přímky a zbytku kritéria
```



```

a = syu/s2u;           % odhad směrnice regresní přímky
b = my-a*mu;          % odhad absolutního členu regresní přímky
la= s2y-a*syu;        % odhad zbytku kritéria (rozptylu šumu)

% zobrazení výsledků
a,b,la                % tisk koeficientů a,b a zbytku kritéria
yr = a*uu+b;          % předpovídané výstupy
figure(1)
plot(uu,yy,'o',uu,yr) % graf dat a reg. přímky
title('Měřené datové dvojice (o) a regresní přímka (-)')

```

Doporučené experimenty

1. Základním experimentem je změna koeficientů simulované soustavy a a b . Pro malé amplitudy šumu $ae=0.1$ je algoritmus odhadu spolehlivě najde.
2. Čím větší je amplituda šumu ae , tím více budou body $[u, y]$ ze simulované soustavy "rozházeny" kolem regresní přímky, tím větší bude hodnota odhadu zbytku kritéria la a tím horší budou odhady regresních koeficientů. Ověřte pro $ae=5$ a $ae=50$.
3. Vypočtete rozptyl chyb predikce ($ep=yy-yz$) a ověřte, že se rovná hodnotě la .

Program pro odhad statického modelu – s informační maticí

Tento příklad ukazuje bayesovské řešení odhadu. Generovaná data se průběžně ukládají do statistiky v , ze které je nakonec možno vypočítat bodové odhady. Apriorní znalost může být vyjádřena pomocí počáteční hodnoty statistiky. Ta je v následujícím programu nastavena jako nulová ($v=zeros(3)$), a tedy žádná. V tomto stavu dává program stejné výsledky jako předchozí varianta, používající momenty. Srovnajte!

Program používá algoritmus odhadování ze strany 29, upravený pro případ statického modelu, tj. modelu řádu 0.

```

% Odhad statického modelu metodou nejmenších čtverců
% (pomocí informační matice)
% -----
clc, clear, clear all
randn('seed',17);
n = 30;           % horizont
a = .9;           % parametry simulované soustavy
b = .2;           % parametry simulované soustavy
ae = 1;           % amplituda šumu simulované soustavy
yg=zeros(1,n); ug=zeros(1,n); v=zeros(3);

for t=1:n
    % simulace a měření dat
    e = ae*randn; % šum s amplitudou ae
    u = t;        % vstup (lineárně narůstá s časem)
    y = a*u+b+e;  % simulovaná soustava
    yg(t) = y;    % y pro graf
    ug(t) = u;    % u pro graf

```

```

% přepočítání informační matice v (nenormované)
v = v+[y u 1]'*[y u 1];
end

% normování informační matice
v=v/n;

% rozdělení informační matice
vd =v(1,1);
vdf=v(2:3,1);
vf =v(2:3,2:3);

% výpočet bodových odhadu
th = inv(vf)*vdf;      % odhady parametrů
la = vd-th*vdf;      % odhad zbytku kritéria

% zobrazení výsledků
th,la,v              % tisk výsledků
yr = th(1)*ug+th(2); % předpovídané výstupy (na regresní přímce)
figure(1)
plot(ug,yg,'o',ug,yr)
title('Měřené datové dvojice (o) a regresní přímka (-)')

```

Apriorní informaci je možno zavést prostřednictvím nenulové apriorní statistiky. Zkuste ji demonstrovat tak, že zvýšíte amplitudu šumu v simulované soustavě, čímž se zhorší kvalita bodových odhadů, a "vnutíte" algoritmu správné (nebo přibližné) hodnoty těchto odhadů.

Bezprostřední otázka je, jaké hodnoty se mají do apriorní statistiky zadat, aby odpovídaly apriorním odhadům parametrů. K tomu je nejlépe použít tzv. **metodu fiktivních regresních vektorů**. Při ní dosadíme do rovnice odhadovaného modelu apriorní hodnoty parametrů a nulový, nebo téměř nulový, šum. Z této rovnice generujeme data, a používáme je pro apriorní identifikaci, tj. před použitím měřených dat. Množství takto použitých fiktivních dat, vzhledem k délce datového vzorku, určuje váhu apriorní informace. Zkuste upravit program sami. Pro kontrolu je dále uvedena změněná část programu od začátku až do cyklu pro čas $t=1:n$.

```

% Odhad statického modelu metodou nejmenších čtverců
% (pomocí informační matice)
% -----
clc, clear, clear all
randn('seed',17);
n0 = 100;          % apriorní horizont
n = 300;          % horizont
a = .9;           % parametry simulované soustavy
b = .2;           % parametry simulované soustavy
ae = 20;          % amplituda šumu simulované soustavy
yg=zeros(1,n); ug=zeros(1,n); v=zeros(3);

a0 = 1; b0 = .5;  % apriorní parametry
for i=1:n0
    u = randn;

```

```

    fi = [a0*u+b0 u 1];
    v = v+fi'*fi;
end

```

Dále následuje vlastní odhadování s měřenými daty.

Doporučené experimenty

1. Stejně jako v předchozím příkladě je možno experimentovat s hodnotami parametrů simulované soustavy a porovnávat je s bodovými odhady.
2. Ověřte vliv apriorní informace pro různé poměry n_0 a n při velkých a malých amplitudách šumu ae .

Program pro odhad parametrů soustavy řádu m

Tento program používá algoritmus odhadování ze strany 29 pro obecný řád odhadovaného modelu.

```

% Odhad metodou nejmenších čtverců
%      skalární regresní model řádu m
% -----
clc, clear, clear all
randn('seed',17);

% zadání simulace
n=300;           % horizont odhadování
m=2;            % řád simulované soustavy
ae=.1;          % amplituda šumu simulované soustavy
a=[.9 -.1]; b=[.5 -.3 .1]; c=1; % parametry simulované soustavy

% sestavení vektoru parametrů simulované soustavy
% odpovídající struktuře regresního vektoru:
%   fi = [u(t) y(t-1) u(t-1) y(t-2) u(t-2) 1]
th=b(1);
for i=1:m
    th=[th; a(i); b(i+1)];
end
th=[th; c];

% deklarace pomocných vektorů a matic
yg=zeros(1,n); % vektor výstupů pro graf
mv=2*m+2; fi=zeros(mv,1); % regresní vektor odhadovaného modelu
mv1=mv+1; v=1e-8*eye(mv1); % rozšířená informační matice

% simulační smyčka + sběr dat
for t=1:n
    % simulace
    e=ae*randn; % generování šumu
    u=.2*randn; % generování vstupu
    fi=[u; fi(1:2*m); 1]; % přepočítání regresního vektoru

```

```

y=th'*fi+e;                % simulace výstupu

% sběr dat do statistiky
v=v+[y; fi]*[y; fi]';     % přepočít rozšířené informační matice

fi=[y; fi(1:2*m); 1];     % přepočít regresního vektoru
yg(t)=y;                  % y pro graf
end

% normování rozšířené informační matice
v=v/n;

% rozdělení informační matice
vd =v(1,1);
vdf=v(2:mv1,1);
vf =v(2:mv1,2:mv1);

% výpočet odhadu parametrů a zbytku kritéria
theta =inv(vf)*vdf;       % odhad parametrů
lambda=vd-theta'*vdf;    % odhad rozptylu šumu

% zobrazení výsledků
figure(1), plot(yg)
title('Datový vzorek použitý pro odhad')
disp('Simulované (th) a odhadnuté (theta) parametry')
disp(' ')
disp('      th      theta')
disp([th, theta])        % tisk simulovaných (th)
                          % a odhadnutých (theta) parametrů

```

Doporučené experimenty

1. Odhad parametrů pro různé řády modelu (ten je zde společný jak pro simulovanou soustavu tak i pro odhadovaný model).
2. Odhad parametrů pro odlišné řády simulované struktury a odhadovaného modelu – tzv. neshoda struktury. V tomto případě je algoritmus odhadování nucen aproximovat. Program je třeba upravit – odlišit řády simulace a odhadování.

Kapitola 3

Pravděpodobnost

V této kapitole připomeneme některé pojmy a tvrzení teorie pravděpodobnosti, které budeme dále používat při vytváření pravděpodobnostních modelů a při práci s nimi. Pro základní informaci o klasickém pojetí pravděpodobnosti se čtenář může obrátit např. na literaturu [2], [5] nebo [6], případně na jinou klasickou učebnici.

V dalších úvahách se budeme setkávat se dvěma typy náhodných veličin. Prvním typem jsou **signály** v daných časových okamžicích (např. vstup u_t , resp. výstup y_t daného systému). Tyto náhodné veličiny jsou blízké klasickému pojetí pravděpodobnosti. Jejich naměřené hodnoty tvoří datový výběr. Náhodnost v hodnotách měřených dat je možno interpretovat jako vliv určité poruchy, např. jako chyby měření.

Dalším typem náhodné veličiny jsou **neznámé parametry** systému. Tady se již jedná o čistě bayesovskou interpretaci náhodnosti, která je způsobena neznalostí příslušných parametrů. Víme sice, že parametry jsou konstanty, ale jejich přesné hodnoty neznáme a popisujeme je – podle Bayesova přístupu – tak, že specifikujeme všechny jejich možné hodnoty a jejich pravděpodobnostní rozdělení. Tím dostávají charakter náhodné veličiny.

V dalším výkladu této kapitoly budeme mít na mysli především signály jako reprezentanty náhodných veličin. Ty jsou totiž pro osvětlení základních pojmů snazší.

3.1 Základní pojmy

Náhodný pokus

Je to experiment, který při opakování i za relativně stálých podmínek dává různé výsledky. V diskutované úloze odhadu a řízení spočívá v měření datového signálu v určitém časovém okamžiku. Např. v čase τ změříme výstup a obdržíme hodnotu y_τ . Tato naměřená hodnota je výsledek náhodného pokusu.

Náhodná veličina (nv) X

Je to proměnná, jejíž hodnoty jsou přiřazeny výsledkům náhodného pokusu. Protože výsledky našeho pokusu jsou číselné povahy, můžeme hodnoty náhodné veličiny s těmito výsledky přímo ztotožnit. Hodnoty náhodné veličiny jsou tedy přímo naměřené hodnoty signálů.

Poznámka: *Podle toho, zda je množina hodnot náhodné veličiny spočetná (konečná) nebo nespočetná, dělíme náhodné veličiny na diskrétní nebo spojitě. U diskrétní nv X má smysl mluvit o pravděpodobnosti její libovolné hodnoty x . Máme tím na mysli pravděpodobnost, s*

níž tato náhodná veličina X nabývá hodnotu x . Pro spojitou nv jsou pravděpodobnosti jejich jednotlivých hodnot vždy nulové, a proto mluvíme zpravidla o pravděpodobnosti daného intervalu. Máme tím na mysli pravděpodobnost, s níž tato náhodná veličina nabývá hodnotu z uvažovaného intervalu.

Příklady

Náhodné pokusy:

házení mincí – výsledek: padl rub, padl líc;

průjezd aut křižovatkou – výsledek: odbočení doprava, jízda rovně, odbočení doleva;

doba čekání na dopravní prostředek – výsledek: nezáporné reálné číslo (doba čekání).

Přiřazení hodnot náhodné veličiny výsledkům pokusu:

padl rub $\rightarrow 0$, *padl líc* $\rightarrow 1$;

odbočení doprava $\rightarrow 1$, *jízda rovně* $\rightarrow 0$, *odbočení doleva* $\rightarrow -1$;

doba čekání \rightarrow *nezáporné reálné číslo*.

Rozdělení náhodné veličiny

Náhodná veličina svým oborem hodnot vymezuje množinu všech možných výsledků. Tato množina výsledků však sama neříká nic o tom, jak jsou jednotlivé výsledky pravděpodobné – tj. nepopisuje náhodný pokus, pouze vymezuje "prostor, ve kterém se budeme pohybovat". Náhodný pokus je plně popsán definicí náhodné veličiny a jejího rozdělení. Toto rozdělení u diskrétní náhodné veličiny je zadáno určením pravděpodobností všech jednotlivých hodnot, u spojitě náhodné veličiny určením pravděpodobností všech možných intervalů hodnot. V našem případě budeme rozdělení diskrétní náhodné veličiny zadávat pomocí **pravděpodobnostní funkce**, rozdělení spojitě náhodné veličiny pomocí **hustoty pravděpodobnosti**. V dalším textu budeme pro oba typy rozdělení používat název hustota pravděpodobnosti a zkracovat jej hp. Definici hp uvedeme dále.

♣ Náhodná veličina a její rozdělení jsou úplným popisem náhodného pokusu.

Distribuční funkce

Distribuční funkce F_X náhodné veličiny X je reálná funkce definovaná na celé reálné ose vztahem

$$F_X(x) = P(X \leq x), \quad (3.1)$$

tj. hodnota distribuční funkce v bodě x udává pravděpodobnost $P(\cdot)$, s níž nv X nabude hodnot menších nebo rovných číslu x . Je to tedy pravděpodobnost intervalu $(-\infty; x)$.

Poznámka: Z důvodu jednoduchosti zápisu budeme tam, kde nemůže dojít k omylu, vynechávat index X . Náhodná veličina bude potom určena pomocí své realizace v argumentu distribuční funkce.

Hustota pravděpodobnosti

Pro diskrétní nv se také nazývá **pravděpodobnostní funkce** a je definovaná pomocí distribuční funkce F_X nv X , a to jako reálná funkce $f(x)$, pro kterou platí

$$F_X(x) = \begin{cases} \int_{-\infty}^x f(\xi) d\xi & \text{pro spojitou nv,} \\ \sum_{i \leq x} f(i) & \text{pro diskrétní nv.} \end{cases}$$

Poznámka: Pomocí definice distribuční funkce a definice hp můžeme pravděpodobnost intervalu (a, b) vyjádřit takto:

$$P(X \in (a, b)) = P(X < b) - P(X \leq a) = F(b) - F(a) = \int_{-\infty}^b f(x) dx - \int_{-\infty}^a f(x) dx = \int_a^b f(x) dx.$$

Vidíme tedy, že platí:

- ♣ Pravděpodobnost intervalu (a, b) hodnot spojitě náhodné veličiny je dána integrálem $\int_a^b f(x) dx$ z hp přes tento interval.

Poznámka: Obdobné tvrzení platí i pro pravděpodobnost množiny hodnot diskrétní náhodné veličiny:

- ♣ Pravděpodobnost množiny M hodnot diskrétní náhodné veličiny je dána součtem $\sum_{x \in M} f(x)$ pravděpodobností těchto hodnot.

Sdružená distribuční funkce

Pro dvouprvkový náhodný vektor $[X, Y]$, kde X a Y jsou nějaké nv, je tato funkce definována vztahem

$$F_{XY}(x, y) = P((X \leq x) \wedge (Y \leq y)).$$

Sdružená hustota pravděpodobnosti

Je pro dvouprvkový náhodný vektor $[X, Y]$ definována pomocí jeho sdružené distribuční funkce jako funkce $f(x, y)$ dvou reálných proměnných, vyhovující vztahu

$$F_{XY}(x, y) = \begin{cases} \int_{-\infty}^x \int_{-\infty}^y f(\xi, \eta) d\xi d\eta & \text{pro spojitou nv,} \\ \sum_{i \leq x} \sum_{j \leq y} f(i, j) & \text{pro diskrétní nv.} \end{cases}$$

- ♣ Pravděpodobnost, že náhodný vektor nabude hodnoty z dané množiny v rovině je dána dvojným integrálem ze sdružené hp přes tuto množinu.

Marginální hustota pravděpodobnosti

Je definována pro sdruženou hp $f(x, y)$ vztahem

$$f(x) = \begin{cases} \int_{-\infty}^{\infty} f(x, y) dy & \text{pro spojitou nv,} \\ \sum_{y \in Y} f(x, y) & \text{pro diskrétní nv.} \end{cases}$$

♣ Představíme-li si sdruženou hp $f(x, y)$ jako kopeček písku, pak vytvoření marginální hp si můžeme představit tak, že jsme písek rovnoměrně nahrnuli do jedné z os soustavy souřadnic rovnoběžně s druhou osou.

Podmíněná hustota pravděpodobnosti

Je to funkce jedné reálné proměnné x s parametrem y , definovaná vztahem

$$f(x|y) = \frac{f(x,y)}{f(y)}.$$

Podmíněná hp je úměrná sdružené. Jmenovatel v předchozí definici, který nezávisí na proměnné x , představuje jen normování na integrál (součet) roven jedné.

♣ Představíme-li si sdruženou hp $f(x, y)$ jako kopeček písku, pak podmíněnou hp $f(x|y_0)$ si můžeme představit jako řez povrchu kopečku rovinou kolmou k podstavě, rovnoběžnou s osou x a protínající osu y v bodě y_0 . Tento řez určuje profil, který je ještě nutno normalizovat na jednotkový integrál.

3.2 Diskrétní rozdělení

Na příkladech ukážeme základní nástroje a jejich vnitřní strukturu pro popis diskrétních náhodných veličin.

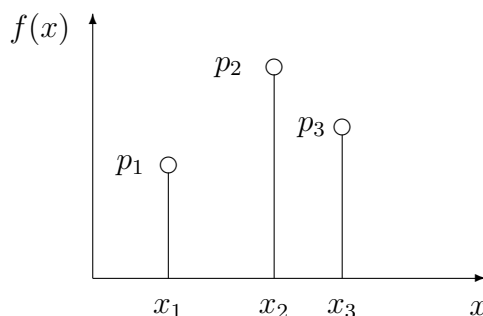
Jednorozměrná náhodná veličina

Popis: Diskrétní nv má za obor hodnot konečnou (spočetnou) množinu nezáporných reálných čísel. Její rozdělení (hp) je proto dáno posloupností hodnot a posloupností odpovídajících pravděpodobností. Hodnoty pravděpodobností lze zadat tabulkou, grafem nebo jako posloupnost předpisem. Uvedeme příklady jednotlivých zadání.

Tabulkou

x	x_1	x_2	x_3
$f(x)$	p_1	p_2	p_3

Grafem, jako např. na obr. 3.1.



Obrázek 3.1: Graf diskrétního rozdělení

Předpisem, jako např. alternativní rozdělení $f(x) = p^x(1-p)^{1-x}$, $x \in \{0, 1\}$.

Význam: Diskrétní hp $f_X(x)$ udává pravděpodobnosti jednotlivých hodnot diskrétní náhodné veličiny X , tj. $f(x_i) = f_X(x_i) = P(X = x_i) = p_i$.

Platí: $\sum_i p_i = 1$.

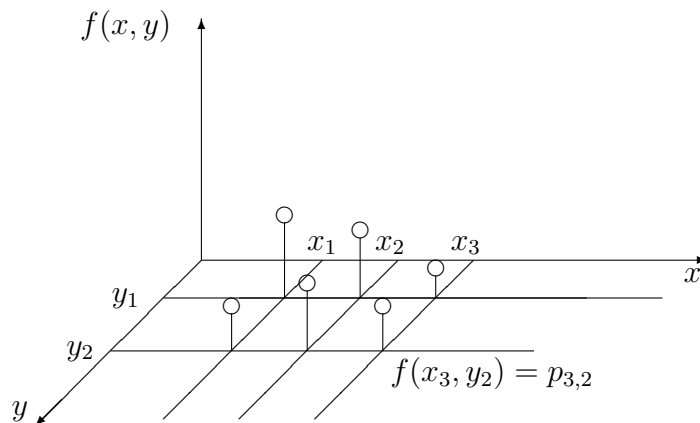
Dvourozměrná náhodná veličina

Popis: Dvourozměrná diskretní nv $[X, Y]$, nebo také dvourozměrný diskretní náhodný vektor $[X, Y]$, je uspořádaná dvojice diskretních náhodných veličin X a Y . Dvourozměrná diskretní hp je funkce dvou proměnných, která každé uspořádané dvojici (x, y) , kde $x \in X$ a $y \in Y$, přiřazuje reálné číslo z intervalu $\langle 0, 1 \rangle$. Toto číslo udává pravděpodobnost, s níž při současné realizaci obou nv X, Y nabude nv X hodnotu x a nv Y hodnotu y . Tuto funkci lze reprezentovat maticí, pomocí tabulky, pomocí grafu nebo nějakým analytickým předpisem.

Tabulkou

(x, y)	y_1	y_2
x_1	$p_{1,1}$	$p_{1,2}$
x_2	$p_{2,1}$	$p_{2,2}$
x_3	$p_{3,1}$	$p_{3,2}$

Grafem, jako např. na obr. 3.2.



Obrázek 3.2: Graf dvourozměrného diskretního rozdělení

Předpisem, jako např. rozdělení beta

$$f(x, y) = \frac{1}{B(x, y)} \alpha^{x-1} (1-\alpha)^{y-1}, \quad x, y \in N, \quad \alpha \in \langle 0, 1 \rangle,$$

kde $B(x, y)$ je funkce beta, definovaná v (11.3).

Význam: Je dán přímo definicí $f(x, y) = P((X = x) \wedge (Y = y))$.

Platí: $\sum_i \sum_j p_{i,j} = 1$.

Marginální pravděpodobnosti náhodného vektoru $[X, Y]$

Jsou to pravděpodobnosti hodnot jednotlivých složek tohoto vektoru, které dostaneme jako součty ve sloupcích nebo řádcích matice reprezentující příslušnou sdruženou hp, tj.

$$f(y_j) = \sum_i p_{i,j} \quad \text{pro všechna } j, \text{ nebo } f(x_i) = \sum_j p_{i,j} \quad \text{pro všechna } i.$$

Podmíněné pravděpodobnosti náhodného vektoru $[X, Y]$

Jsou to pravděpodobnosti hodnot jedné složky za předpokladu, že je pevně zvolena hodnota druhé složky. Můžeme je počítat jako normalizované řezy sdruženou hp, tedy jako normalizované sloupce nebo řádky příslušné matice reprezentující sdruženou hp, tj.

$$f(x_i|Y = y_j) = \frac{p_{i,j}}{\sum_i p_{i,j}} \quad \text{pro všechna } i, \text{ nebo } f(y_j|X = x_i) = \frac{p_{i,j}}{\sum_j p_{i,j}} \quad \text{pro všechna } j.$$

3.3 Spojité rozdělení

Podobně, jako jsme to učinili pro diskrétní nv, ukážeme nyní strukturu základních nástrojů také pro spojité nv.

Jednorozměrná náhodná veličina**Popis:**

Spojité nv může nabývat libovolných reálných hodnot. Její rozdělení je zpravidla zadáno pomocí hp, což je reálná funkce jedné reálné proměnné. Tu je možno zadat funkčním předpisem nebo grafem.

Předpisem, např. normální rozdělení

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2}, \quad x \in X, \quad \mu, \sigma \in R, \quad \sigma > 0$$

nebo exponenciální rozdělení

$$f(x) = \frac{1}{\delta} e^{-\frac{x}{\delta}}, \quad x \geq 0 \quad \delta \in R, \quad \delta > 0.$$

Grafem, jako je např. pro normální rozdělení načrtnut na obr. 3.3.

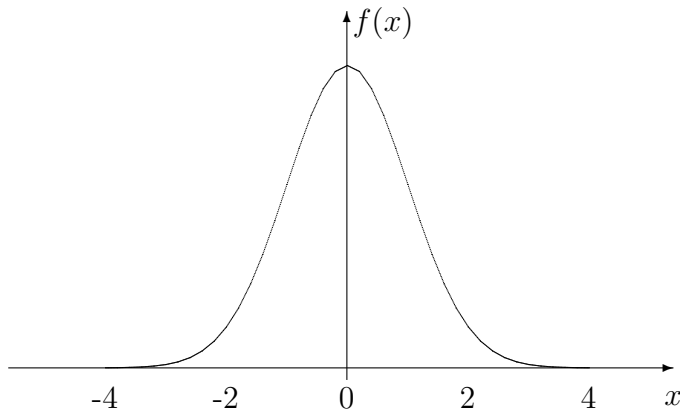
Význam: Pomocí hp lze určit pravděpodobnost, se kterou při realizaci náhodné veličiny bude její hodnota ležet v daném intervalu $\langle a, b \rangle$

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Speciálně pro $a = -\infty$ dostaneme pravděpodobnost, která se rovná hodnotě distribuční funkce v bodě b

$$P(-\infty \leq X \leq b) = \int_{-\infty}^b f(x) dx = F(b).$$

Platí: $\int_{-\infty}^{\infty} f(x) dx = 1.$



Obrázek 3.3: Graf normálního rozdělení

Dvourozměrná náhodná veličina

Popis: Dvourozměrná spojitá nv $[X, Y]$, nebo také dvourozměrný spojitý náhodný vektor $[X, Y]$, je uspořádaná dvojice spojitých náhodných veličin X a Y . Dvourozměrná spojitá hp je funkce dvou proměnných, která každé uspořádané dvojici (x, y) , kde $x \in X$ a $y \in Y$, přiřazuje nezáporné číslo. Zadává se zpravidla analytickým funkčním předpisem. Např. dvourozměrné rovnoměrné rozdělení je zadáno na obdélníkové oblasti $(a, b) \times (c, d)$ předpisem

$$f(x, y) = \frac{1}{(b-a)(d-c)} \chi_{(a,b) \times (c,d)}(x, y),$$

kde funkce χ_M je **indikátor** množiny M , tj.

$$\chi_M(a) = \begin{cases} 1 & \text{pro } a \in M, \\ 0 & \text{pro } a \notin M. \end{cases}$$

Význam: Pomocí hp lze určit pravděpodobnost obdélníkové podoblasti $(\alpha, \beta) \times (\gamma, \delta)$

$$P((\alpha \leq X \leq \beta) \wedge (\gamma \leq Y \leq \delta)) = \int_{\alpha}^{\beta} \int_{\gamma}^{\delta} f(x, y) \, dy \, dx.$$

Obecně lze pomocí hp určit pravděpodobnost libovolné ("rozumné") množiny $M \subset \mathbb{R}^2$

$$P([x, y] \in M) = \iint_M f(x, y) \, dx \, dy.$$

Platí:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1.$$

Marginální a podmíněná hp: Pro tyto hp platí integrální definiční vztahy z úvodu kapitoly (viz str. 39). Často jsme ale v situaci, kdy známe marginální hp $f(y)$ a podmíněnou hp $f(x|y)$, a potřebujeme vypočítat druhou marginální hp $f(x)$ a druhou podmíněnou hp $f(y|x)$. K tomu lze v obecném vzorci použít tzv. **řetězové pravidlo** $f(x, y) = f(x|y)f(y)$. Dostaneme pak

marginální hp

podmíněná hp

$$f(x) = \int_{-\infty}^{\infty} f(x|y)f(y) dy,$$

$$f(y|x) = \frac{f(x|y)f(y)}{\int_{-\infty}^{\infty} f(x|y)f(y) dy}.$$

Poznámka: Vztah pro výpočet podmíněné hp se nazývá **Bayesův vzorec**.

Praktická možnost, jak ze sdružené hp dostat marginální a podmíněnou hp, je rozložit sdruženou hp na součin dvou členů tak, aby první člen závisel jen na jedné proměnné a měl jednotkový integrál. Tento člen potom představuje marginální hp, zbývající člen je podmíněná hp. V případě normálního rozdělení, kdy v exponentu je kvadratická forma, lze pro rozklad použít doplnění na čtverec (viz 11.1). Tuto situaci ilustrujeme na následujícím příkladě.

Příklad [Dvourozměrné normální rozdělení]

Sdružená hustota pravděpodobnosti

Sdruženou hp je zvykem psát ve dvojitým různém vyjádření: skalárně a vektorově. Protože normální rozdělení je velmi časté a oba zápisy hp se podle okolností užívají, uvedeme obě verze.

Vektorový zápis:

$$f(z) = \frac{1}{2\pi\sqrt{|R|}} \exp\left(-\frac{1}{2}(z - \mu)^T R^{-1}(z - \mu)\right),$$

kde

$$z = \begin{bmatrix} x \\ y \end{bmatrix}, \quad \mu = E[z] = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix},$$

$\sigma_x^2 = E[(X - \mu_x)^2]$, $\sigma_y^2 = E[(Y - \mu_y)^2]$ jsou rozptyly X a Y ,

$\sigma_{xy} = E[(X - \mu_x)(Y - \mu_y)]$ je kovariance nv X a Y ,

$R = D[z] = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}$ je kovarianční matice,

$|R| = \sigma_x^2\sigma_y^2 - (\sigma_{xy})^2$ je determinant kovarianční matice R .

Skalární zápis:

$$f(z) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)} \left[\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho \frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right]\right\},$$

kde $\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y}$ je korelační koeficient, který je vždy v intervalu $\langle -1, 1 \rangle$.

Návod k odvození:

$$\begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix}^{-1} = \frac{1}{1-\rho^2} \begin{bmatrix} \frac{1}{\sigma_x^2} & -\frac{\rho}{\sigma_x\sigma_y} \\ -\frac{\rho}{\sigma_x\sigma_y} & \frac{1}{\sigma_y^2} \end{bmatrix}.$$

(konec příkladu)

Cvičení: Ověřit inverzi a souvislost obou vyjádření hp.

Marginální a podmíněná hustota pravděpodobnosti

Sdružená hp:

$$f(x, y) = \mathcal{N}(\mu, R), \quad \mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \quad R = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}.$$

Marginální hp:

$$f(x) = \mathcal{N}(\mu_x, \sigma_x^2), \quad f(y) = \mathcal{N}(\mu_y, \sigma_y^2).$$

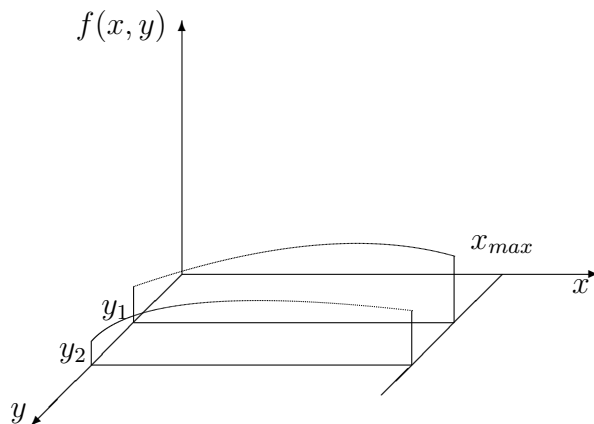
Podmíněná hp:

$$f(x|y) = \mathcal{N}\left(\mu_x + \rho\frac{\sigma_x}{\sigma_y}(y - \mu_y), \sigma_x^2(1 - \rho^2)\right),$$

$$f(y|x) = \mathcal{N}\left(\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x), \sigma_y^2(1 - \rho^2)\right).$$

Cvičení: Napsat hustoty a ověřit jejich tvar.Návod: doplnit na čtverec (viz 11.1), nejdříve zkusit pro $\mu = 0$.**3.4 Kombinované rozdělení**

Sdružené rozdělení dvou náhodných veličin, kde jedna nv je diskrétní a druhá spojitá, nazýváme **kombinovaným rozdělením**. Hp je potom funkce dvou proměnných, z nichž jedna je diskrétní a druhá spojitá. Příklad takové hp s první nv diskrétní a druhou spojitou je uveden na obr. 3.4.



Obrázek 3.4: Graf sdruženého kombinovaného rozdělení

Opět musí platit základní podmínka o normování pravděpodobnosti. Ta má pro hp na obrázku následující tvar:

$$\sum_{y \in \{y_1, y_2\}} \int_0^{x_{max}} f(x, y) dx = 1.$$

3.5 Operace s hustotami pravděpodobností

Uvedeme ty operace s hp, které budou dále základním pracovním nástrojem při odvození algoritmu bayesovského odhadu a řízení. Proto se jimi budeme zabývat poněkud podrobněji, než bývá zvykem. Budeme se snažit nejen formulovat jejich matematický smysl, ale také demonstrovat jejich praktickou interpretaci. Tu ukážeme na velmi jednoduchých reálných situacích. Nejprve poukážeme na jednotlivé pojmy, které se k takovým operacím vztahují.

Tyto pojmy budeme demonstrovat na následující situaci:

Mám jet z Prahy do Brna. Víím, že do Brna pojede buď řidič Hondy, nebo řidič Fiata (někdo určitě pojede, ale nepojedou současně). Přitom obě možnosti jsou stejně pravděpodobné. Pojede-li řidič Hondy, zastaví se pro mě s pravděpodobností 0.4, řidič Fiata s pravděpodobností 0.3. Je tedy možné, že ani jeden z řidičů se pro mne nezastaví.

Otázka A

S jakou pravděpodobností pojedou do Brna s řidičem Hondy?

Otázka B

S jakou pravděpodobností pojedou do Brna?

Otázka C

Před domem se pro mě zastavilo auto. S jakou pravděpodobností to byl řidič Hondy?

Řešení A

Kdybych věděl určitě, že do Brna pojede řidič Hondy, byla by míra mého očekávání v cestu Hondou rovna pravděpodobnosti 0.4. Avšak pravděpodobnost, že pojede Honda jen polovina (0.5), takže i míra mého očekávání je poloviční, a tedy hledaná pravděpodobnost je $0.4 \times 0.5 = 0.2$.

Řešení B

Na rozdíl od předchozího případu nás nyní zajímá, jaká je možnost, že mě někdo vůbec vezme, ať už Honda, nebo Fiat. Protože pojede právě jeden z řidičů, bude moje míra důvěry v úspěch při stopu dána mírou důvěry v úspěch při stopu od Hondy plus mírou důvěry v úspěch při stopu od Fiata. Hledaná pravděpodobnost je tedy $0.4 \times 0.5 + 0.3 \times 0.5 = 0.35$.

Řešení C

Moje očekávání úspěšného stopu od Hondy a od Fiata je v poměru 0.4 : 0.3. Protože jsem úspěch při stopu již dosáhl, musí být celková míra úspěšného stopu rovna jedné. Proto musí platit, že k úspěchu při stopu mi řidič Hondy přispěl s pravděpodobností $0.4/(0.4 + 0.3)$ a řidič Fiata s pravděpodobností $0.3/(0.4 + 0.3)$. Toto řešení vyhovuje, neboť poměr pravděpodobností je správný a jejich součet je jedna.

♣ Bayesovský přístup chápe pravděpodobnost jevu jako míru důvěry vložené do splnění jevu (z celkového množství jednotkové důvěry). Důvěru čerpáme ze dvou zdrojů:

1. z prvotní (apriorní) znalosti zkoumaného problému a
2. z dat, naměřených při experimentu.

Pravděpodobnosti kombinujeme s užitím pravidel pro hp.

V našem příkladě pravděpodobnosti jízdy jednotlivých řidičů, i pravděpodobnosti úspěšného stopu od jednotlivých řidičů vyjadřují naši apriorní znalost. Zjištění, že některé z aut jelo, představuje naměření dat.

První a druhá úloha je jen kombinací apriorní znalosti, třetí úloha již využívá informaci o naměřených datech (úspěch při stopu).

Nyní ukážeme pravděpodobnostní interpretaci jednotlivých pojmů. Definujeme jevy:

$A \equiv$ přijede řidič auta ..., $A \in \{h = \text{řidič Hondy}, f = \text{řidič Fiata}\}$,

$V \equiv$ výsledek stopu, $V \in \{u = \text{úspěšný stop}, n = \text{neúspěšný stop}\}$,

pravděpodobnosti jevů, které jsme zavedli jsou

$$P(A = h) = P(h) = 0.5, \quad P(A = f) = P(f) = 0.5.$$

Pravděpodobnosti hodnot jevu V zatím neznáme, nebyly zadány. O pravděpodobnostech úspěšného stopu se však v zadání hovoří. Pojede-li řidič Hondy, bude stop úspěšný s pravděpodobností 0.4, atd. To jsou pravděpodobnosti vázané na splnění určité podmínky. Tato podmínka je formulována za pomoci jevu A . Jedná se o podmíněné pravděpodobnosti, a ty se vztahují na podmíněné jevy – ”jestliže jede řidič Hondy, úspěch při stopu bude ...”.

Definujeme *podmíněné jevy*:

$V|A \equiv$ výsledek stopování za podmínky, že známe řidiče; $V|A \in \{u = \text{úspěšný stop}, n = \text{neúspěšný stop}\}$.

Např. $(V = u)|(A = h)$ znamená: výsledek stopování bude úspěch při stopu, jestliže vím, že jede řidič Hondy.

Pravděpodobnosti podmíněných jevů jsou zadány

$$P(V = u|A = h) = P(u|h) = 0.4, \quad P(V = u|A = f) = P(u|f) = 0.3,$$

$$P(V = n|A = h) = P(n|h) = 0.6, \quad P(V = n|A = f) = P(n|f) = 0.7.$$

Ve výše uvedených otázkách a při jejich řešení se vyskytovaly další pojmy. Je to především výrok ”úspěchu při stopování bude dosaženo od Hondy”. Pozor! Nezaměnit s výrokem ”bude úspěch při stopování, jestliže vím, že jede řidič Hondy”. Ve druhém výroku víme, že řidič Hondy jede, a předmětem zkoumání je jen to, zda přinesl při stopu úspěch nebo neúspěch. V prvním výroku nevíme ani jedno, ani druhé, a zajímá nás, zda obě tyto skutečnosti nastaly současně. Výrok ”úspěchu při stopování bude dosaženo od Hondy” je jedním z možných výsledků **sduženého jevu** (V, A) a jeho pravděpodobnost je **sdužená pravděpodobnost**. Výrok ”bude úspěch při stopování, jestliže vím, že jede řidič Hondy”, je výrokem podmíněným.

Dalším výrokem je výrok ”bude dosaženo úspěchu při stopování”, a je nám jedno, s čí pomocí. Tento výrok je výsledkem **marginálního jevu** V (vzhledem k předchozímu sduženému jevu (V, A)). Z tohoto hlediska je marginálním jevem také jev A .

Nakonec se objevuje výrok ”jede řidič Hondy, jestliže jsem dosáhl úspěchu při stopování”. To je opět výsledek podmíněného jevu, ale opačného k podmíněnému jevu, jehož pravděpodobnost je zadána v příkladě. Tento jev označíme $A|V$. Týká se pouze jména auta, protože výsledek jevu ”úspěch při stopování” je znám.

Jednotlivé otázky a jimi formulované úlohy jsou interpretacemi základních vztahů Bayesovské statistiky.

A. Řetězové pravidlo

$$P(V, A) = P(V|A)P(A), \tag{3.2}$$

tj. pravděpodobnost úspěchu při stopování s pomocí daného řidiče = pravděpodobnost úspěchu při stopování, jestliže jede daný řidič \times pravděpodobnost, že jede daný řidič.

Například $P(u, h) = P(u|h)P(h)$ udává pravděpodobnost úspěchu při stopování od Hondy a je dána jako součin pravděpodobnosti úspěchu při stopování za předpokladu, že jede řidič Hondy \times pravděpodobnost, že jede řidič Hondy.

B. Úplná pravděpodobnost

$$P(V) = \sum_A P(V|A)P(A), \quad \text{nebo podrobněji} \quad P(V) = \sum_{a \in A} P(V|A = a)P(A = a) \quad (3.3)$$

znamená: pravděpodobnost úspěchu při stopování = pravděpodobnost úspěchu při stopování, jede-li daný řidič \times pravděpodobnost, že jede tento daný řidič, a to sečteno přes všechny možné řidiče.

Pro konkrétní hodnoty má vzorec (3.3) tvar

$$P(u) = P(u|h)P(h) + P(u|f)P(f), \quad (3.4)$$

a zřejmě udává pravděpodobnost úspěšného stopování.

Poznámka: Předchozí výpočet je ve skutečnosti výpočtem marginální pravděpodobnosti $P(V) = \sum_A P(V, A)$. Jestliže sdruženou pravděpodobnost ještě rozložíme podle řetězového pravidla, dostaneme vztah, který se nazývá pravidlem úplné pravděpodobnosti.

C. Bayesův vzorec

$$P(A|V) = \frac{P(V|A)P(A)}{\sum_A P(V|A)P(A)}. \quad (3.5)$$

Tento vzorec udává pravděpodobnost, že jede daný řidič, jestliže víme, jaký byl úspěch při stopování. Pro výpočet se používají opačné podmíněné pravděpodobnosti, tj. pravděpodobnosti úspěchu při stopování za předpokladu, že jede daný řidič a pravděpodobnosti jízdy daných řidičů.

Pro konkrétní hodnoty má vzorec (3.5) tvar

$$P(h|u) = \frac{P(u|h)P(h)}{P(u|h)P(h) + P(u|f)P(f)}. \quad (3.6)$$

Poznámka: Jmenovatel Bayesova vzorce je vlastně marginální pravděpodobnost. Odvození Bayesova vzorce v tomto tvaru je zvláště jednoduché. Stačí dvojnásobkem rozložit sdruženou pravděpodobnost

$$P(A|V)P(V) = P(A, V) = P(V, A) = P(V|A)P(A),$$

a vyjádřit požadovanou podmíněnou pravděpodobnost.

Příklad [Operace s hp]

Na vybraném místě silniční komunikace byla sledována rychlost projíždějících vozidel. Bylo zjištěno, že 30% projíždějících vozidel má cizí SPZ a zbytek naši. Z cizích vozidel překročilo povolenou rychlost 20% vozidel, z našich 25%. Určete:

- A Pravděpodobnost, že právě projíždějící vozidlo je cizí a překračuje povolenou rychlost.
- B Pravděpodobnost, že právě projíždějící vozidlo překračuje povolenou rychlost.
- C Zjistili jsme, že právě projíždějící vozidlo překračuje povolenou rychlost. Jaká je pravděpodobnost, že je to cizí vozidlo?

Označíme jevy a jejich hodnoty:

$Z \equiv$ vozidlo má Značku ..., $Z \in \{c = \text{cizí}, n = \text{naší}\}$

$R \equiv$ vozidlo jede Rychlostí ..., $R \in \{v = \text{vyšší}, s = \text{správnou}\}$.

Pravděpodobnosti hodnot jevu Z jsou

$$P(c) = 0.3, \quad P(n) = 0.7.$$

Podmíněný jev je $R|Z$ a pravděpodobnosti jeho hodnot jsou

$$P(v|c) = 0.2, \quad P(v|n) = 0.25,$$

$$P(s|c) = 0.8, \quad P(s|n) = 0.75.$$

Řešení A

Ptáme se na sdruženou pravděpodobnost $P(Z = c, R = v)$, která je dána řetězovým pravidlem

$$P(Z, R) = P(R|Z)P(Z).$$

Po dosazení hodnot a jejich pravděpodobností dostaneme

$$P(c, v) = P(v|c)P(c) = 0.2 \times 0.3 = 0.06.$$

Řešení B

Hledáme marginální pravděpodobnost pomocí vzorce pro úplnou pravděpodobnost

$$P(R) = \sum_Z P(R|Z)P(Z).$$

Po dosazení

$$P(v) = P(v|c)P(c) + P(v|n)P(n) = 0.2 \times 0.3 + 0.25 \times 0.7 = 0.235.$$

Řešení C

Máme určit obráceně podmíněnou pravděpodobnost a k tomu použijeme Bayesův vzorec

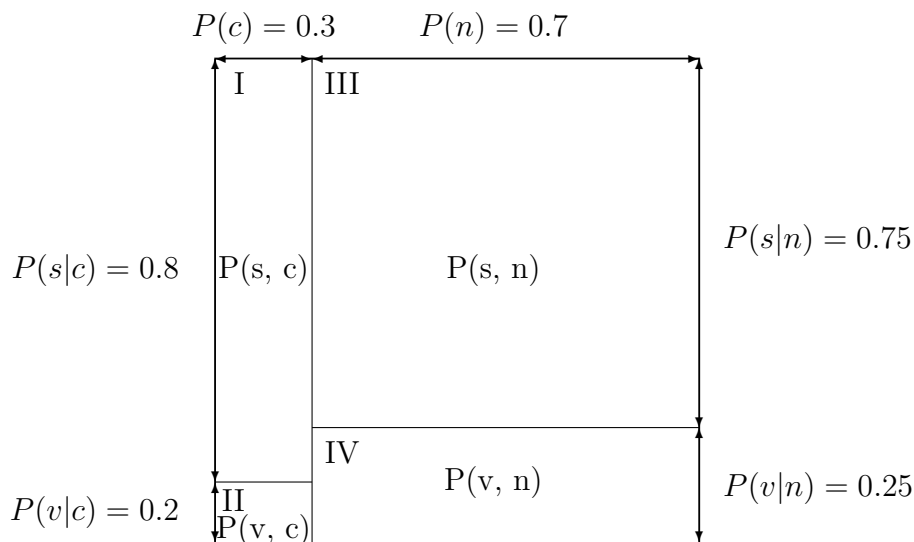
$$P(Z|R) = \frac{P(R|Z)P(Z)}{\sum_Z P(R|Z)P(Z)}.$$

Dosadíme a dostaneme

$$P(c|v) = \frac{P(v|c)P(c)}{P(v|c)P(c) + P(v|n)P(n)} = \frac{0.2 \times 0.3 + 0.25 \cdot 0.7}{0.2 \times 0.3 + 0.25 \cdot 0.7} \doteq 0.255.$$

(konec příkladu)

Strukturu předchozích operací budeme ještě ilustrovat pomocí grafických schémat. Použijeme dvojí vyjádření, vhodné pro demonstraci podmíněné pravděpodobnosti a operací s ní spojených. Jedná se o znázornění pravděpodobností pomocí obdélníkového schématu a pomocí pravděpodobnostního stromu.



Obrázek 3.5: Ilustrace operací s pravděpodobnostmi

Obdélníkové schéma

Při tomto zobrazení vycházíme z **řetězového pravidla**, které říká, že sdružená pravděpodobnost je dána jako součin podmíněné a marginální pravděpodobnosti. Zobrazíme-li tedy podmíněnou a marginální pravděpodobnost jako strany obdélníku, bude jeho plocha představovat sdruženou pravděpodobnost. Základní prostor pro sdruženou pravděpodobnost bude čtverec o straně jedna. V tomto základním čtverci lze zobrazovat jednotlivé typy pravděpodobností a demonstrovat operace mezi nimi.

V našem příkladě na vodorovné strany nanášíme pravděpodobnost (marginálního) jevu Z - značka a na svislé strany nanášíme pravděpodobnosti podmíněného jevu $R|Z$ - rychlost podmíněná značkou. Dostáváme schéma načrtnuté na obr. 3.5. Základní čtverec se tím rozdělí na čtyři obdélníky I. až IV., jejichž plošné obsahy udávají příslušné sdružené pravděpodobnosti. Skutečně, pro plošné obsahy jednotlivých obdélníků dostáváme

- I. pro jev "cizí auto nepřekročilo rychlost" $P(s, c) = P(s|c)P(c)$;
- II. pro jev "cizí auto překročilo rychlost" $P(v, c) = P(v|c)P(c)$;
- III. pro jev "naše auto nepřekročilo rychlost" $P(s, n) = P(s|n)P(n)$;
- IV. pro jev "naše auto překročilo rychlost" $P(v, n) = P(v|n)P(n)$.

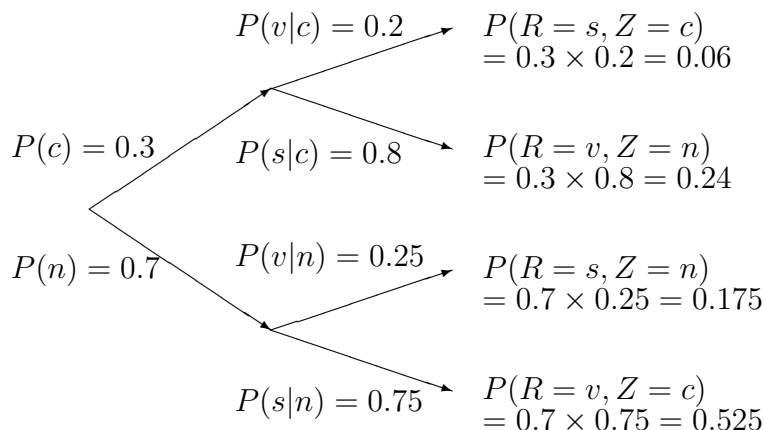
Z těchto pravděpodobností můžeme vystavět libovolnou pravděpodobnost, týkající se jevů Z a R . Na tomto místě bych rád citoval větu svého školitele a předního odborníka v oblasti bayesovské teorie řízení u nás Ing. V. Peterky, DrSc:

♣ Známe-li (tj. umíme-li zkonstruovat) sdruženou pravděpodobnost, známe všechno potřebné pro úlohy odhadu a řízení.

Marginální pravděpodobnost např. pro jev $R = v$ je dána součtem ploch všech obdélníků s boční stranou odpovídající hodnotě v . Jsou to obdélníky II. a IV. Vyjádříme-li jejich plochy jako součiny stran a sečteme je, dostáváme vzorec pro **úplnou pravděpodobnost**.

Obrácená **podmíněná pravděpodobnost** např. jevu $Z|R = c|v$ se odvodí takto: Protože jev $R = v$ nastal, musí mít pravděpodobnost 1. To znamená, že musíme definovat nový základní

čtverec pro jevy $Z|R = v \in \{c|v, n|v\}$, a přitom poměr pravděpodobností hodnot $c|v$ a $n|v$ by měl být dán poměrem příslušných sdružených pravděpodobností jevů (c, v) a (n, v) . Tedy $P(c|v) : P(n|v) = P(v, c) : P(v, n) = 0.06 : 0.175$. Tento poměr je dán obsahy dolních obdélníků. Stačí tedy jejich obsahy normovat tak, aby součet byl jedna a obdržíme příslušné podmíněné pravděpodobnosti. Ta, kterou zjišťujeme, je dána levým obdélníkem. Její normování provedeme tak, že plochu tohoto obdélníku dělíme součtem ploch obou dolních obdélníků. Tomu při vyjádření ploch součinem stran obdélníků odpovídá Bayesův vztah.



Obrázek 3.6: Pravděpodobnostní strom

Pravděpodobnostní strom

Jiným nástrojem nejen pro popis, ale i pro výpočet sdružených pravděpodobností pomocí podmíněných a marginálních je pravděpodobnostní strom. Je také založen na řetězovém pravidlu a je dokonce vhodný i pro jeho vícenásobné použití. Strom začíná v jednom bodě a postupně se větví podle hodnot jednotlivých jevů. Hodnoty předchozích jevů při větvení se přitom dostávají do podmínek následujících jevů.

V našem příkladě bude strom velmi jednoduchý. První větvení je podle hodnot jevu Z na větve $Z = c$ a $Z = n$. K větvím zároveň připisujeme pravděpodobnosti těchto jevů. Na konci obou větví dojde k dalšímu větvení, podle hodnot jevu R . Protože ale konci horní větve přísluší $Z = c$, budou další větve z tohoto místa příslušet podmíněným jevům $R|(Z = c)$. Podobně větvení ze spodního bodu bude reprezentovat podmíněné jevy $R|(Z = n)$. Na konci dostáváme 4 body – konce větvení. Každému bodu přiřadíme pravděpodobnost rovnu součinu jednotlivých pravděpodobností od začátku větvení. Podle řetězového pravidla konce větvení odpovídají jednotlivým hodnotám pravděpodobnosti sdruženého jevu (Z, R) . A podobně jako v předchozím případě: známe-li sdružené pravděpodobnosti, známe všechno.

Kapitola 4

Model jako hustota pravděpodobnosti

Bayesovský přístup k odhadování a řízení je založen na **modelu soustavy**, který představuje základní prvek pro popis chování modelované veličiny [3]. Umožňuje určit hustotu pravděpodobnosti (dále jen hp) výstupu soustavy y_t na základě známých veličin (starších dat) a konstant (parametrů modelu). Předpověď je potom základem pro další odhadování i řízení.

Nejčastěji používaný lineární model je **regresní model** (2.2). Ten se skládá ze dvou základních částí. První část je deterministická. Ta je dána skalárním součinem **regresního vektoru**, který obsahuje známá data, a **vektoru regresních koeficientů**. Druhá část představuje náhodnou část modelu. Je to náhodná veličina se specifikovaným rozdělením (většinou se uvažuje normální rozdělení), která se nazývá **šum**. Parametry této náhodné veličiny (většinou je to rozptyl) spolu s regresními koeficienty tvoří **parametry regresního modelu**. Regresní model tedy představuje jakýsi ideální deterministický model, který je pod vlivem poruch (šumu), jak je naznačeno na obrázku 2.3.

To, co jsme právě uvedli, platí však pouze tehdy, jestliže známe parametry modelu. Tento předpoklad však většinou není splněn a parametry je třeba odhadovat z apriorní znalosti a z měřených dat, viz odstavec 5.3.

Model tedy obecně obsahuje dva neurčité prvky – šum a neznámé parametry. Oba tyto objekty bayesovská statistika chápe jako náhodné veličiny, které jsou popsány svými hp. Přitom mohou být buď diskrétní (tj. s konečným/spočetným počtem hodnot), nebo spojité (s reálnými hodnotami). Ačkoliv rozdíl ve vyjádření takových modelů je dán jen rozdílem ve vyjádření hp diskrétní a spojité náhodné veličiny, některé z těchto modelů mohou působit nezvykle. Proto ukážeme na příkladech jednotlivé typy modelů a v další kapitole také jejich použití pro odhadování a dále pro řízení.

Značení

I v této kapitole budeme používat symbolické značení pro pravděpodobnosti. Tak např. pro pravděpodobnost daného náhodného jevu X místo podrobného označení $P(X = x)$ budeme používat stručnější zápis $P(x)$. Podobně budeme používat stručnější symboliku i pro hp popisující signály nebo parametry. Taková jednodušší symbolika je v literatuře běžně používaná, zvyšuje přehlednost, ale někdy může vést k nedorozumění. Zjednodušení spočívá v tom, že ve značení přestaneme rozlišovat mezi náhodnou veličinou X a její hodnotou x . Budeme používat malá písmena pro signály a malá nebo velká řecká písmena pro parametry. Zároveň u hp nebudeme vyznačovat indexem náhodnou veličinu a bude se rozumět, že hp náleží té náhodné

veličině, jejíž hodnota se vyskytuje v argumentu. Tedy např. symbolem $f(y_t|u_t)$ nahrazujeme komplikovanější symbol $f_{Y_t|U_t}(y_t|u_t)$.

P o z n á m k a: Signály jsou náhodné posloupnosti, tedy indexované náhodné veličiny. Proto se v našich úvahách u náhodných veličin popisujících signály vyskytují indexy.

4.1 Regresní model

Rovnice modelu

V obecném tvaru je rovnice regresního modelu součtem dvou členů - deterministického (m) a stochastického (e)

$$y_t = m(\varphi_t, \theta) + e_t, \quad (4.1)$$

kde

y_t je modelovaná veličina – výstup soustavy,

$\varphi_t = [u_t, y_{t-1}, u_{t-1}, \dots, y_{t-\nu}, u_{t-\nu}, 1]^T$ je regresní vektor, obsahující aktuální řídicí veličinu u_t , data zpožděná až do ν kroků zpět (ν je řád modelu) a jedničku pro respektování konstanty modelu,

$\theta = [b_0, a_1, b_1, \dots, a_\nu, b_\nu, c]^T$ je vektor regresních koeficientů (známé nebo odhadnuté konstanty), jejichž pořadí koresponduje s pořadím dat v regresním vektoru,

$m(\varphi_t, \theta)$ je funkce dat z regresního vektoru φ_t a parametrů θ , která pro známé parametry představuje deterministickou transformaci dat z regresního vektoru na modelovaný výstup soustavy. Tato funkce je nejčastěji lineární $m(\varphi_t, \theta) = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + \dots + a_\nu y_{t-\nu} + b_\nu u_{t-\nu} + c$. Pak dostaneme **lineární regresní model** ve tvaru

$$y_t = \varphi_t^T \theta + e_t, \quad (4.2)$$

e_t je náhodná veličina (šum), o které předpokládáme, že pro všechna t platí

- je nekorelovaná se svými zpožděnými hodnotami $E[e_t|e_{t-i}] = E[e_t]$ pro $i = 1, 2, \dots$ a s daty obsaženými v regresním vektoru φ_t , tj. $E[e_t|\varphi_t] = E[e_t]$,
- má nulovou střední hodnotu $E[e_t] = 0$ a konstantní rozptyl $D[e_t] = \sigma^2 = \text{konst.}$,

Symbolem Θ označíme souhrn všech parametrů modelu, tj. symbolicky píšeme

$$\Theta = \{\theta, \sigma^2\}.$$

Tento zápis budeme chápat ve smyslu cell-vektoru, zavedeného toolboxem MATLAB, pro vektor (strukturu) s prvky nestejného typu.

Hp modelu

Z rovnice modelu (4.1) vidíme, že y_t závisí na regresním vektoru φ_t , regresních koeficientech θ a rozptylu šumu σ^2 , takže i rozdělení náhodné veličiny y_t bude dáno podmíněnou hp

$$f(y_t|\varphi_t, \Theta). \quad (4.3)$$

Jde skutečně o podmíněnou hp, protože jedině při znalosti veličin φ_t a Θ jsme schopni rovnici (4.1) pro předpověď výstupu y_t využít. Pro regresní vektor φ_t a parametry θ umíme vypočítat hodnotu funkce $m(\varphi_t, \theta)$. Hp výstupu y_t je pak dána posunutím hp náhodné veličiny e_t (šumu) do nové střední hodnoty dané hodnotou funkce $m(\varphi_t, \theta)$.

Poznámka: Tato transformace odpovídá základní transformaci náhodné veličiny do normovaného tvaru $Z \rightarrow X$ pomocí transformační rovnice

$$X = \mu + \sigma Z,$$

kteřá spočívá v tom, že v hp náhodné veličiny dosadíme za proměnnou z podíl $\frac{x - \mu}{\sigma}$, jak plyne z transformační rovnice, a celou hp vynásobíme Jakobíánem $1/\sigma$ této transformace.

Charakteristiky modelu

Hp modelu určuje charakteristiky výstupu y_t . Tyto charakteristiky jsou rovněž podmíněné, protože jsou určeny pomocí podmíněné hp modelu.

Střední hodnota y_t :

$$E[y_t|\varphi_t, \Theta] = E[m(\varphi_t, \theta)|\varphi_t, \Theta] + E[e_t|\varphi_t, \Theta] = m(\varphi_t, \theta). \quad (4.4)$$

Rozptyl y_t :

$$D[y_t|\varphi_t, \Theta] = D[y_t|\Theta] = \sigma^2. \quad (4.5)$$

Druhý obecný moment y_t :

$$E[y_t^2|\varphi_t, \Theta] = (E[y_t|\varphi_t, \Theta])^2 + D[y_t|\Theta] = (m(\varphi_t, \theta))^2 + \sigma^2. \quad (4.6)$$

Poznámka: Uvedené charakteristiky jsou podmíněné. Samozřejmě existují i nepodmíněné, ty ale nejsme schopni určit jen pomocí modelu, a musíme je konstruovat jinak. Na první pohled je např. zřejmé, že při jejich konstrukci bude zapotřebí odhadovat parametry, tj. provádět úlohu odhadu. Model vyžaduje znalost parametrů, ty ale v praxi většinou známé nejsou.

Typy regresních modelů

Uvedené vztahy pro rovnici modelu a jeho charakteristiky jsou zcela obecné. Model je konstruován jako součet deterministické funkce a náhodné veličiny, přičemž se předpokládá, že náhodná veličina e_t je nekorelovaná s daty obsaženými v regresním vektoru φ_t . Takto definovaný model se nazývá **obecný regresní model**. Podle konkrétní specifikace funkce m , případně rozdělení náhodné veličiny e_t , dostáváme některý z následujících typů regresních modelů.

- Je-li funkce m lineární vzhledem k parametrům, potom má model tvar (4.2), srv. též (2.2), a nazývá se **lineární regresní model**. V případě, že m není lineární, nazývá se **nelineární regresní model**.
- Jestliže regresní vektor obsahuje také zpožděné modelované veličiny, tj. y_{t-1}, y_{t-2}, \dots , jde o **dynamický regresní model**. V opačném případě, je-li např. výstup modelován jako konstanta plus šum, nebo řízení plus šum, jde o **statický regresní model**.

- Má-li náhodná veličina e_t normální rozdělení, nazýváme model **normálním regresním modelem**. Má-li jiné rozdělení, je jméno modelu odvozeno podle jména rozdělení. Např. rovnoměrný regresní model, nebo také regresní model s rovnoměrným šumem.
- Model může pracovat se spojitými, diskrétními nebo kombinovanými veličinami. V případě spojitých veličin nazýváme model **spojitým regresním modelem**, jinak **diskrétním regresním modelem**.
- Jsou-li parametry θ konstanty, nazýváme tento model **regresní model s konstantními koeficienty**. Jsou-li funkcemi času, jde o **regresní model s proměnnými koeficienty**.
- Jsou-li signály, se kterými model pracuje, vícerozměrné, tj. v každém časovém okamžiku tvoří vektor, nazývá se takový model **mnohorozměrný regresní model**. Jsou-li signály skalární veličiny, nazývá se **skalární regresní model**.

My se budeme zabývat skalárními lineárními regresními modely s konstantními koeficienty.

4.1.1 Dynamický regresní model s normálním šumem

Jde o skalární lineární dynamický regresní model řádu ν s konstantními koeficienty a normálním rozdělením šumu (4.2). Při studiu lineárních modelů je takový model používán nejčastěji. Vyjádření střední hodnoty pomocí lineární funkce a normální rozdělení šumu pokrývá mnoho praktických úloh a má i značné výhody z hlediska jeho použití pro odhad a při řízení. Je to stochastická obdoba diferenční rovnice.

Rovnice modelu

$$y_t = \varphi_t^T \theta + e_t, \quad (4.7)$$

kde

$\varphi_t^T = [u_t, y_{t-1}, u_{t-1}, \dots, y_{t-\nu}, u_{t-\nu}, 1]$ je regresní vektor,

$\theta^T = [b_0, a_1, b_1, \dots, a_\nu, b_\nu, c]$ jsou regresní koeficienty,

e_t je skalární náhodná veličina s rozdělením $N(0, \sigma^2)$, tj.

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

kde $\exp\{x\} = e^x$ značí exponenciální funkci.

Model lze rovněž rozepsat do tvaru diferenční rovnice

$$y_t = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + \dots + a_\nu y_{t-\nu} + b_\nu u_{t-\nu} + c + e_t. \quad (4.8)$$

Momenty modelu

Model umožňuje najít rozdělení, nebo, je-li třeba, charakteristiky modelovaného výstupu y_t .

Střední hodnota a rozptyl:

$$E[y_t | \varphi_t, \Theta] = \varphi_t^T \theta, \quad D[y_t | \varphi_t, \Theta] = \sigma^2. \quad (4.9)$$

Druhý obecný moment:

$$E[y_t^2 | \varphi_t, \Theta] = E[y_t | \varphi_t, \Theta]^2 + D[y_t | \varphi_t, \Theta] = (\varphi_t^T \theta)^2 + \sigma^2. \quad (4.10)$$

Hp modelu

Má stejný tvar jako hp šumu, ale se střední hodnotou $\varphi_t^T \theta$. Je tedy

$$f(y_t | \varphi_t, \Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_t - \varphi_t^T \theta)^2 \right\} \quad (4.11)$$

Poznámka: Prakticky transformaci provedeme tak, že z rovnice modelu vyjádříme e_t a dosadíme do hp šumu. Jakobián transformace je v tomto případě roven jedné.

4.1.2 Statický regresní model s normálním šumem**A. Řízený systém**

Statický model je speciálním případem dynamického regresního modelu (4.1), kdy regresní vektor neobsahuje žádné zpožděné výstupy. Regresní vektor může tedy obsahovat pouze řídicí veličinu a případně jedničku pro respektování konstanty.

Nejprve budeme uvažovat model bez konstanty, kde modelovaná veličina je úměrná řídicí veličině s neznámým koeficientem úměrnosti – neznámým parametrem modelu θ . Pro $u_t = 1$ pro všechna t dostaneme speciální případ odhadování střední hodnoty.

Navíc budeme předpokládat známý rozptyl šumu, např. $\sigma^2 = 1$. Tento model bude mít tedy jediný vstup u_t a jediný neznámý parametr θ , tj. $\Theta = \{\theta\}$.

Rovnice modelu

$$y_t = \theta u_t + e_t, \quad (4.12)$$

kde

θ je konstantní skalární parametr,

y_t , resp. u_t je výstup, resp. vstup soustavy,

e_t je náhodná veličina s normálním rozdělením s nulovou střední hodnotou a jednotkovým rozptylem, nezávislá na u_t .

Poznámka: Není-li dáno rozdělení náhodných veličin e_t , pak (4.12) nepopisuje situaci úplně.

Momenty modelu

Jsou to střední hodnota θu_t , rozptyl 1 a druhý obecný moment $\theta^2 u_t^2 + 1$.

Hp modelu

$$\begin{aligned} f(y_t | u_t, \theta) &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} (y_t^2 - 2\theta y_t u_t + \theta^2 u_t^2) \right\} = \\ &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \begin{bmatrix} -1 & \theta \end{bmatrix} \begin{pmatrix} y_t \\ u_t \end{pmatrix} \begin{bmatrix} y_t & u_t \end{pmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}. \end{aligned} \quad (4.13)$$

Exponent prvního vyjádření hp přímo ukazuje, jaké budou statistiky pro odhadování. Budou to kvadráty y_t^2 , u_t^2 a součiny $y_t u_t$ pro t na časovém intervalu odhadování. Druhé vyjádření reprezentuje model jako funkci neznámých parametrů a je vhodné pro konstrukci věrohodnostní funkce (5.5).

B. Neřízený systém

Ještě jednodušší variantou tohoto modelu je model neřízeného systému, který obdržíme, jestliže v předchozím modelu (4.12) položíme $u_t = 1$ pro všechna t .

Rovnice modelu

$$y_t = \theta + e_t, \quad e_t \sim N(0, \sigma^2). \quad (4.14)$$

Hp modelu

Pro známý rozptyl, např. $\sigma^2 = 1$, je

$$f(y_t|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y_t^2 - 2\theta y_t + \theta^2)\right\} = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y_t - \theta)^2\right\}. \quad (4.15)$$

Poznámka: V případě, kdy by rozptyl šumu σ^2 nebyl znám, měl by šum rozdělení $N(0, \sigma^2)$ a hp modelu by byla

$$f(y_t|\Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_t - \theta)^2\right\}. \quad (4.16)$$

Neznámé parametry v tomto případě tvoří dvojici $\Theta = \{\theta, \sigma^2\}$.

4.1.3 Statický regresní model s rovnoměrným šumem

Pro vyjádření hp rovnoměrného rozdělení budeme používat tzv. **indikátor** χ_I intervalu $I = \langle a, b \rangle \subset R$ (srovnej str. 43)

$$\chi_I(x) = \begin{cases} 1 & \text{pro } x \in I, \\ 0 & \text{pro } x \notin I, \end{cases}$$

kde argument x , pokud nemůže dojít k omylu, budeme vynechávat.

Rovnice modelu

$$y_t = \theta + e_t, \quad (4.17)$$

kde

θ je neznámý spojitý parametr,

e_t je náhodná veličina s rovnoměrným rozdělením s hp

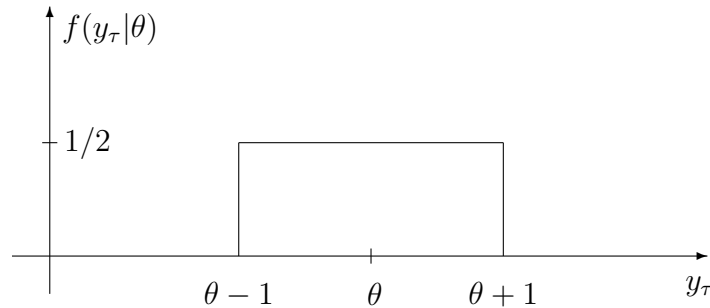
$$f(e_t) = \frac{1}{2} \chi_{(-1,1)}(e_t).$$

Poznámka: Pro jednoduchost uvažujeme hp šumu se známými mezemi od -1 do 1. Pokud meze pro hp nebudou jednotkové, ale budou známé, je situace obdobná. Složitější model dostaneme v případě, kdy tyto meze nejsou známé a představují další parametry, které je třeba odhadovat. Tím získáme analogii případu s neznámou směrodatnou odchylkou.

Hp modelu

Má také rovnoměrné rozdělení, které opět dostaneme transformací hp šumu

$$f(y_t|\theta) = \frac{1}{2}\chi_{(\theta-1,\theta+1)}(y_t), \quad y_t \in R. \quad (4.18)$$



Obrázek 4.1: Graf rovnoměrného rozdělení

4.1.4 Statický regresní model s diskretním šumem

V tomto odstavci se budeme zabývat modely, popisujícími diskretní signály, tj. signály, které mohou nabývat jen konečného počtu hodnot. Typickým příkladem takového signálu je výsledek pokusu při hodu mincí nebo kostkou, ale patří sem např. také intenzita dopravního proudu (počet vozidel za jednotku času) nebo číslo řídicího plánu světelné signalizace v dané mikrooblasti (v Praze se používají plány č. 1, 2, ..., 10). Parametry takového modelu bývají obvykle spojité. Modely s diskretními parametry se budeme pro ukázkou zabývat v části 4.2.

Zde uvedeme model diskretního signálu s dvěma možnými hodnotami pro známé a neznámé parametry modelu.

A. Známý parametr modelu

Jedná se snad o nejjednodušší statistický model, popisující např. házení mincí. Abychom tento příklad udělali trochu zajímavější, uvažujeme poškozenou minci, kde pravděpodobnost padnutí rubu se nerovná pravděpodobnosti padnutí líce. Předpokládáme ale, že pravděpodobnost padnutí rubu (a tedy i líce) je známa (např. z dřívějších experimentů) a rovná se $\theta_0 \in (0, 1)$ pro rub a $1 - \theta_0$ pro líc.

Uvažujeme posloupnost nezávislých hodů v časech $\tau = 1, 2, \dots, t$, s výsledky $V_\tau = R$ (rub) nebo $V_\tau = L$ (líc). V okamžicích τ definujeme náhodné veličiny y_τ předpisem

$$y_\tau = \begin{cases} 0 & \text{pro } V_\tau = R, \\ 1 & \text{pro } V_\tau = L. \end{cases}$$

S tímto označením dostáváme následující model.

Rovnice modelu

Je triviální

$$y_t = e_t, \quad (4.19)$$

kde hp šumu formálně zapíšeme ve tvaru alternativního rozdělení s parametrem θ_0 , tj.

$$f(e_t) = \theta_0^{1-e_t}(1 - \theta_0)^{e_t}$$

pro $e_t \in \{0, 1\}$.

Hp modelu

Je shodná s hp šumu

$$f(y_t) = \theta_0^{1-y_t}(1 - \theta_0)^{y_t} \quad (4.20)$$

pro $y_t \in \{0, 1\}$.

Poznámka: V tomto případě je zřejmě zbytečné zavádět model jako rovnici. Lépe je o modelu rovnou hovořit jako o příslušné hp.

B. Neznámý parametr modelu

Tento model je obdobný jako předchozí, ale předpokládáme, že pravděpodobnost padnutí rubu (a tedy i líce) není známa. Takový model má velmi dobrou dopravní interpretaci, kterou jsme se zabývali v úvodním příkladu 1.1. Při ní uvažujeme křižovatku ve tvaru T, do které přijíždějí automobily a odbočují buď vpravo nebo vlevo. Evidujeme počty vozidel a zjišťujeme procento automobilů odbočujících do jednotlivých směrů. Na základě takto naměřených dat potom odhadujeme skutečný podíl odbočujících vozidel. Jedná se o nejjednodušší verzi odhadu směrových vztahů na křižovatce.

Rovnice modelu

Je shodná s rovnicí z předchozího příkladu. Jak jsme již uvedli, v takto jednoduchém příkladě není zajímavá.

Hp modelu

Je to podmíněná hp s alternativním rozdělením. Závisí na jednom neznámém parametru $\theta \in \langle 0, 1 \rangle$ a má tvar

$$f(y_t|\theta) = \theta^{1-y_t}(1 - \theta)^{y_t}, \quad y_t \in \{0, 1\}. \quad (4.21)$$

Odtud $f(0|\theta) = P(y_t = 0|\theta) = \theta$, $f(1|\theta) = P(y_t = 1|\theta) = 1 - \theta$, takže pravděpodobnost, s níž padne rub ($y_t = 0$), je θ a pravděpodobnost, s níž padne líc ($y_t = 1$), je $1 - \theta$.

4.1.5 Mnohorozměrný regresní model

Mnohorozměrným odhadem a řízením se zde podrobně nezabýváme. Protože však ve většině praktických úloh je třeba modelovat několik veličin, a protože mnohorozměrný model je možno převést na soustavu tzv. jednorozměrných faktorů, uvedeme pro úplnost také zápis mnohorozměrného regresního modelu. Pro hlubší studium algoritmů s mnohorozměrnými regresními modely lze doporučit např. [3].

Uvažujeme model s n_u vstupy a n_y výstupy, tj. vstup, výstup i šum jsou vektory odpovídající délky

$$u_t^T = [u_{1,t}, u_{2,t}, \dots, u_{n_u,t}], \quad y_t^T = [y_{1,t}, y_{2,t}, \dots, y_{n_y,t}], \quad e_t^T = [e_{1,t}, e_{2,t}, \dots, e_{n_y,t}].$$

Rovnice modelu

Má obdobný vektorový tvar jako rovnice skalárního modelu, pouze pořadí parametru θ a regresního vektoru φ_t musíme ve skalárním součinu formálně přehodit, abychom modelovaný výstup obdrželi jako sloupcový vektor

$$y_t = \theta^T \varphi_t + e_t, \quad (4.22)$$

kde $\theta^T = [b_0, a_1, b_1, \dots, a_\nu, b_\nu, c]$, kde

a_i, c jsou matice typu $(n_y \times n_y)$ a b_i jsou matice typu $(n_y \times n_u)$,

$\varphi_t^T = [u_t^T, y_{t-1}^T, u_{t-1}^T, \dots, y_{t-\nu}^T, u_{t-\nu}^T, 1]$ je vektor délky $n_u + \nu(n_y + n_u) + 1$,

e_t je n_y -rozměrná náhodná veličina s rozdělením s nulovou střední hodnotou a kovarianční maticí R .

Střední hodnota modelu je vektor $\mu_y = \theta^T \varphi_t$ a neurčitost vyjadřuje kovarianční matice R .

Hp modelu

Je to hp mnohorozměrného normálního rozdělení se střední hodnotou μ a kovarianční maticí R

$$f(y_t | \varphi_t, \Theta) = (2\pi)^{-\frac{n_y}{2}} |R|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y_t - \theta^T \varphi_t)^T R^{-1} (y_t - \theta^T \varphi_t) \right\}. \quad (4.23)$$

Příklad [Rovnice modelu pro $n_y = 2$, $n_u = 1$, $C = 0$, $\nu = 1$]

Modelujeme 2 veličiny měřené na soustavě s jedním vstupem pomocí dynamického modelu řádu 1. Rovnice modelu je

$$\begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} b_{1,0} \\ b_{2,0} \end{bmatrix} u_t + \begin{bmatrix} a_{11;1} & a_{12;1} \\ a_{21;1} & a_{22;1} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} b_{1;1} \\ b_{2;1} \end{bmatrix} u_{t-1} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

Parametry a regresní vektor

$$\theta^T = \begin{bmatrix} b_{1;0} & a_{11;1} & a_{12;1} & b_{1;1} \\ b_{2;0} & a_{21;1} & a_{22;1} & b_{2;1} \end{bmatrix}, \quad \varphi_t = \begin{bmatrix} u_t \\ y_{1,t-1} \\ y_{2,t-1} \\ u_{t-1} \end{bmatrix}$$

Střední hodnota a kovarianční matice šumu

$$E[e_t] = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \quad E[e_t^T e_t] = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = R$$

(konec příkladu)

4.2 Modely s diskrétními parametry

Neznámé parametry modelu považujeme za náhodné veličiny popsané svými hp. Pokud tyto náhodné veličiny mohou nabývat jen konečného počtu předem specifikovaných hodnot, jedná se o diskrétní náhodné veličiny – viz str. 40. Odpovídající model nazveme modelem s diskrétními parametry.

V dalším výkladu budeme uvažovat statický regresní model (viz str. 55 – typy modelů) popsany pomocí hp šumu $f(e)$ a parametru $f(\theta)$, které předpokládáme známé. Na základě znalosti modelu v daném čase provedeme předpověď výstupu, a po změření skutečné hodnoty výstupu přepočteme hodnotu odhadovaného parametru. Protože se zabýváme jen jedním krokem odhadování, nebudeme u signálů uvádět časový index.

A. Diskrétní parametr a diskrétní signál

Rovnice modelu

$$y = \theta + e, \quad (4.24)$$

kde hp $f(e)$ a $f(\theta)$ jsou dány tabulkami, viz část 3.2

$$\begin{array}{c|cc} e & -1 & 1 \\ \hline f(e) & 0.4 & 0.6 \end{array}, \quad \begin{array}{c|ccc} \theta & -2 & 0 & 2 \\ \hline f(\theta) & 0.2 & 0.5 & 0.3 \end{array}.$$

Obě náhodné veličiny jsou nezávislé.

Hp modelu

Je to funkce $f(y|\theta)$ zadaná opět pomocí tabulky, viz druhý odstavec v části 3.2. Ve sloupcích tabulky jsou modely vždy pro určitou hodnotu parametru v podmínce.

$$\begin{array}{c|ccc} y \backslash \theta & -2 & 0 & 2 \\ \hline -3 & 0.4 & 0 & 0 \\ -1 & 0.6 & 0.4 & 0 \\ 1 & 0 & 0.6 & 0.4 \\ 3 & 0 & 0 & 0.6 \end{array} \quad (4.25)$$

Tuto tabulku sestavíme z tabulek pro $f(e)$ a $f(\theta)$ takto: Nejdříve určíme všechny možné hodnoty výstupu y tak, že pro danou hodnotu θ dosazujeme do rovnice modelu jednotlivé hodnoty e . Potom určíme pravděpodobnosti tak, že pro dané hodnoty θ a y nalezneme pomocí (4.24) hodnotu e , a vezmeme její pravděpodobnost. Pravděpodobnosti hodnot θ se neuplatní. Proč? ¹

Ve sloupcích tabulky udávající hp modelu jsou jednotlivé hp výstupu (jejichž pravděpodobnosti se rovnají pravděpodobnostem šumu) vždy pro jednu konkrétní hodnotu θ v podmínce, tedy vektory $f(y|\theta = -2)$, $f(y|\theta = 0)$ a $f(y|\theta = 2)$, případně doplněné nulami.

B. Diskrétní parametr a spojitý signál

Rovnice modelu

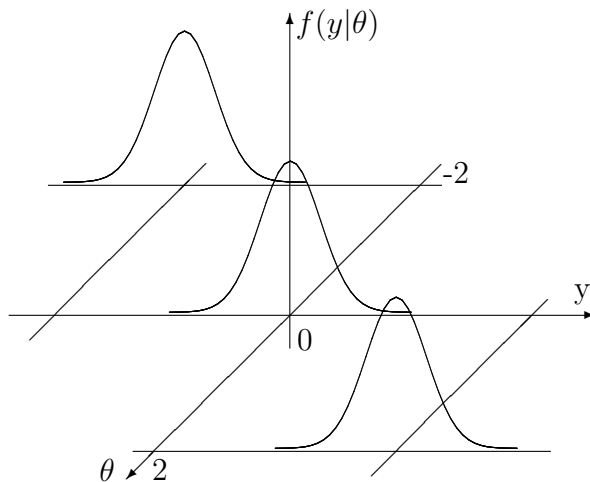
Má opět tvar

$$y = \theta + e, \quad (4.26)$$

kde θ je diskrétní a e spojitá náhodná veličina s rozděleními – viz (3.2 a (3.3)

$$\begin{array}{c|ccc} \theta & -2 & 0 & 2 \\ \hline f(\theta) & 0.2 & 0.5 & 0.3 \end{array}, \quad f(e) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}e^2\right\}.$$

¹Protože modelem je hp podmíněná parametrem θ . Ten zde tedy vystupují ne jako náhodná veličina, ale prostřednictvím svých hodnot, jako by byly známé. Proto model nezávisí na pravděpodobnostech těchto hodnot.



Obrázek 4.2: Hustota kombinovaného modelu

Hp modelu

Dostaneme ji jako transformaci nv e na nv y

$$f(y|\theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y - \theta)^2 \right\}, \quad \theta \in \{-2, 0, 2\}. \quad (4.27)$$

Tento kombinovaný model můžeme interpretovat jako tři spojité modely pro jednotlivé hodnoty parametru θ

$$\begin{aligned} f(y|\theta = -2) &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y + 2)^2 \right\} && \text{pro } \theta = -2, \\ f(y|\theta = 0) &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}y^2 \right\} && \text{pro } \theta = 0, \\ f(y|\theta = 2) &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y - 2)^2 \right\} && \text{pro } \theta = 2. \end{aligned}$$

Kapitola 5

Předpovídání a odhadování

5.1 Úvodní poznámky a označení

V této kapitole se budeme věnovat odhadování pravděpodobnostních modelů. Ta je těsně spjata, podobně jako i řízení, s předpovědí budoucího výstupu soustavy. Pojednáme tedy o předpovědi a odhadování společně. Základní literaturou k této oblasti je [3].

Nejdříve uvedeme dvě **důležité poznámky**, na které se budeme v dalším textu často odvolávat.

Značení integrálů a datové historie

Bayesovská statistika pracuje se sdruženými, marginálními a podmíněnými hp. Při výpočtech s nimi se často používá integrace – viz str. 39. Jedná se o integraci přes celý obor hodnot náhodné veličiny nebo náhodných veličin, které příslušná hp popisuje. Jsou to tedy integrály určité. Pro lepší přehlednost nebudeme vyznačovat jejich meze a zavedeme následující dohodu: Nejsou-li vyznačeny meze integrace a není-li uvedeno jinak, máme na mysli integraci přes celý obor hodnot náhodné veličiny. Např. pro normální rozdělení je to celá reálná osa, nebo pro sdruženou hp pro n náhodných veličin je to celý prostor \mathbb{R}^n .

Protože pomocí podmíněných hp popisujeme zpravidla kauzální závislosti, vyskytuje se často v podmínce všechna historická data, tj. data změřená od začátku odhadování (řízení) až do současného nebo minulého okamžiku. Do těchto dat zahrnujeme i apriorní informaci, např. ve formě apriorních dat. Pro tuto datovou historii, zahrnující data do času τ zavedeme značení:

$d_\tau = [y_\tau, u_\tau]^T$ jsou data (datová dvojice vstupu a výstupu) v čase τ ,

$D(\tau, \dots, t) = \{d_\tau, d_{\tau+1}, \dots, d_t\}$ je množina datových dvojic počínaje indexem τ až do t .

Potom

$$D(t) = \{d_t, d_{t-1}, d_{t-2}, \dots\} \quad (5.1)$$

značí souhrn všech datových dvojic až do (současného) okamžiku s časovým indexem t , $D(t-1)$ zahrnuje všechna data minulá. Symbol $D(0)$ označuje apriorní data (všechna data, naměřená před začátkem řízení). Data $D(t)$, získaná až do okamžiku t , lze rozdělit takto: $D(t) = D(1, \dots, t) \cup D(0)$, tj. na data apriorní a data získaná v průběhu řízení až do času t .

Přirozené podmínky řízení

Odhady parametrů Θ modelu soustavy řízené ve zpětné vazbě prostřednictvím řídicí veličiny u_t předpokládáme platnost tzv. **přirozených podmínek řízení** (p.p.ř.). Ty jsou vyjádřeny následujícími vztahy

$$f(u_t|D(t-1), \Theta) = f(u_t|D(t-1)), \quad (5.2)$$

$$f(\Theta|u_t, D(t-1)) = f(\Theta|D(t-1)), \quad (5.3)$$

kteřé jsou navzájem svázány Bayesovým vzorcem. Tyto podmínky vycházejí ze skutečnosti, že řídicí veličina smí být konstruována jen na základě dat $D(t-1)$, a tedy neznámé parametry neovlivňují pravidlo (5.2) jejího generování.

Objekty a vzorce v předpovědi a odhadování

Úlohu předpovědi a odhadování budeme dále demonstrovat na příkladech pro jednotlivé druhy modelů, diskutované v předchozí kapitole. Jsou to spojitě modely s normálním 4.1.1 a 4.1.2, rovnoměrným 4.1.3 a diskrétním šumem 4.1.4 a diskrétní modely 4.2. V těchto příkladech budeme sledovat vždy stejné schéma s následujícími prvky:

Model

$$y_t = \varphi_t^T \theta + e_t \quad \text{nebo jeho hp} \quad f(y_t|\varphi_t, \Theta) \quad (5.4)$$

podle (4.2), kde φ_t je regresní vektor, θ je vektor regresních koeficientů, e_t je šum s nulovou střední hodnotou a rozptylem σ^2 a $\Theta = \{\theta, \sigma^2\}$ je vektor (struktura) označující souhrn všech parametrů modelu.

Věrohodnostní funkce

$$L_t(\Theta) = \prod_{\tau=1}^t f(y_\tau|\varphi_\tau, \Theta) \quad (5.5)$$

je součin hp modelu s **dosazenými hodnotami naměřených** dat y_τ a regresních vektorů φ_τ , pro $\tau = 1, 2, \dots, t$.

Apriorní hp

$$f(\Theta|D(0)) \propto \prod_{\tau=-\kappa_0}^0 f(y_\tau|\varphi_\tau, \Theta) \quad (5.6)$$

je hp parametrů, konstruovaná jen na základě apriorní znalosti která je zde vyjádřena symbolem $D(0)$ - jako data získaná do začátku vlastního odhadování.

Aposteriorní hp

$$f(\Theta|D(t)) \propto f(y_t|\varphi_t, \Theta)f(\Theta|D(t-1)) \propto L_t(\Theta)f(\Theta|D(0)) \quad (5.7)$$

je hp parametrů, ve které je zahrnuta jednak apriorní znalost, jednak informace obsažená v datech změřených během odhadování, tj. pro časové okamžiky $\tau = 1, 2, \dots, t$. Tato data, včetně apriorních, označujeme symbolem $D(t)$.

Hp $f(\Theta|D(t-1))$ někdy označujeme jako apriorní z minulého kroku.

Hp předpovědi

$$f(y_t|u_t, D(t-1)) \underset{\text{p.p.ř.}}{=} \int f(y_t|u_t, D(t-1), \Theta) f(\Theta|D(t-1)) d\Theta^1 \quad (5.8)$$

je to hp výstupu, podmíněná vstupem a staršími daty. Je podobná hp modelu, ale nemá v podmínce parametry a závisí na všech minulých datech, nikoli jen na regresním vektoru.

Bodový odhad

$$\hat{\Theta}_t = E[\Theta|D(t-1)]. \quad (5.9)$$

Bodová předpověď

$$\hat{y}_t = E[y_t|u_t, D(t-1)]. \quad (5.10)$$

5.2 Bodová předpověď a bodové odhady

Pro motivaci rekurzivního odhadování parametrů vyjdeme z praktické úlohy – bodové předpovědi výstupu, což je přímá předpověď budoucích hodnot výstupu soustavy. Nejdříve budeme specifikovat, co máme na mysli **optimálním bodovým odhadem**, ať už je to odhad výstupu, kterému říkáme předpověď¹, nebo bodový odhad neznámého parametru.

Řekneme, že odhad je optimální, jestliže minimalizuje zvolené kritérium optimality. Z hlediska výpočetního je nejzajímavější kritérium kvadratické, které pro předpověď výstupu \hat{y}_t , resp. odhad parametrů $\hat{\Theta}_t$ má tvar

$$J_{\text{pre}} = E[(\hat{y}_t - y_t)^2], \quad \text{resp.} \quad J_{\text{odh}} = E[(\hat{\Theta}_t - \Theta_t)^2].$$

Tento typ kritéria vede na bodové odhady ve formě střední hodnoty, podmíněné minulými daty. Předpověď výstupu, resp. bodový odhad parametrů je

$$\hat{y}_t = E[y_t|u_t, D(t-1)], \quad \text{resp.} \quad \hat{\Theta}_t = E[\Theta|D(t-1)], \quad (5.11)$$

kde ve vyjádření bodového odhadu parametrů jsou využity přirozené podmínky řízení, zmíněné na začátku této kapitoly. Odvození těchto výsledků je ponecháno čtenáři jako cvičení. V případě nouze je uvedeno v kapitole s dodatky, část 11.5.

♣ Optimální bodové odhady minimalizující kvadratické kritérium lze vyjádřit jako podmíněné střední hodnoty. V podmínce jsou jen starší, již naměřená data.

Bodová předpověď výstupu soustavy

Bodová předpověď podle kvadratického kritéria je dána podmíněnou střední hodnotou podle (5.11)

$$\hat{y}_t = E[y_t|u_t, D(t-1)] = \int_{-\infty}^{\infty} y_t f(y_t|u_t, D(t-1)) dy_t.$$

¹Označení p.p.ř. připomíná platnost přirozených podmínek řízení (5.2) nebo (5.3)

Pro její výpočet potřebujeme **hp předpovědi** $f(y_t|u_t, D(t-1))$. V případě známých parametrů je tato hp rovna přímo hp modelu a situace je velmi jednoduchá. Stačí dosadit měřený regresní vektor do modelu a "šum položit roven nule". Složitější je předpověď v případě, kdy parametry modelu jsou neznámé. Obě možnosti ukážeme na příkladě.

Příklad [Bodová předpověď pro model se známými parametry]

Uvažujme model podle (5.4) ve tvaru

$$y_t = \varphi_t^T \theta + e_t \quad \text{s hp} \quad f(y_t|\varphi_t) \propto N(\varphi_t^T \theta, \sigma^2),$$

kde $\Theta = \{\theta, \sigma^2\}$ známe.

Abychom byli ještě konkrétnější, zvolme následující číselné zadání modelu

$$y_t = 0.6y_{t-1} + 0.1u_t + 0.5 + 0.3\epsilon_t$$

kde $\epsilon_t \sim N(0, 1)$, $e_t = 0.3\epsilon_t$ a $D[e_t] = D[0.3\epsilon_t] = 0.09$.

Hp, odpovídající tomuto modelu je

$$f(y_t|u_t, y_{t-1}) = f(y_t|\varphi_t),$$

kde $\varphi_t = [u_t, y_{t-1}, 1]$.

(konec příkladu)

Poznámka: Na tomto číselném zadání je dobře patrné, že hp modelu v tomto případě neobsahuje v podmínce parametry. Žádná náhodná veličina popisující parametry totiž neexistuje. Parametry jsou známá čísla.

Bodová předpověď je (5.11)

$$\hat{y}_t = E[y_t|u_t, D(t-1)] = \int y_t f(y_t|\varphi_t) dy_t = \varphi_t^T \theta \quad (5.12)$$

a protože jak θ , tak i φ_t známe, lze dosadit a předpověď vypočítat. To lze provést pro číselné zadání

$$\hat{y}_t = [2; 2.1; 1] \cdot [0.1; 0.6; 0.5]^T = 1.96$$

♣ Bodová předpověď výstupu při známých parametrech regresního modelu je dána skalárním součinem regresního vektoru s vektorem parametrů.

Příklad [Bodová předpověď pro model s neznámými parametry]

Zde je situace složitější než v případě známých parametrů. Předchozí výsledek nelze použít, protože parametry modelu neznáme. K výpočtu střední hodnoty podle (5.11) potřebujeme hp předpovědi, podmíněnou daty a nikoliv parametry. Tu vypočteme takto

$$\begin{aligned} f(y_t|u_t, D(t-1)) &= \int f(y_t, \Theta|u_t, D(t-1)) d\Theta = && \text{[marginální pravděp.]} \\ &= \int f(y_t|u_t, D(t-1), \Theta) f(\Theta|u_t, D(t-1)) d\Theta. && \text{[podmíněná pravděp.]} \end{aligned}$$

V posledním řádku můžeme ještě provést dvě úpravy:

1. $f(y_t|u_t, D(t-1), \Theta) = f(y_t|\varphi_t, \Theta)$, protože při daných parametrech závisí y_t jen na φ_t a ne na starších datech – viz předpoklad o regresním modelu (4.1),

2. $f(\Theta|u_t, D(t-1)) = f(\Theta|D(t-1))$, plyne z přirozených podmínek řízení (5.2), (5.3).

Je tedy

$$f(y_t|u_t, D(t-1)) = \int f(y_t|\varphi_t, \Theta)f(\Theta|D(t-1)) d\Theta. \quad (5.13)$$

Poznámka: Interpretace tohoto výsledku je názorná a odpovídá metodě úplné pravděpodobnosti (3.3). Mohu použít jen model s dosazenými parametry, přičemž hodnoty parametrů neznám, znám jen jejich rozdělení, tj. "všechny možné hodnoty a jejich pravděpodobnosti". Výsledná hp je dána váženým součtem (integrálem) "modelů" s dosazenými hodnotami parametrů v podmínce $f(y_t|\varphi_t, \Theta)$, váženými "pravděpodobnostmi hodnot parametrů" $f(\Theta|D(t-1))$.

Nyní můžeme přistoupit k výpočtu střední hodnoty $E[y_t|u_t, D(t-1)]$, jak je uvedeno v (5.11). Použijeme předchozí vyjádření hp předpovědi a dostaneme

$$\begin{aligned} E[y_t|u_t, D(t-1)] &= \int_{y_t} y_t f(y_t|u_t, D(t-1)) dy_t = \\ &= \int_{y_t} y_t \int_{\Theta} f(y_t|\varphi_t, \Theta) f(\Theta|D(t-1)) d\Theta dy_t = && \text{[dosazeno z (5.13)]} \\ &= \int_{\Theta} \left[\int_{y_t} y_t f(y_t|\varphi_t, \Theta) dy_t \right] f(\Theta|D(t-1)) d\Theta = && \text{[záměna integrace]} \\ &= \int_{\Theta} \varphi_t^T \theta f(\Theta|D(t-1)) d\Theta = && \text{[dosazeno z (5.12)]} \\ &= \varphi_t^T \int_{\Theta} \theta f(\theta, \sigma|D(t-1)) d\theta d\sigma = && \text{[integrováno přes } \sigma \text{]} \\ &= \varphi_t^T \int_{\theta} \theta f(\theta|D(t-1)) d\theta = \varphi_t^T \hat{\theta}_{t-1}, \end{aligned}$$

kde $\hat{\theta}_{t-1}$ značí bodový odhad θ v časovém okamžiku $t-1$. Protože bodová předpověď se rovná této střední hodnotě, dostáváme

$$\hat{y}_t = \varphi_t^T \hat{\theta}_{t-1}. \quad (5.14)$$

(konec příkladu)

♣ Bodová předpověď výstupu v čase t při neznámých parametrech regresního modelu je dána skalárním součinem regresního vektoru v čase t s vektorem bodových odhadů parametrů z času $t-1$.

5.3 Předpovídání a odhadování

Podobně, jako jsme u modelu soustavy uvažovali jeho dvojí možné vyjádření, a to buď pomocí hp (4.3), nebo rovnic (4.1), můžeme takové dvě pomyslné linie sledovat i při vytváření algoritmů předpovědi a odhadování. První linie počítá s obecnými hustotami pravděpodobnosti, druhá tyto hustoty konkretizuje a počítá se speciálním rozdělením a jeho statistikami, tj. měřená data akumuluje do statistik a z nich případně jako výsledky počítá odhady. První cesta je obecná, druhá dává praktické výsledky.

Vlastní odhad často není cílem, ale jen prostředkem k dosažení jiného cíle. V předchozím odstavci jsme jako cíl postavili předpověď výstupu pro model s neznámými parametry. Jako podúloha, která vznikla při řešení tohoto problému, se objevil právě odhad. Protože jsme ale požadovali bodovou předpověď, potřebovali jsme jako výsledek odhadu také bodové odhady

parametrů (5.14). Vezmeme-li ale za cílovou úlohu např. optimální řízení s pravděpodobnostním modelem, minimalizující střední hodnotu kritéria penalizujícího kvadráty vstupu a výstupu (tak, jak jej budeme uvažovat na konci této kapitoly), dostaneme také jako podúlohu odhad parametrů. V tomto případě se ale úloha odhadování neomezí jen na pouhý výpočet bodových odhadů a jejich dosazení za neznámé parametry. Pracuje se zde s celými hp parametrů pomocí pravidla o úplné pravděpodobnosti – viz poznámka ke vztahu (5.13). Volně řečeno, místo výběru jedné hodnoty odhadovaného parametru se pracuje se všemi možnými hodnotami a jejich pravděpodobnostmi. Tím se zachová informace nejen o hodnotě parametru, ale také o jeho neurčitosti. Právě neurčitost odhadnutého parametru se při dosazení bodového odhadu ztrácí.

- ♣ Možnost vyjádření veškeré informace o odhadovaném parametru včetně jeho neurčitosti je velkou předností bayesovského přístupu.

Konstrukce hp předpovědi

Jak jsme ukázali v předchozí části (viz (5.13)), lze hp předpovědi (5.8) konstruovat podle vzorce pro úplnou pravděpodobnost (3.3) v integrální podobě

$$f(y_t|u_t, D(t-1)) = \int f(y_t|\varphi_t, \Theta) f(\Theta|D(t-1)) d\Theta. \quad (5.15)$$

První hp na pravé straně je hp definovaná modelem s danými parametry a druhá je hp parametrů, podmíněná znalostí dat až do okamžiku $t-1$. Tuto hp parametrů neznáme a musíme ji postupně napočítávat s měřenými daty pomocí odhadu, jak ukážeme ihned v následujícím textu.

- ♣ Konstrukce hp předpovědi probíhá podle vzorce pro úplnou pravděpodobnost. Vstupuje do ní model soustavy a hp parametrů kterou neznáme a kterou musíme postupně napočítávat z dat a apriorní hp v odhadu.

Poznámka: Hp předpovědi je možno počítat průběžně, jak jsme to zatím ukázali, ale také jednorázově pro všechna dosud naměřená data najednou. Odvození této jednorázové předpovědi provedeme v dodatku, v odstavci 11.4. Zde pro úplnost výkladu uvedeme jen výsledek

$$f(y_t|u_t, D(t-1)) = \frac{I_t}{I_{t-1}}, \quad (5.16)$$

kde

$$I_\vartheta = \int L_\vartheta(\Theta) f(\Theta|D(0)) d\Theta \quad \text{pro } \vartheta = t, t-1,$$

$L_t(\vartheta)$ je věrohodnostní funkce (5.5) a $f(\Theta|D(0))$ je apriorní hp (5.6) z počátku odhadu.

Konstrukce hp parametrů

Ve vztahu (5.15) pro hp předpovědi jsme použili hp parametrů $f(\Theta|D(t-1))$, kterou přímo neznáme. Její konstrukci je možno provést rekurzivně tak, jak postupně měříme data. Vychází z tzv. **apriorní hp** (5.6), která vyjadřuje naše předběžné znalosti parametrů, nebo se opírá o předem změřená data. K této hp je postupně s využitím Bayesova vzorce (3.5) přidávána informace získaná z měřených dat, a tak je apriorní hp přepočítávána na tzv. **aposteriorní hp** (5.7). Ta je buď žádaným výsledkem, jako například pro výpočet předpovědi (5.8), nebo jsou z ní počítány bodové odhady (5.9), tj. hodnoty, které mohou být za parametry za určitých okolností dosazeny.

Pro odvození rekurze vyjdeme z časového okamžiku t a aposteriorní hp $f(\Theta|D(t))$ vyjádříme pomocí aposteriorní hp z minulého kroku $f(\Theta|D(t-1))$

$$\begin{aligned} f(\Theta|D(t)) &= f(\Theta|y_t, u_t, D(t-1)) = \\ &= \frac{f(y_t|u_t, D(t-1), \Theta)f(\Theta|u_t, D(t-1))}{f(y_t|u_t, D(t-1))}. \quad [\text{Bayesův vzorec}] \end{aligned} \quad (5.17)$$

Pomocí stejné argumentace jako pro předpověď (5.15) lze z podmínky první hp v čitateli odstranit data starší než jsou v regresním vektoru a z podmínky druhé hp v čitateli odstranit veličinu u_t , jak plyne z přirozených podmínek řízení (5.3). Hp ve jmenovateli nezávisí na Θ , představuje tedy jen konstantu, která zajišťuje jednotkový integrál aposteriorní hp, tj. $\int f(\Theta|D(t)) d\Theta = 1$. Z této podmínky ji lze kdykoli dopočítat, nemusíme ji tedy uvažovat.

Jmenovatel uvedeného vztahu nezávisí na Θ . Pomocí znaménka pro úměrnost \propto lze tento vztah zapsat takto

$$f(\Theta|D(t)) \propto f(y_t|\varphi_t, \Theta)f(\Theta|D(t-1)). \quad (5.18)$$

Jedná se o rekurentní výpočet hp parametrů, který začíná s apriorní hustotou $f(\Theta|D(0)) = f(\Theta)$ a v jednotlivých časových okamžicích $\tau = 1, 2, \dots, t$ poskytuje požadovanou (aposteriorní) hp.

Uvedený postup provádí výpočty postupně (on-line), tj. jak měříme data, tak přepočítáváme aposteriorní hp. Celý výpočet je ale možno provést i jednorázově (off-line) až po naměření všech dat. Pak, jak plyne z (5.18),

$$f(\Theta|D(t)) \propto L_t(\Theta)f(\Theta), \quad (5.19)$$

kde $L_t(\Theta)$ je věrohodnostní funkce (5.5).

♣ Konstrukce hp parametrů probíhá rekurentně podle Bayesova vzorce. Vychází z apriorní hp a s použitím naměřených dat ji přepočítává na aposteriorní hp.

ALGORITMUS

PŘEDPOVĚD' VÝSTUPU A ODHAD PARAMETRŮ

Start apriorní hp

$$f(\Theta|D(0)) = f(\Theta) \quad [\text{viz (5.6)}]$$

for $\tau = 1 : t$

předpověď

$$f(y_\tau|u_\tau, D(\tau-1)) = \int f(y_\tau|\varphi_\tau, \Theta)f(\Theta|D(\tau-1)) d\tau \quad [\text{viz (5.15)}]$$

odhad

$$f(\Theta|D(\tau)) = \frac{f(y_\tau|\varphi_\tau, \Theta)f(\Theta|D(\tau-1))}{f(y_\tau|u_\tau, D(\tau-1))} \quad [\text{viz (5.17)}]$$

end % for

Poznámka: Uvedený algoritmus představuje rekurzi pro funkce. V tomto tvaru není prakticky realizovatelný.

5.4 Příklady

V následujících příkladech ukážeme, jak probíhá předpověď a odhad na konkrétních modelech. Budeme sledovat typy modelů, o nichž jsme hovořili v příkladové části 4.1 a 4.2 předchozí kapitoly. Uvidíme, že konečný algoritmus odhadování je velmi podobný algoritmům pro deterministický model, uvedeným v kapitole 2.

5.4.1 Regresní model obecného řádu

V tomto příkladě navážeme na příklad modelu zadaného rovnicí (4.7). Budeme přitom demonstrovat postup odhadování pro nejběžnější případ skalárního regresního modelu obecného řádu ν , zadaného rovnicí

$$y_t = \varphi_t^T \theta + e_t, \quad (5.20)$$

s hp modelu (4.11), tj.

$$f(y_t | \varphi_t, \Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2\sigma^2} (y_t - \varphi_t^T \theta)^2 \right\}. \quad (5.21)$$

Součin modelů

Nejprve vyjádříme součin modelů od času 1 do t jako funkci dat i parametrů

$$\prod_{\tau=1}^t f(y_\tau | \varphi_\tau, \Theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^t \exp \left\{ -\frac{1}{2\sigma^2} \sum_{\tau=1}^t (y_\tau - \varphi_\tau^T \theta)^2 \right\}. \quad (5.22)$$

Věrohodnostní funkce

Součin modelů, vyjádřený jako funkce parametrů, je věrohodnostní funkce (5.5). Zavedeme-li sloupcové vektory $[-1, \theta^T]^T$ a $[y_t, \varphi_t^T]^T$, můžeme předchozí součin modelů uspořádat jako funkci parametrů θ, σ^2 . Jde o stejný postup a značení, jako jsme použili při studiu modelu (2.17) (viz (2.19), (2.20) a (2.21))

$$L_t(\Theta) = (2\pi\sigma^2)^{-\frac{t}{2}} \exp \left\{ -\frac{1}{2\sigma^2} [-1 \ \theta^T] \sum_{\tau=1}^t \left(\begin{bmatrix} y_\tau \\ \varphi_\tau \end{bmatrix} [y_\tau \ \varphi_\tau^T] \right) \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}. \quad (5.23)$$

Apriorní hp

Na začátku odhadování volíme apriorní hp (5.6) tak, aby měla podobný tvar jako věrohodnostní funkce (z výpočetních důvodů potřebujeme, aby se její tvar v Bayesově vzorci reprodukoval)

$$f(\Theta | D(0)) \propto \sigma^{-\frac{\kappa(0)}{2}} \exp \left\{ -\frac{1}{2\sigma^2} [-1 \ \theta^T] V(0) \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}, \quad (5.24)$$

kde $V(0)$ je pozitivně definitní symetrická čtvercová matice odpovídajícího stupně a $\kappa(0)$ je skalár.

Aposteriorní hp

Je to, podle (5.7), normovaný součin věrohodnostní funkce a apriorní hp (5.6)

$$f(\Theta | D(t)) \propto L_t(\Theta) f(\Theta | D(0)).$$

Často nám stačí určit jen funkci úměrnou aposteriorní hp, protože normování nemění potřebné charakteristiky hp (např. extrém). Pro náš model dostaneme

$$f(\Theta|D(t)) \propto \sigma^{-\frac{\kappa(t)}{2}} \exp \left\{ -\frac{1}{2\sigma^2} \begin{bmatrix} -1 & \theta^T \end{bmatrix} V(t) \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}, \quad (5.25)$$

kde – viz (2.21), (2.22)

$$V(t) = V(t-1) + \begin{bmatrix} y_t \\ \varphi_t \end{bmatrix} \begin{bmatrix} y_t & \varphi_t^T \end{bmatrix} = V(0) + \sum_{\tau=1}^t \left(\begin{bmatrix} y_\tau \\ \varphi_\tau \end{bmatrix} \begin{bmatrix} y_\tau & \varphi_\tau^T \end{bmatrix} \right), \quad (5.26)$$

$$\kappa(t) = \kappa(t-1) + 1 = \kappa(0) + t. \quad (5.27)$$

Označíme, viz (2.26)

$$V(t) = \begin{bmatrix} V_y(t) & V_{y\varphi}^T(t) \\ V_{y\varphi}(t) & V_\varphi(t) \end{bmatrix}, \quad (5.28)$$

kde V_y je skalár, $V_{y\varphi}$ je sloupcový vektor a V_φ je čtvercová matice. Rozměry submatic odpovídají rozměrům y a φ , ze kterých matice V vznikla. S takto zavedeným označením lze exponent aposteriorní hp doplnit na čtverec v proměnné θ

$$-\frac{1}{2\sigma^2} \begin{bmatrix} -1 & \theta^T \end{bmatrix} \begin{bmatrix} V_y(t) & V_{y\varphi}^T(t) \\ V_{y\varphi}(t) & V_\varphi(t) \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix} = -\frac{1}{2\sigma^2} [(\theta - \hat{\theta}_t)^T V_\varphi(t) (\theta - \hat{\theta}_t) + \Lambda_t],$$

kde (srv. (2.27) a (2.28))

$$\Lambda_t = V_y(t) - V_{y\varphi}^T(t) V_\varphi^{-1}(t) V_{y\varphi}(t), \quad (5.29)$$

$$\hat{\theta}_t = V_\varphi^{-1}(t) V_{y\varphi}(t). \quad (5.30)$$

Bodové odhady parametrů

Z doplněného tvaru exponentu plyne, že podmíněná střední hodnota θ v čase t je $\hat{\theta}_t$, což je bodový odhad regresních koeficientů. Pro určení bodového odhadu σ^2 můžeme použít např. derivaci k určení extrému (prakticky se stejným výsledkem, jako kdybychom počítali střední hodnotu) a odhad rozptylu šumu σ^2 v čase t závisí na funkci Λ_t . Platí

$$\hat{\theta}_t = V_\varphi^{-1}(t) V_{y\varphi}(t), \quad [\text{odhad regresních koeficientů}] \quad (5.31)$$

$$\hat{\sigma}_t^2 = \Lambda_t / \kappa(t). \quad [\text{odhad rozptylu šumu}] \quad (5.32)$$

Hp předpovědi

Je dána vztahem (5.15)

$$f(y_t|u_t, D(t-1)) = \int f(y_t|\varphi_t, \Theta) f(\Theta|D(t-1)) d\Theta.$$

Výsledek integrace, a tedy hp předpovědi je

$$f(y_t|u_t, D(t-1)) = \frac{\Gamma\left(\frac{\kappa_t - \rho + 2}{2}\right)}{2^{\frac{\kappa_t + 2}{2}} \pi^\rho} |V_{\varphi,t}|^{\frac{1}{2}} \Lambda_{t-1}^{\frac{\kappa_t - \rho + 3}{2}}, \quad (5.33)$$

kde ρ je dimenze regresního vektoru φ_t .

Poznámka: Výpočet uvedeného integrálu je analogický tomu, které je uvedeno na str. 80 při integraci aposteriorní hp.

Bodová předpověď výstupu

Je dána jako střední hodnota y_t , podmíněná staršími daty (nikoliv parametry). Podle (5.14) je

$$\hat{y}_t = \varphi_t^T \hat{\theta}_{t-1}, \quad (5.34)$$

kde $\hat{\theta}_{t-1}$ je odhad regresních koeficientů z minulého kroku.

ALGORITMUS

BAYESOVSKÉ ODHADOVÁNÍ

(REGRESNÍ MODEL, NORMÁLNÍ ŠUM)

Před sběrem dat

Příprav nebo změř apriorní data

$$y_0, u_0, y_{-1}, u_{-1}, \dots, y_{-n}, u_{-n}$$

Sestav apriorní statistiky (např. pomocí apriorních dat)

$$\kappa, \quad V$$

Sběr dat

for $\tau = 1 : t$

Generuj vstup a změř výstup

$$u_\tau = \dots, \quad y_\tau = \dots$$

Sestav rozšířený regresní vektor, (4.1)

$$\phi^T = [y_\tau, u_\tau, \dots, y_{\tau-n}, u_{\tau-n}]$$

Proved' aktualizaci statistik, (5.26), (5.27)

$$\kappa = \kappa + 1, \quad V = V + \phi\phi^T$$

end % for

Po sběru dat

Rozděl informační matici, (5.28)

$$V_y = V_{11}, \quad V_{y\varphi} = V_{2:\dim V, 1}, \quad V_\varphi = V_{2:\dim V, 2:\dim V}$$

Vypočti

– zbytek kritéria (5.29)

$$\Lambda = V_y - V_{y\varphi}^T V_\varphi^{-1} V_{y\varphi}$$

– odhad parametrů (5.30)

$$\hat{\theta} = V_\varphi^{-1} V_{y\varphi}, \quad \hat{\sigma}^2 = \frac{\Lambda}{\kappa}$$

Cvičení: Naprogramovat a zkusit v MATLABu pro různé vstupní signály a úrovně šumu.

5.4.2 Statický regresní model s normálním šumem

A. Model se známým rozptylem

Statický model (4.14) je velmi jednoduchý. I když samozřejmě i v tomto případě platí právě uvedený obecný algoritmus odhadu, není zde třeba zavádět vektorovou symboliku jako v případě regresního modelu obecného řádu (4.7) a odvození algoritmu odhadování je průhlednější. Jak uvidíme, budeme pracovat v podstatě s prvním a druhým momentem datového výběru, tj. s průměrem a rozptylem výběru.

Při odvození budeme postupovat tak, že nejprve pomocí hp modelu a aposteriorní hp z minulého kroku určíme sdruženou hp pro výstup y_t a parametry Θ a z ní pak doplněním exponentu normálního rozdělení na čtverec určíme vše ostatní – aposteriorní hp i hp předpovědi. Budeme tedy sledovat následující schéma pro hp v jednom časovém okamžiku t

$$\underbrace{f(y_t|\theta)}_{\text{model}} \underbrace{f(\theta|D(t-1))}_{\text{apriorní}} = \underbrace{f(y_t, \theta|D(t-1))}_{\text{sdružená}} = \underbrace{f(\theta|D(t))}_{\text{aposteriorní}} \underbrace{f(y_t|D(t-1))}_{\text{předpověď}}, \quad (5.35)$$

kteří není nic jiného než postupně rozepsaný Bayesův vzorec (5.17).

Jednotlivé odvozené vzorce budou představovat speciální případy vzorců z předchozího příkladu pro obecný regresní model.

Model

Je dán rovnicí (4.14), tj.

$$y_t = \theta + e_t \quad (5.36)$$

s hp podle (4.15). Je to model s jediným neznámým parametrem θ , protože uvažujeme známý rozptyl $\sigma^2 = 1$. Je tedy $\Theta = \{\theta\}$.

Věrohodnostní funkce

V obecném čase ϑ je věrohodnostní funkce (5.5)

$$\begin{aligned} L_{\vartheta}(\theta) &= \prod_{\tau=1}^{\vartheta} f(y_{\tau}|\theta) = (2\pi)^{-\frac{\vartheta}{2}} \exp \left\{ -\frac{1}{2} \sum_{\tau=1}^{\vartheta} (y_{\tau}^2 - 2\theta y_{\tau} + \theta^2) \right\} = \\ &= (2\pi)^{-\frac{\vartheta}{2}} \exp \left\{ -\frac{1}{2} \left(\vartheta\theta^2 - 2\theta \sum_{\tau=1}^{\vartheta} y_{\tau} + \sum_{\tau=1}^{\vartheta} y_{\tau}^2 \right) \right\}. \end{aligned} \quad (5.37)$$

Zřejmě exponent součinu modelů lze upravit jako kvadratickou formu v proměnné θ . Celý součin je určen statistikami

$$\vartheta, \quad \sum_{\tau=1}^{\vartheta} y_{\tau}, \quad \sum_{\tau=1}^{\vartheta} y_{\tau}^2, \quad (5.38)$$

udávajícími po řadě počet dat, součet dat a součet kvadrátů dat (srv. struktura matic na str. 26).

Apriorní hp

Volíme ji podle předchozího vyjádření součinu modelů a v souladu s regresním modelem obecného řádu ve tvaru (5.6)

$$f(\theta|D(0)) \propto \exp \left\{ -\frac{1}{2} \left(\kappa(0)\theta^2 - 2\theta S_{y1}(0) + S_y^2(0) \right) \right\}, \quad (5.39)$$

kde $\kappa(0)$, $S_{y1}(0)$, $S_y^2(0)$ jsou statistiky této hp (jednička v indexu druhé statistiky formálně označuje $\varphi = 1$).

Poznámka: Je-li apriorní hp získána z apriorních dat měřených před začátkem odhadování pro $\tau = (-\kappa(0) + 1) : 0$, mají uvedené statistiky tento význam:

$$S_{y1}(0) = \sum_{\tau=-\kappa(0)+1}^0 y_\tau \quad \text{a} \quad S_y^2(0) = \sum_{\tau=-\kappa(0)+1}^0 y_\tau^2.$$

Pro apriorní hp s rovnoměrným rozdělením mají všechny statistiky hodnotu nula.

Jiná volba apriorní hp, pro kterou ovšem musí platit $\kappa(0) > 0$ a $S_y^2(0) > 0$, vyjadřuje naši subjektivní apriorní informaci.

Dále ukážeme jeden krok přepočtu aposteriorní hp z času $t-1$ na čas t a zabudování informace z dat y_t změřených v čase t . Pro přehlednost budeme uvažovat apriorní statistiky nulové.

Apsteriorní hp pro $t-1$

V čase $t-1$ má tato hp stejný tvar jako apriorní (5.39) se statistikami v příslušném časovém okamžiku, zde $t-1$, srv. (5.7)

$$f(\theta|D(t-1)) \propto \exp\left\{-\frac{1}{2}\left(\kappa(t-1)\theta^2 - 2\theta S_{y1}(t-1) + S_y^2(t-1)\right)\right\} = \quad (5.40)$$

$$= \exp\left\{-\frac{1}{2}\left[\kappa(t-1)\left(\theta - \hat{\theta}_{t-1}\right)^2 + \Lambda_{t-1}\right]\right\}, \quad (5.41)$$

kde $\hat{\theta}_{t-1} = \frac{S_{y1}(t-1)}{\kappa(t-1)}$, a pro $\kappa(0) = 0$ je

$$\Lambda_{t-1} = S_y^2(t-1) - \frac{(S_{y1}(t-1))^2}{\kappa(t-1)} = \sum_{\tau=1}^{t-1} \left(y_\tau - \frac{S_{y1}(t-1)}{\kappa(t-1)}\right)^2, \quad \kappa(t-1) = t-1,$$

$$S_{y1}(t-1) = \sum_{\tau=1}^{t-1} y_\tau, \quad S_y^2(t-1) = \sum_{\tau=1}^{t-1} y_\tau^2, \quad S_{y1}(0) = 0, \quad S_y^2(0) = 0$$

Poznámka: Všimněme si, že statistika Λ po vydělení počtem dat κ představuje rozptyl šumu. Tento výsledek jsme odvodili použitím vzorce pro výpočetní tvar rozptylu $D[x] = E[x^2] - E[x]^2$.

Poznámka: Při použití nenulových apriorních statistik bychom je museli k uvedeným statistikám ještě přičíst jako v (5.26) a (5.27).

Sdružená hp

Pro výstup y_t a parametr θ ji konstruujeme podle schématu hp (5.35) tak, že vynásobíme hp modelu (4.15) a apsteriorní hp (5.41) pro čas $t-1$

$f(y_t, \theta|D(t-1)) \propto$

$$\propto (2\pi)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y_t - \theta)^2\right\} (2\pi)^{-\frac{\kappa(t-1)}{2}} \exp\left\{-\frac{1}{2}\left(\kappa(t-1)(\theta - \hat{\theta}_{t-1})^2 + \Lambda_{t-1}\right)\right\} = \quad (5.42)$$

$$= (2\pi)^{-\frac{\kappa(t-1)+1}{2}} \exp\left\{-\frac{1}{2}\left((y_t - \theta)^2 + \kappa(t-1)(\theta - \hat{\theta}_{t-1})^2 + \Lambda_{t-1}\right)\right\}. \quad (5.43)$$

Upravíme exponent tak, že jej doplníme na čtverec v proměnné θ – viz (11.1)

$$\begin{aligned}
& y_t^2 - 2\theta y_t + \theta^2 + \kappa(t-1) (\theta^2 - 2\theta\hat{\theta}_{t-1} + \hat{\theta}_{t-1}^2) + \Lambda_{t-1} = \\
& = (\kappa(t-1) + 1) \theta^2 - 2\theta (y_t + \kappa(t-1)\hat{\theta}_{t-1}) + y_t^2 + \kappa(t-1)\hat{\theta}_{t-1}^2 + \Lambda_{t-1} = \\
& = (\kappa(t-1) + 1) \left(\theta - \frac{y_t + \kappa(t-1)\hat{\theta}_{t-1}}{\kappa(t-1) + 1} \right)^2 - \frac{(y_t + \kappa(t-1)\hat{\theta}_{t-1})^2}{\kappa(t-1) + 1} + \\
& \quad + y_t^2 + \kappa(t-1)\hat{\theta}_{t-1}^2 + \Lambda_{t-1} = \\
& = \kappa(t) (\theta - \hat{\theta}_t)^2 + \frac{\kappa(t-1)}{\kappa(t)} (y_t - \hat{\theta}_{t-1})^2 + \Lambda_{t-1},
\end{aligned}$$

kde $\kappa(t) = \kappa(t-1) + 1$ je počet dat, $\hat{\theta}_t = \frac{y_t + \kappa(t-1)\hat{\theta}_{t-1}}{\kappa(t-1) + 1} = \hat{\theta}_{t-1} + \frac{1}{\kappa(t)}(y_t - \hat{\theta}_{t-1})$ je odhad parametru, pro který podle (5.30) a (5.29) při $\kappa(0) = 0$ platí

$$\hat{\theta}_t = \frac{1}{\kappa(t)} \sum_{\tau=1}^t y_\tau \quad \text{a} \quad \Lambda_t = \Lambda_{t-1} + \frac{\kappa(t-1)}{\kappa(t)} (y_t - \hat{\theta}_{t-1})^2.$$

Po této úpravě exponentu a dosazení $\kappa(t-1) + 1 = \kappa(t)$ bude mít sdružená hp tvar

$$\begin{aligned}
f(y_t, \theta | D(t-1)) & \propto (2\pi)^{-\frac{\kappa(t)}{2}} \exp \left\{ -\frac{1}{2} \left[\kappa(t) (\theta - \hat{\theta}_t)^2 + \frac{\kappa(t-1)}{\kappa(t)} (y_t - \hat{\theta}_{t-1})^2 + \Lambda_{t-1} \right] \right\} = \\
& = (2\pi)^{-\frac{\kappa(t)}{2}} \exp \left\{ -\frac{1}{2} \kappa(t) (\theta - \hat{\theta}_t)^2 \right\} \exp \left\{ -\frac{1}{2} \frac{\kappa(t-1)}{\kappa(t)} (y_t - \hat{\theta}_{t-1})^2 \right\} \exp \left\{ -\frac{1}{2} \Lambda_{t-1} \right\} \quad (5.44)
\end{aligned}$$

Tímto jsme se dostali k hlavnímu bodu celého příkladu, ve kterém jsme provedli přepočítání (5.35). Tento přepočítání je dán výrazy (5.42) \rightarrow (5.43) \rightarrow (5.44). V jednotlivých faktorech vystupuje člen s Λ_{t-1} . Tento člen nezávisí ani na y_t ani na θ , a představuje tedy konstantu. Určuje zbytek po minimalizaci kritéria nejmenších čtverců.

Výraz v (5.42) obsahuje součin dvou exponenciál. První z nich je hp y_t , podmíněná parametry θ , tj. model soustavy. Druhá je úměrná hp θ v čase $t-1$, tj. není podmíněná výstupem y_t . Součinem těchto dvou hp dostaneme sdruženou hp, která je vyjádřena v (5.43). Poslední výraz (5.44) dostaneme obráceným rozkladem sdružené hp. První exponenciála v tomto výrazu je hp θ podmíněná naměřenými daty y_t , a až na konstantu představuje přepočtenou hp parametrů v čase t . Druhá exponenciála je úměrná hp výstupu y_t a je nezávislá na parametrech θ . Je to marginální hp, která je funkcí starých dat, komprimovaných v odhadu parametrů $\hat{\theta}_{t-1}$ a statistice $\kappa(t)$. Ta odpovídá hp předpovědi, která odhaduje výstup jen na základě starých dat, nikoliv parametrů.

Apsteriorní hp pro t

V čase t má podle předchozího odvození normální rozdělení a po normování má tvar (5.7)

$$f(\theta | D(t)) = \frac{1}{\sqrt{2\pi/\kappa(t)}} \exp \left\{ -\frac{1}{2} \kappa(t) (\theta - \hat{\theta}_t)^2 \right\}, \quad (5.45)$$

což je - až na posun času - formálně stejný tvar, jako měla apsteriorní hp v čase $t-1$. Výpočet tedy probíhá rekurentně – krok, který jsme provedli, bude stejný ve všech časech.

Bodové odhady parametrů

Dostaneme je jako podmíněnou střední hodnotu parametrů, kterou spočteme pomocí aposteriorní hp z času t . Vyjádříme je jednak rekurentně (pomocí odhadu z minulého kroku), jednak pomocí jednorázového výpočtu (ze všech minulých dat najednou). Obecně pro nenulové počáteční statistiky platí (5.9)

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{\kappa(t)} (y_t - \hat{\theta}_{t-1}) = \frac{S_{y1}(t)}{\kappa(t)} = \frac{\sum_{\tau=-\kappa(0)+1}^t y_\tau}{\kappa(0) + t}, \quad (5.46)$$

kde $\hat{\theta}_0$ je apriorní odhad parametru θ a $\kappa(0)$ je počáteční statistika, která má význam počtu apriorních dat, tj. dat, z nichž byla čerpána apriorní informace.

Rozptyl odhadu parametrů je

$$C_t = \frac{1}{\kappa(t)}.$$

- ♣ Bodovým odhadem parametru θ je výběrový průměr z y .
- ♣ Rozptyl odhadu parametrů jde postupně k nule. S rostoucím počtem dat jsme schopni tento odhad určit se stále větší přesností.

Výsledek odpovídá známému poznatku z klasické statistiky: rozptyl výběrového průměru (odhad parametrů) je rozptyl základního souboru (šumu, zde 1), dělený rozsahem výběru (počtem dat).

Hp předpovědi

Je určena marginálou ze sdružené hp (5.43) po doplnění exponentu na čtverec v proměnné θ , podle (5.8). Oddělíme podmíněnou hp a po normování dostaneme

$$f(y_t | D(t-1)) = \frac{1}{\sqrt{2\pi \frac{\kappa(t)}{\kappa(t-1)}}} \exp \left\{ -\frac{1}{2} \frac{\kappa(t-1)}{\kappa(t)} (y_t - \hat{\theta}_{t-1})^2 \right\}. \quad (5.47)$$

Část sdružené hp (5.44), která obsahuje funkci Λ_{t-1} , je zahrnuta v konstantě.

Bodová předpověď

Je bodovým odhadem budoucího výstupu vzhledem k změřeným datům. Je to opět střední hodnota výstupu určená pomocí hp předpovědi (5.10)

$$\hat{y}_t = \hat{\theta}_{t-1} = \frac{1}{\kappa(t-1)} \sum_{\tau=-\kappa(0)+1}^{t-1} y_\tau. \quad (5.48)$$

- ♣ Bodovým odhadem budoucího výstupu je tedy, stejně jako pro parametr, průměrná hodnota všech dosud změřených dat (případně modifikovaná počátečními statistikami).

Rozptyl bodové předpovědi je

$$C_{\hat{y}} = \frac{\kappa(t)}{\kappa(t-1)} = \frac{\kappa(0) + t}{\kappa(0) + t - 1}.$$

- ♣ Rozptyl bodové předpovědi postupně klesá (čím více máme dat, tím přesněji jsme schopni předpovídat) a ustaluje se na jedničce. (S hodně velkým počtem dat narůstá jistota ve střední hodnotě – tj. odhadovaný parametr, a jediná nejistota je způsobena šumem s rozptylem jedna).

Funkce Λ_t

Z hlediska prosté metody nejmenších čtverců, kterou jsme se zabývali v části 2.3 (vztah (2.28)), je funkce Λ zbytek kritéria po minimalizaci. Z tohoto pohledu by tato funkce měla být rovna součtu kvadrátů reziduí. Ukážeme, že to tak je. Pro přehlednost použijeme znovu nulové počáteční podmínky. Pomocí označení použitého ve vztahu (5.41) můžeme psát

$$\begin{aligned}\Lambda_t &= S_y^2(t) - \frac{(S_{y1}(t))^2}{\kappa(t)} = \sum_{\tau=1}^t y_\tau^2 - \frac{\left(\sum_{\tau=1}^t y_\tau\right)^2}{t} = t\bar{y}_\tau^2 - t\bar{y}^2 = \\ &= t(\overline{y - \bar{y}})^2 = \sum_{\tau=1}^t (y_\tau - \bar{y})^2,\end{aligned}\quad (5.49)$$

kde $\bar{y} = \frac{1}{t} \sum_{\tau=1}^t y_\tau$ je operátorová symbolika pro výběrový průměr, $S_y^2(t) = t\bar{y}_\tau^2$ a $S_{y1}(t) = t\bar{y}$ jsou statistiky, vyjádřené v operátorové symbolice pro nulové počáteční podmínky.

Protože regresní přímka pro statický model je konstanta $y = \hat{\theta}_t$, a protože $\bar{y}_t = \hat{\theta}_t$, potvrzuje předchozí výsledek tvrzení, že Λ_t je součtem kvadrátů reziduí. V tomto příkladě $y_\tau - \hat{\theta}_t = \hat{e}_\tau$ je residuum v bodě τ pro parametry odhadnuté z dat až do času t . Tvrzení ale platí obecně.

♣ Hodnota funkce Λ_t je rovna součtu kvadrátů reziduí lineární regrese.

Poznámka: Stejný výsledek, opět pro nulové počáteční podmínky, lze snadno a rychle odvodit následujícím způsobem.

Exponent modelu je $(y_\tau - \theta)^2$. Exponent součtinu t modelů je

$$\begin{aligned}\sum_{\tau=1}^t (y_\tau - \theta)^2 &= \sum_{\tau=1}^t (y_\tau - \bar{y} + \bar{y} - \theta)^2 = \\ &= \sum_{\tau=1}^t (y_\tau - \bar{y})^2 - 2 \sum_{\tau=1}^t (y_\tau - \bar{y})(\theta - \bar{y}) + \sum_{\tau=1}^t (\theta - \bar{y})^2 = \\ &= \sum_{\tau=1}^t (\theta - \bar{y})^2 + \sum_{\tau=1}^t (y_\tau - \bar{y})^2,\end{aligned}$$

kde prostřední člen předposledního výrazu je roven nule, protože $\theta - \bar{y}$ je vzhledem ke sčítání konstanta a $\sum_{\tau=1}^t y_\tau = t\bar{y}$.

První člen posledního výrazu odpovídá exponentu aposteriorní hp, druhý člen exponentu hp předpovědi. Exponent této hp tvoří hodnota Λ_t .

B. Model s neznámým rozptylem

Protože rozptyl jako neznámý parametr má určité zvláštnosti proti regresním koeficientům a pro odvození odhadu vyžaduje některé zvláštní postupy, uvedeme nyní stručně ještě jednu předchozí vzorce, ale pro statický model s neznámým rozptylem $\sigma^2 = r$. Neznámé parametry nyní budou $\Theta = \{\theta, r\}$.

Model

Je zadán stejnou rovnicí jako v předchozím příkladu (5.36)

$$y_t = \theta + e_t, \quad (5.50)$$

ale hp modelu bude jako v (4.16)

$$f(y_t|\Theta) = \frac{1}{\sqrt{2\pi r}} \exp\left\{-\frac{1}{2r}(y_t^2 - 2\theta y_t + \theta^2)\right\} \quad (5.51)$$

s neznámými parametry $\Theta = \{\theta, r\}$.

Věrohodnostní funkce

Podle (5.5) s modelem (5.51) je (srv. (5.37))

$$L_t(\Theta) \propto r^{-\frac{t}{2}} \exp\left\{-\frac{1}{2r}\left(t\theta^2 - 2\theta \sum_{\tau=1}^t y_\tau + \sum_{\tau=1}^t y_\tau^2\right)\right\}. \quad (5.52)$$

Aposteriorní hp

Podle (5.7) je (srv. (5.45))

$$f(\Theta|D(t)) \propto r^{-\frac{\kappa(t)}{2}} \exp\left\{-\frac{1}{2r}\left[\kappa(t)(\theta - \hat{\theta}_t)^2 + \Lambda_t\right]\right\}. \quad (5.53)$$

Dalším potřebným krokem je integrace aposteriorní hp (5.53), která je úměrná sdružené hp (5.43), přes neznámé parametry a její normalizace. Tak dostaneme hp předpovědi. V předešlém případě se známým rozptylem měla aposteriorní hp normální rozdělení a integraci jsme provedli "známým trikem" pomocí doplnění exponentu v hp na úplný čtverec v proměnné, podle které máme integrovat. Tyto úpravy byly provedeny ve výrazech (5.42) - (5.44). V případě neznámého rozptylu má hp předpovědi Studentovo rozdělení a předchozí postup nelze použít. Musíme tedy "skutečně" integrovat. Nejdříve budeme integrovat podle θ (to je stejný krok jako v případě známého rozptylu), potom podle rozptylu r .

I_1 – integrace podle θ

$$\begin{aligned} I_1 &\propto r^{-\frac{\kappa(t)}{2}} \exp\left\{-\frac{\Lambda_t}{2r}\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{\kappa(t)}{2r}(\theta - \hat{\theta}_t)^2\right\} d\theta \propto \\ &\propto r^{-\frac{\kappa(t)}{2}} \exp\left\{-\frac{\Lambda_t}{2r}\right\} \left(\frac{r}{\kappa(t)}\right)^{\frac{1}{2}} = \kappa(t)^{-\frac{1}{2}} r^{-\frac{\kappa(t)-1}{2}} \exp\left\{-\frac{\Lambda_t}{2r}\right\}. \end{aligned}$$

I_2 – integrace podle r

$$\begin{aligned} I_2 &\propto \kappa(t)^{-\frac{1}{2}} \int_0^{\infty} r^{-\frac{\kappa(t)-1}{2}} \exp\left\{-\frac{\Lambda_t}{2r}\right\} dr = \\ &\left| \begin{array}{l} \frac{\Lambda_t}{2r} = z, \quad r = \frac{\Lambda_t}{2z} \\ -\frac{\Lambda_t}{2r^2} dr = dz, \quad \text{meze: } \infty \text{ a } 0 \end{array} \right| \\ &= \kappa(t)^{-\frac{1}{2}} \int_0^{\infty} \left(\frac{\Lambda_t}{2z}\right)^{-\frac{\kappa(t)-1}{2}} \exp\{-z\} \frac{2}{\Lambda_t} \left(\frac{\Lambda_t}{2z}\right)^2 dz = \\ &= \kappa(t)^{-\frac{1}{2}} \int_0^{\infty} \left(\frac{\Lambda_t}{2}\right)^{-\frac{\kappa(t)-1}{2}+2-1} z^{\frac{\kappa(t)-1}{2}-2} \exp\{-z\} dz = \\ &= \kappa(t)^{-\frac{1}{2}} \left(\frac{\Lambda_t}{2}\right)^{-\frac{\kappa(t)-3}{2}} \int_0^{\infty} z^{\frac{\kappa(t)-3}{2}-1} \exp\{-z\} dz = \end{aligned}$$

$$\begin{aligned}
&= \kappa(t)^{-\frac{1}{2}} \left(\frac{\Lambda_t}{2}\right)^{-\frac{\kappa(t)-3}{2}} \Gamma\left(\frac{\kappa(t)-3}{2}\right) = \\
&= \kappa(t)^{-\frac{1}{2}} 2^{\frac{\kappa(t)-3}{2}} \Gamma\left(\frac{\kappa(t)-3}{2}\right) \Lambda_t^{-\frac{\kappa(t)-3}{2}}.
\end{aligned}$$

Hp předpovědi

Vyjádříme-li Λ_t rekurzivně pomocí Λ_{t-1} , můžeme hp předpovědi (5.8) vyjádřit ve formě Studentova rozdělení jako funkci proměnné y_t (srv. (5.47))

$$f(y_t|D(t-1)) \propto \kappa(t)^{-\frac{1}{2}} 2^{\frac{\kappa(t)-3}{2}} \Gamma\left(\frac{\kappa(t)-3}{2}\right) \left\{ \frac{\kappa(t-1)}{\kappa(t)} (y_t - \hat{\theta}_{t-1})^2 + \Lambda_{t-1} \right\}^{-\frac{\kappa(t)-3}{2}}. \quad (5.54)$$

5.4.3 Statický regresní model s rovnoměrným šumem

V tomto příkladě ukážeme strukturu vzorců odhadu pro občas používaný užitečný a realistický model s rovnoměrně rozloženým šumem. Připomeňme, že rovnoměrné rozdělení se používá tam, kde množina realizací je zdola i shora omezená (patří do konečného intervalu) a na tomto intervalu nejsou žádné preference, všechny hodnoty jsou stejně pravděpodobné. Rovnoměrně rozdělená apriorní hp modeluje počáteční neznalost.

Základní schéma odhadování pro hp je stejné jako pro normální šum, konstrukce statistik pro akumulaci dat je však značně odlišná.

Model

Odhadovanou soustavu popíšeme modelem (4.17)

$$y_t = \theta + e_t$$

s hp (4.18)

$$f(y_t|\theta) = \frac{1}{2} \chi_{\langle \theta-1, \theta+1 \rangle}(y_t) = \begin{cases} \frac{1}{2} & \text{pro } y_t \in \langle \theta-1, \theta+1 \rangle, \\ 0 & \text{pro } y_t \notin \langle \theta-1, \theta+1 \rangle. \end{cases} \quad (5.55)$$

Součin modelů

Pro časové okamžiky $\tau = 1, 2, \dots, t$ určíme součin hp modelu tak, abychom odhadli strukturu apriorní a aposteriorní hp, které s tímto součinem musí být strukturálně shodné. Součin je

$$\prod_{\tau=1}^t f(y_\tau|\theta) = \prod_{\tau=1}^t \frac{1}{2} \chi_{\langle \theta-1, \theta+1 \rangle}(y_\tau). \quad (5.56)$$

Podle (5.55) musí pro $\tau = 1, 2, \dots, t$ platit $\theta - 1 \leq y_\tau \leq \theta + 1$,

a tedy $\theta - 1 \leq \min\{y_\tau\} \leq \max\{y_\tau\} \leq \theta + 1$. Odtud již bezprostředně plyne

$$\max\{y_\tau\} - 1 \leq \theta \leq \min\{y_\tau\} + 1, \quad \text{tj. } \theta \in \langle \max\{y_\tau\} - 1, \min\{y_\tau\} + 1 \rangle.$$

Odtud a z (5.56) dostáváme

$$\prod_{\tau=1}^t f(y_\tau|\theta) = \frac{1}{2^t} \chi_{\langle \max\{y_\tau\} - 1, \min\{y_\tau\} + 1 \rangle}(\theta), \quad (5.57)$$

protože součin konstantních funkcí je opět konstantní funkce.

Věrohodnostní funkce

Je podle (5.5) a (5.57)

$$L_t(\theta) = \frac{1}{2^t} \chi_{\langle \max\{y_\tau\} - 1, \min\{y_\tau\} + 1 \rangle}(\Theta). \quad (5.58)$$

Poznámka: Je-li $\max\{y_\tau\} - \min\{y_\tau\} > 2$, je $L_t(\theta) = 0$! Proč? ²

Apriorní hp

Volíme ji tak, aby měla strukturu věrohodnostní funkce, tj. rovnoměrnou, s apriorními statistickými B_0^L a B_0^U ,

$$f(\theta) = \frac{1}{B_0^U - B_0^L} \chi_{\langle B_0^L, B_0^U \rangle}(\Theta). \quad (5.59)$$

Aposteriorní hp

Dostaneme ji jako součin apriorní hp a věrohodnostní funkce (5.7)

$$f(\theta|Y(t)) = \frac{1}{B_t^U - B_t^L} \chi_{\langle B_t^L, B_t^U \rangle}(\Theta), \quad (5.60)$$

s označením $Y(t) = [y_t, y_{t-1}, \dots]$ obdobným jako pro $D(t)$ a

$$\begin{aligned} B_t^L &= \max\{B_0^L; \max\{y_\tau\}_{\tau=1}^t - 1\} = \max\{B_{t-1}^L; y_t - 1\}, \\ B_t^U &= \min\{B_0^U; \min\{y_\tau\}_{\tau=1}^t + 1\} = \min\{B_{t-1}^U; y_t + 1\}, \end{aligned} \quad (5.61)$$

s počáteční podmínkou B_0^L a B_0^U .

Bodový odhad parametrů

Tento odhad můžeme určit jako střední hodnotu s hp (5.60) podle (5.9)

$$\hat{\theta}_t = E[\theta|Y(t)] = \int \theta f(\theta|Y(t)) d\theta = \frac{B_t^L + B_t^U}{2}. \quad (5.62)$$

Skutečně,

$$\begin{aligned} \hat{\theta}_t &\stackrel{\text{definice}}{=} \int_{-\infty}^{\infty} \theta f(\theta|Y(t)) d\theta \stackrel{\text{dosazeno za hp}}{=} \int_{-\infty}^{\infty} \theta \frac{1}{B_t^U - B_t^L} \chi_{\langle B_t^L, B_t^U \rangle}(\Theta) d\theta = \\ &= \frac{1}{B_t^U - B_t^L} \int_{B_t^L}^{B_t^U} \theta d\theta \stackrel{\text{integrace přes } \theta}{=} \frac{1}{B_t^U - B_t^L} \left[\frac{1}{2} \theta^2 \right]_{B_t^L}^{B_t^U} = \\ &= \frac{1}{2} \frac{1}{B_t^U - B_t^L} \left[(B_t^U)^2 - (B_t^L)^2 \right] = \frac{B_t^U + B_t^L}{2}. \end{aligned}$$

Výsledek je v souladu s očekáváním. Při rovnoměrném rozdělení výskytu parametru na intervalu $[B_t^L, B_t^U]$ budeme odhadovat jeho správnou hodnotu uprostřed.

²Protože interval v argumentu indikátorové funkce v (5.58) je prázdný.

Hp předpovědi

Je to hp, která bude předpovídat budoucí výstup na základě minulých dat. Dostaneme ji podle obecného schématu (5.8)

$$\begin{aligned}
f(y_t|Y(t-1)) &\stackrel{\text{úplná pravděpodobnost}}{=} \int_{-\infty}^{\infty} f(y_t|\theta)f(\theta|Y(t-1)) d\theta = \\
&\stackrel{\text{dosazeno za hp}}{=} \int_{-\infty}^{\infty} \frac{1}{2} \chi_{(\theta-1, \theta+1)}(y_t) \frac{1}{B_{t-1}^U - B_{t-1}^L} \chi_{(B_{t-1}^L, B_{t-1}^U)}(\Theta) d\theta = \\
&= \frac{1}{2(B_{t-1}^U - B_{t-1}^L)} \int_{-\infty}^{\infty} \chi_{(\max\{B_{t-1}^L, y_t-1\}, \min\{B_{t-1}^U, y_t+1\})}(\Theta) d\theta = \\
&= \frac{1}{2} \frac{\min\{B_{t-1}^U, y_t+1\} - \max\{B_{t-1}^L, y_t-1\}}{B_{t-1}^U - B_{t-1}^L} = \\
&= \begin{cases} \frac{1}{2} \frac{y_t+1 - B_{t-1}^L}{B_{t-1}^U - B_{t-1}^L} & \text{pro } y_t \in (B_{t-1}^L - 1; B_{t-1}^U - 1), \\ \frac{1}{2} & \text{pro } y_t \in (B_{t-1}^U - 1; B_{t-1}^L + 1), \\ \frac{1}{2} \frac{B_{t-1}^U + 1 - y_t}{B_{t-1}^U - B_{t-1}^L} & \text{pro } y_t \in (B_{t-1}^L + 1; B_{t-1}^U + 1), \\ 0 & \text{jinak.} \end{cases} \tag{5.63}
\end{aligned}$$

Poznámka: Se stejným výsledkem bychom počítali i pomocí integrálů (5.16), pro které v tomto případě platí

$$I_t \propto 2^{\kappa_t} (B_t^U - B_t^L), \quad I_{t-1} \propto 2^{(\kappa_t-1)} (B_{t-1}^U - B_{t-1}^L).$$

Je tedy

$$f(y_t|Y(t-1)) = \frac{I_t}{I_{t-1}} = \frac{1}{2} \frac{B_t^U - B_t^L}{B_{t-1}^U - B_{t-1}^L}.$$

Použijeme-li definici mezi posteriorní hp (5.60), pak čitatel předchozího zlomku je

$$B_t^U - B_t^L = \min\{B_{t-1}^U; y_t+1\} - \max\{B_{t-1}^L; y_t-1\}.$$

Tak dostáváme stejný výraz jako v předchozím odstavci (před rozepsáním zlomku do jednotlivých intervalů), a tedy i stejný výsledek.

Bodová předpověď' výstupu

Je to střední hodnota výstupu, podmíněná minulými výstupy (5.10)

$$\hat{y}_t = E[y_t|Y(t-1)] = \int y_t f(y_t|Y(t-1)) dy_t = \hat{\theta}_{t-1}. \tag{5.64}$$

Skutečně, hodnota předpovídaného výstupu soustavy je dána podmíněnou střední hodnotou výstupu

$$\hat{y}_t = \int y_t f(y_t|Y(t-1)) dy_t,$$

kde hp předpovědi $f(y_t|Y(t-1))$ je (5.63). Tvzení (5.64) bychom mohli dokázat pomocí naznačené integrace přes jednotlivé úseky hp. Integrace je sice přímočará, ale poněkud zdlouhavá. Proto je dokážeme jinak.

Z funkčního předpisu (5.63) je patrné, že hp předpovědi je symetrická funkce se středem symetrie v bodě $m = \frac{B_{t-1}^L + B_{t-1}^U}{2}$. Dokážeme intuitivně zřejmé tvrzení, že střední hodnotou náhodné veličiny se symetrickou hp je střed symetrie.

Uvažujme symetrickou hp $f(x)$ se středem symetrie v bodě m a integrál

$$\int_{-\infty}^{\infty} (x - m)f(x) dx.$$

V integrálu provedeme substituci: $x - m = t$, $dx = dt$. Dostaneme

$$\int_{-\infty}^{\infty} tf(t + m) dt.$$

Integrand je nyní lichou funkcí, neboť je součinem liché funkce t a sudé funkce $f(t + m)$ (transformací se střed symetrie hp posunul do počátku). Podle známé věty o integraci liché funkce je tento integrál roven nule.

Vrátíme-li se nyní k jeho původnímu vyjádření, dostaneme

$$\int_{-\infty}^{\infty} (x - m)f(x) dx = \int_{-\infty}^{\infty} xf(x) dx - m = 0,$$

a tedy

$$\int_{-\infty}^{\infty} xf(x) dx = m.$$

Střed symetrie m je tedy střední hodnotou hp $f(x)$. Pro hp předpovědi je střední hodnotou, a tedy bodovou předpovědí, hodnota $\hat{y}_t = (B_{t-1}^L + B_{t-1}^U)/2$. Odtud a z (5.62) plyne dokazovaná rovnost (5.64).

Výsledek je opět to, co lze čekat. Výstup je modelován jako konstanta, na kterou je superponován šum s nulovou střední hodnotou. Nejlepším odhadem výstupu je tedy nejlepší odhad $\hat{\theta}_{t-1}$ parametru θ .

5.4.4 Statický regresní model s diskretním šumem

V příkladu se budeme zabývat náhodným pokusem se dvěma možnými výsledky V , které nastanou s pravděpodobnostmi θ a $1 - \theta$, kde $\theta \in (0, 1)$. Lze si představit házení falešnou korunou, kde θ a $1 - \theta$ jsou pravděpodobnosti padnutí rubu ($V = R$) a lícu ($V = L$). Pokusu přiřadíme náhodnou veličinu y a výsledkům hodnoty $y(R) = 0$ a $y(L) = 1$. Budeme-li pokus opakovat v čase $\tau = 1, 2, \dots, t$, bude

$$y_\tau \in \{0, 1\}, \quad f(0) = P(V_\tau = R) = \theta \quad \text{a} \quad f(1) = P(V_\tau = L) = 1 - \theta,$$

kde parametr modelu $\theta \in (0, 1)$ značí pravděpodobnost padnutí rubu (R) a f je pravděpodobnostní funkce diskretní náhodné veličiny y .

Měřená data jsou $D(t) = Y(t) = \{y_t, y_{t-1}, y_{t-2}, \dots\}$.

Model

Je popsán pomocí hp (4.21). Pro její přehlednější vyjádření a jednodušší rozšíření na náhodné veličiny s více než dvěma hodnotami vyjádříme tuto hp pomocí diskretního **indikátoru** δ (Kroneckerovo delta). Tato funkce je analogií indikátoru χ (5.55) pro množiny, použitého v předchozím příkladě. Je definován pro diskretní argument předpisem

$$\delta(i, j) = \begin{cases} 1 & \text{pro } i = j, \\ 0 & \text{pro } i \neq j. \end{cases} \quad (5.65)$$

S takto zavedeným značením můžeme hp modelu zapsat ve tvaru

$$f(y_\tau|\theta) = \theta^{\delta(y_\tau,0)}(1-\theta)^{\delta(y_\tau,1)}. \quad (5.66)$$

Součin modelů

Hledáme vhodný tvar pro vyjádření apriorní a aposteriorní hp ve tvaru součinu hp modelů

$$\prod_{\tau=1}^t \theta^{\delta(y_\tau,0)}(1-\theta)^{\delta(y_\tau,1)} = \theta^{\sum_{\tau=1}^t \delta(y_\tau,0)}(1-\theta)^{\sum_{\tau=1}^t \delta(y_\tau,1)}.$$

Zavedeme statistiky $\kappa_0(t)$ a $\kappa_1(t)$, kde $\kappa_0(t)$ udává počet padlých rubů a $\kappa_1(t)$ udává počet padlých líců. Pomocí indikátorové funkce je lze vyjádřit takto

$$\kappa_0(t) = \sum_{\tau=1}^t \delta(y_\tau, 0) \quad \text{a} \quad \kappa_1(t) = \sum_{\tau=1}^t \delta(y_\tau, 1).$$

Věrohodnostní funkce

S použitím zavedených statistik můžeme součin modelů vyjádřit jako funkci parametrů, a tedy věrohodnostní funkci podle (5.5)

$$L_t(\theta) = \theta^{\kappa_0(t)}(1-\theta)^{\kappa_1(t)}. \quad (5.67)$$

Apriorní hp

Volíme ji v podobném tvaru jako věrohodnostní funkci

$$f(\theta) = \frac{1}{K_0} \theta^{\kappa_0(0)-1} (1-\theta)^{\kappa_1(0)-1}, \quad (5.68)$$

kde $\kappa_0(0)$ a $\kappa_1(0)$ jsou apriorní statistiky a $1/K_0$ je normalizační konstanta.

P o z n á m k a: Minus jednička v exponentech je zavedena formálně pro pozdější vhodné označení pomocí beta funkce.

P o z n á m k a: Jiná možná volba statistik je např. počet padlých rubů κ_0 a počet pokusů t .

Aposteriorní hp

Dostaneme ji vynásobením apriorní hp a věrohodnostní funkce a normalizací (5.7)

$$f(\theta|Y(t)) = \frac{1}{K(t)} \theta^{\kappa_0(t)-1} (1-\theta)^{\kappa_1(t)-1}, \quad (5.69)$$

kde $\kappa_0(t)$ a $\kappa_1(t)$ jsou statistiky, pro které platí rekurze

$$\kappa_0(t) = \kappa_0(t-1) + \delta(y_t, 0), \quad \kappa_1(t) = \kappa_1(t-1) + \delta(y_t, 1) \quad (5.70)$$

s počátečními hodnotami $\kappa_0(0)$ a $\kappa_1(0)$; tj. po hodu mincí přičteme jedničku k té statistice, jejíž strana padla a $K(t) = \int \theta^{\kappa_0(t)-1} (1-\theta)^{\kappa_1(t)-1} d\theta = B(\kappa_0(t), \kappa_1(t))$, kde $B(\cdot, \cdot)$ je beta funkce, definovaná vztahem (11.3).

Bodový odhad parametru θ

Podle (5.9)

$$\hat{\theta}_t = E[\theta|Y(t)] = \int \theta f(\theta|Y(t)) d\theta = \frac{\kappa_0(t)}{\kappa_0(t) + \kappa_1(t)}. \quad (5.71)$$

Skutečně, za aposteriorní hp dosadíme z (5.69) a dostaneme

$$\begin{aligned} \hat{\theta}_t &= \int_0^1 \theta f(\theta|Y(t)) d\theta = \\ &= \frac{1}{B(\kappa_0(t), \kappa_1(t))} \int_0^1 \theta \times \theta^{\kappa_0(t)-1} (1-\theta)^{\kappa_1(t)-1} d\theta = \\ &= \frac{1}{B(\kappa_0(t), \kappa_1(t))} \int_0^1 \theta^{\kappa_0(t)} (1-\theta)^{\kappa_1(t)-1} d\theta = \\ &= \frac{B(\kappa_0(t) + 1, \kappa_1(t))}{B(\kappa_0(t), \kappa_1(t))} = \frac{\kappa_0(t)}{\kappa_0(t) + \kappa_1(t)}, \end{aligned}$$

kde při přechodu z třetí na čtvrtou řádku jsme využili definici beta funkce $B(\kappa_0(t) + 1, \kappa_1(t))$ (11.3) a vlastnosti beta funkce $B(x+1, y)/B(x, y) = x/(x+y)$ (11.5).

Hp předpovědi

Dostaneme ji integrací sdružené hp pro výstup a parametr dané pomocí hp modelu a apriorní hp z kroku $t-1$ podle (5.8)

$$f(y_t|Y(t-1)) = \begin{cases} \frac{\kappa_0(t-1)}{\kappa_0(t-1) + \kappa_1(t-1)} = \hat{\theta}_{t-1} & \text{pro } y_t = 0, \\ \frac{\kappa_1(t-1)}{\kappa_0(t-1) + \kappa_1(t-1)} = 1 - \hat{\theta}_{t-1} & \text{pro } y_t = 1. \end{cases} \quad (5.72)$$

Skutečně, použijeme opět model (5.66) a aposteriorní hp (5.69) v čase $t-1$

$$\begin{aligned} f(y_t|Y(t-1)) &= \int f(y_t|\theta) f(\theta|Y(t-1)) d\theta = \\ &= \int_0^1 \theta^{\delta(y_t, 0)} (1-\theta)^{\delta(y_t, 1)} \frac{1}{B(\kappa_0(t-1), \kappa_1(t-1))} \theta^{\kappa_0(t-1)-1} (1-\theta)^{\kappa_1(t-1)-1} d\theta = \\ &= \frac{1}{B(\kappa_0(t-1), \kappa_1(t-1))} \int_0^1 \theta^{\kappa_0(t-1)+\delta(y_t, 0)-1} (1-\theta)^{\kappa_1(t-1)+\delta(y_t, 1)-1} d\theta = \\ &= \begin{cases} \frac{1}{B(\kappa_0(t-1), \kappa_1(t-1))} \int_0^1 \theta^{\kappa_0(t-1)-1} (1-\theta)^{\kappa_1(t-1)} d\theta = \frac{\kappa_0(t-1)}{\kappa_0(t-1) + \kappa_1(t-1)} & \text{pro } y_t = 0, \\ \frac{1}{B(\kappa_0(t-1), \kappa_1(t-1))} \int_0^1 \theta^{\kappa_0(t-1)} (1-\theta)^{\kappa_1(t-1)-1} d\theta = \frac{\kappa_1(t-1)}{\kappa_0(t-1) + \kappa_1(t-1)} & \text{pro } y_t = 1. \end{cases} \end{aligned}$$

Bodová předpověď výstupu

Je to střední hodnota výstupu vypočtená pomocí hp předpovědi podle (5.10)

$$\hat{y}_t = E[y_t|Y(t-1)] = 1 - \hat{\theta}_{t-1}. \quad (5.73)$$

Jelikož tato hp je diskrétní, musíme místo integrace použít sumaci

$$\begin{aligned} \hat{y}_t = E[y_t|Y(t-1)] &= \sum_{y_t=0}^1 y_t f(y_t|Y(t-1)) = 0 \frac{\kappa_0(t-1)}{\kappa_0(t-1) + \kappa_1(t-1)} + 1 \frac{\kappa_1(t-1)}{\kappa_0(t-1) + \kappa_1(t-1)} = \\ &= \frac{\kappa_1(t-1)}{\kappa_0(t-1) + \kappa_1(t-1)} = 0 \hat{\theta}_{t-1} + 1(1 - \hat{\theta}_{t-1}) = 1 - \hat{\theta}_{t-1}. \end{aligned}$$

Poznámka: Pokud bychom chtěli, aby předpověď byla přípustnou hodnotou, mohli bychom za předpověď vzít např. maximálně pravděpodobnou hodnotu. Kritérium optimality je v tomto případě argument maxima (tj. modus) hp předpovědi

$$\hat{y}_t = \begin{cases} 0 & \text{pro } \hat{\theta}_{t-1} \geq 0.5, \\ 1 & \text{pro } \hat{\theta}_{t-1} < 0.5. \end{cases}$$

Všimněme si, že pro poškozenou minci, kdy skutečné pravděpodobnosti rubu a líce nejsou stejné, bude odhad $\hat{\theta}$ trvale buď větší nebo menší než 0.5. Potom předpověď bude trvale ta strana, která má vyšší pravděpodobnost. Tento výsledek konec konců odpovídá zdravému rozumu.

5.4.5 Modely s diskrétními parametry

Budeme se zabývat poněkud umělým příkladem modelu s diskrétním parametrem. O neznámém parametru modelu předpokládáme, že může nabývat jen konečný počet různých hodnot. Předpokládáme, že pravděpodobnosti těchto hodnot jsou známé. Odhadováním takových pravděpodobností jsme se zabývali v předchozím příkladě 5.4.4, věnovaném odhadu spojitého parametru z diskrétních dat. Známe tedy rozdělení pravděpodobnosti parametrů (stará aposteriorní hp) a naším úkolem je před změřením nového výstupu určit jeho předpověď a na základě změřeného výstupu provést přepočítání této hp na novou aposteriorní hp. Na jejím základě odhadneme skutečnou hodnotu parametru, která se podílela na generování změřeného výstupu.

Abychom dobře demonstrovali poněkud odlišné struktury hp a práci s nimi, budeme uvažovat oba modely (4.24) a (4.26), uvedené v části 4.2. Jsou to modely s diskrétním parametrem, a jak s diskrétním, tak i se spojitým signálem.

A. Diskrétní parametr i signál**Model**

Popíšeme jej pomocí (4.24), tj.

$$y = \theta + e,$$

s diskrétním šumem e i parametrem θ , jejichž rozdělení je zadáno tabulkami

$$\frac{e}{f(e)} \begin{array}{c|cc} & -1 & 1 \\ \hline & 0.4 & 0.6 \end{array}, \quad \frac{\theta}{f(\theta)} \begin{array}{c|ccc} & -2 & 0 & 2 \\ \hline & 0.2 & 0.5 & 0.3 \end{array}.$$

Tento model má podmíněnou diskretní hp (4.25)

$f(y \theta)$			
$y \setminus \theta$	-2	0	2
-3	0.4	0	0
-1	0.6	0.4	0
1	0	0.6	0.4
3	0	0	0.6

Hp předpovědi

Je to hp výstupu, která není podmíněná znalostí parametrů (5.8)

$$\begin{aligned} f(y) &= \sum_{\theta} f(y|\theta)f(\theta) = \\ &= f(y|\theta = -2)f_{\theta}(-2) + f(y|\theta = 0)f_{\theta}(0) + f(y|\theta = 2)f_{\theta}(2) \end{aligned} \quad (5.74)$$

Poznámka: Z rozepsaného vzorce je dobře vidět princip Bayesovské statistiky. Kdybychom znali správnou hodnotu θ , byl by výstup předpovídán jedinou hp. Protože správnou hodnotu θ neznáme, je pro nás náhodnou veličinou, pro kterou známe pouze množinu všech možných hodnot a jejich pravděpodobností. Proto vezmeme v úvahu všechny podmíněné hp, vážíme je jejich pravděpodobnostmi $f(\theta)$ a ze všech dohromady sestavíme předpověď.

Zavedené hp ve formě tabulek lze výhodně reprezentovat pomocí vektorů a matic. Toto vyjádření chápeme tak, že (např. v následujícím vyjádření) pro $y = i$ a $\theta = j$ je $f(y|\theta) = f_{i,j}$, kde f je příslušná matice.

$$f(y|\theta) = \begin{bmatrix} 0.4 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 0.6 \end{bmatrix} \quad \text{a} \quad f(\theta) = \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix}.$$

Potom předchozí vzorec lze vyjádřit jako součin matice a vektoru

$$\begin{aligned} f(y) &= \sum_{j=1}^3 f(y|\theta_j)f(\theta_j) = \begin{bmatrix} 0.4 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 0.6 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix} = \\ &= 0.2f(y|\theta = -2) + 0.5f(y|\theta = 0) + 0.3f(y|\theta = 2) = \begin{bmatrix} 0.08 \\ 0.32 \\ 0.42 \\ 0.18 \end{bmatrix}. \end{aligned}$$

♣ Pravděpodobnost předpovědi je vážená kombinace jednotlivých modelů (hp), váhy jsou pravděpodobnosti parametrů (tj. pravděpodobnosti jednotlivých modelů).

Bodová předpověď výstupu

Podle (5.10) je

$$\hat{y} = E[y] = \sum_y yf(y) = [-3, -1, 1, 3] \begin{bmatrix} 0.08 \\ 0.32 \\ 0.42 \\ 0.18 \end{bmatrix} = 0.4. \quad (5.75)$$

- ♣ Bodová předpověď je vážená kombinace všech hodnot y , váhy jsou pravděpodobnosti předpovědí.

Poznámka: Kdyby se požadovalo, aby předpovídaná hodnota byla skutečně přípustnou hodnotou, mohli bychom jako předpověď vzít hodnotu s maximální pravděpodobností

$$\hat{y} = \arg \max f(y) = 1.$$

Odhad parametru

Máme řešit následující úlohu (která odpovídá konstrukci aposteriorní hp (5.7)). Naměřili jsme výstup $y = -1$. Jaká hodnota parametru θ a s jakou pravděpodobností byla realizována?

Známe $f(y|\theta)$ a $f(\theta)$. Sdružená hp je $f(y, \theta) = f(y|\theta)f(\theta)$ a marginální $f(y) = \sum_{\theta} f(y|\theta)f(\theta)$. Sdruženou hp počítáme tak, že vždy celý sloupec hp modelu násobíme příslušným prvkem vektoru $f(\theta)$. Hodnoty marginální hp $f(y)$ dostáváme jako součty prvků řádkových vektorů sdružené hp $f(y, \theta)$.

$y \backslash \theta$	-2	0	2	$f(y)$
-3	0.08	0	0	0.08
-1	0.12	0.2	0	0.32
1	0	0.3	0.12	0.42
3	0	0	0.18	0.18

Tabulka sdružené hustoty $f(y, \theta)$

Protože hledáme $f(\theta|y = -1)$, sestrojíme podmíněnou hp $f(\theta|y) = \frac{f(y, \theta)}{f(y)}$ tak, že normujeme řádky sdružené hp, tj. každý řádek sdružené hp dělíme příslušným prvkem marginální hp $f(y)$. Dostaneme

$y \backslash \theta$	-2	0	2	
-3	1	0	0	← $f(\theta y = -3)$
-1	0.375	0.625	0	← $f(\theta y = -1)$
1	0	0.714	0.286	← $f(\theta y = 1)$
3	0	0	1	← $f(\theta y = 3)$

Tabulka podmíněné hustoty $f(\theta|y)$

Z druhého řádku tabulky dostáváme hledanou podmíněnou pravděpodobnost

$$\frac{\theta}{f(\theta|y = -1)} \mid \begin{array}{ccc} -2 & 0 & 2 \\ 0.375 & 0.625 & 0 \end{array},$$

a tím i odpověď na naši otázku.

- ♣ Hodnota výstupu $y = -1$ vzešla nejpravděpodobněji z hodnoty $\theta = 0$ (s pravděpodobností 0.625), nebo méně pravděpodobně z hodnoty $\theta = -2$ (s pravděpodobností 0.375),

nebo

- ♣ Na hodnotě $y = -1$ se 62.5% podílela hodnota $\theta = 0$ a 37.5% hodnota $\theta = -1$.

B. Diskrétní parametr a spojité signál

Dále si všimneme práce s modelem, který má diskrétní parametr, ale pracuje se spojitými funkcemi. Hp takového modelu je kombinovaná, viz obrázek 4.2, v některých proměnných diskrétní a v jiných spojitá. Představuje parametrizovaný systém spojitých funkcí.

Model

Je popsán předpisem (4.26)

$$y = \theta + e,$$

kde diskrétní parametr θ má rozdělení

$$\begin{array}{c|ccc} \theta & -2 & 0 & 2 \\ \hline f(\theta) & 0.2 & 0.5 & 0.3 \end{array}.$$

Model má podmíněnou kombinovanou hp (4.27)

$$f(y|\theta) = \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y - \theta)^2 \right\}, \quad \theta \in \{-2, 0, 2\}.$$

Sdružená hp

Pro jednotlivé hodnoty θ se tato hp vypočte podle vzorce $f(y, \theta) = f(\theta)f(y|\theta)$ tak, že každou spojitou hp modelu násobíme příslušným prvkem vektoru $f(\theta)$

$$f(y, \theta) = \begin{cases} 0.2f(y|\theta = -2) & = \frac{0.2}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y + 2)^2 \right\} & \text{pro } \theta = -2, \\ 0.5f(y|\theta = 0) & = \frac{0.5}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}y^2 \right\} & \text{pro } \theta = 0, \\ 0.3f(y|\theta = 2) & = \frac{0.3}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y - 2)^2 \right\} & \text{pro } \theta = 2. \end{cases} \quad (5.76)$$

Hp předpovědi

Je to marginální hp z příslušné sdružené (5.8) a můžeme ji určit podle vzorce pro úplnou pravděpodobnost

$$f(y) = \sum_j f(y, \theta_j) = \sum_j f(y|\theta_j)f(\theta_j).$$

Dostaneme ji jako součet prvků z předchozího vyjádření sdružené hp (5.76)

$$f(y) = \frac{1}{\sqrt{2\pi}} \left[0.2 \exp \left\{ -\frac{1}{2}(y + 2)^2 \right\} + 0.5 \exp \left\{ -\frac{1}{2}y^2 \right\} + 0.3 \exp \left\{ -\frac{1}{2}(y - 2)^2 \right\} \right]. \quad (5.77)$$

Bodová předpověď výstupu

Je to opět střední hodnota $\hat{y} = E[y] = \int_{-\infty}^{\infty} y f(y) dy$ podle (5.10)

$$\begin{aligned} \hat{y} &= 0.2E[y|\theta = -2] + 0.5E[y|\theta = 0] + 0.3E[y|\theta = 2] = \\ &= 0.2 \times (-2) + 0.5 \times 0 + 0.3 \times 2 = 0.2. \end{aligned} \quad (5.78)$$

Aposteriorní hp

Obecně je dána v (5.7). Zde ji dostaneme ze sdružené hp podle vzorce $f(\theta|y) = \frac{f(\theta,y)}{f(y)}$ tak, že hodnoty sdružené hp dělíme hodnotou příslušné marginální hp $f(y)$. Formálně ji můžeme zapsat takto (y_0 je pevné y z podmínky)

$$f(\theta|y = y_0) = \begin{cases} \frac{1}{f(y_0)} \frac{0.2}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y_0 + 2)^2 \right\} & \text{pro } \theta = -2, \\ \frac{1}{f(y_0)} \frac{0.5}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}y_0^2 \right\} & \text{pro } \theta = 0, \\ \frac{1}{f(y_0)} \frac{0.3}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(y_0 - 2)^2 \right\} & \text{pro } \theta = 2. \end{cases} \quad (5.79)$$

Kapitola 6

Řízení při známých parametrech modelu

V této kapitole navážeme na 2. kapitolu, v níž jsme se zabývali základními výpočetními principy optimálního řízení založeného na kvadratickém kritériu optimality – viz část 2.2. Pro větší názornost jsme tam uvažovali deterministický model. Zde naznačíme odvození syntézy řízení pro stochastický model obecného řádu (4.7) se známými parametry. Řízením s modelem, jehož parametry neznáme, se budeme zabývat v následující kapitole. O problematice řízení, založeného na dynamickém programování, kterým se zde zabýváme, dosti podrobně pojednává seriál článků [7]. Ten lze doporučit pro hlubší studium.

Při syntéze řízení, podobně jako tomu bylo i u odhadování, je možno sledovat dvě úrovně, srv. algoritmy na str. 71 a str. 74. Na jedné úrovni obdržíme rekurzivní algoritmus pro hustoty pravděpodobností, aniž specifikujeme jejich konkrétní tvar (rozdělení a jeho statistiky). Takový algoritmus je ale prakticky nerealizovatelný, protože počítače neumějí pracovat s funkcemi na úrovni proměnných. Proto jsme nuceni nejdříve zadat konkrétní tvar rozdělení jednotlivých hustot pravděpodobností (zde budeme předpokládat normální rozdělení) a pak algoritmus pro hustoty pravděpodobností převést na algoritmus pro jejich statistiky. Odvození tohoto algoritmu na druhé úrovni je shodné s odvozením algoritmu řízení pro deterministický model, kterým jsme se zabývali ve 2. kapitole. Rovněž výsledky odvození jsou pro odpovídající si modely a kritéria prakticky shodné. Jediným rozdílem je, že zbytek kritéria po minimalizaci je pro stochastický model větší, neboť je do něho kromě chyb řízení zahrnut ještě rozptyl šumu, který nelze řídit.

Vlastní algoritmus sestává ze dvojice kroků, které se vždy v každém časovém okamžiku opakují. Jsou to kroky

Středování, při němž se pomocí modelu řízené soustavy počítá střední hodnota příslušné části kritéria (té části, která obsahuje nejnovější data a zbytku kritéria z předchozí minimalizace). Tento krok převádí stochastickou úlohu na deterministickou.

Minimalizace, která hledá optimální hodnotu nejnovějšího členu akční veličiny, pomocí které soustavu řídíme. Tento krok koresponduje s minimalizací kritéria pro deterministický model.

6.1 Odvození v integrální formě

Model

Syntézu řízení odvodíme pro stejný tvar modelu, jaký jsme uvažovali v kapitole 4, tj. pro model (4.7)

$$y_t = \varphi_t^T \theta + e_t, \quad (6.1)$$

kde y_t je výstup modelu a θ jsou parametry modelu,

$\varphi_t = [u_t, y_{t-1}, u_{t-1}, \dots, y_{t-\nu}, u_{t-\nu}]^T$ je regresní vektor,

u_t je vstup a e_t je porucha modelu (šum).

Budeme používat konvence pro značení integrálů ze strany 65 a datové historie zavedené v (5.1).

Kritérium

Uvažované kritérium nyní zahrnuje signály ve formě posloupností náhodných veličin a má tvar (srv. (2.8))

$$J = E \left[\sum_{\tau=1}^N \{y_\tau^2 + \omega u_\tau^2\} | D(0) \right], \quad (6.2)$$

kde $E[\bullet | D(0)]$ značí operátor podmíněné střední hodnoty definovaný pomocí hp, podmíněné apriorními daty $f(\bullet | D(0))$. Např. pro středovanou funkci $h(x)$ je

$$E[h(x) | D(0)] = \int h(x) f(x | D(0)) dx.$$

A. Syntéza řízení pro hp

Naším úkolem je nalézt v obecném případě posloupnost hp $f(u_t | D(t-1))$ pro $t = 1, \dots, N$, které minimalizují zvolené kritérium. Postup minimalizace jen naznačíme, podobně, jako jsme to udělali v části 2.2 na horizontu řízení $N = 3$.

Hledáme tedy posloupnost hp

$$f_1, f_2, f_3, \quad f_\tau = f(u_\tau | D(\tau-1)) \quad \text{pro } \tau = 1, 2, 3$$

tak, aby bylo dosaženo minima

$$\begin{aligned} J &= \min_{f_1, f_2, f_3} E \left[\sum_{\tau=1}^3 (y_\tau^2 + \omega u_\tau^2) | D(0) \right] = \\ &= \min_{f_1, f_2, f_3} \int \sum_{\tau=1}^3 (y_\tau^2 + \omega u_\tau^2) f(y_3, u_3, y_2, u_2, y_1, u_1 | D(0)) d(y_3, u_3, y_2, u_2, y_1, u_1), \end{aligned}$$

kde symbol $\int \dots d(y_3, u_3, \dots, y_1, u_1)$ představuje integraci přes celé obory hodnot veličin $y_3, u_3, \dots, y_1, u_1$.

Minimalizaci začneme podobně jako v části 2.2 od konce, tj. pro časový okamžik $t = N = 3$. V kritériu oddělíme veličiny s indexem 3 a sdruženou hp rozdělíme podle řetězového pravidla (3.2)

$$J = \min_{f_1, f_2, f_3} \int \int \left[\sum_{\tau=1}^2 (y_\tau^2 + \omega u_\tau^2) + (y_3^2 + \omega u_3^2) \right] \times$$

$$\times f(y_3, u_3|D(2)) d(y_3, u_3) f(y_2, u_2, y_1, u_1|D(0)) d(y_2, u_2, y_1, u_1).$$

První část integrace se týká pouze veličin v čase $t = 3$. V součtu (po vynásobení závorky v integrandu) můžeme integrovat každý sčítanec zvlášť, a protože první sčítanec nezávisí na veličinách v čase 3, zůstane po integraci nezměněn. Také operaci \min_{f_3} můžeme přesunout až k druhému sčítanci, neboť jen v něm je volená hp $f(u_3|D(2))$. Dostaneme

$$J = \min_{f_1, f_2} \int \left\{ \sum_{\tau=1}^2 (y_\tau^2 + \omega u_\tau^2) + \min_{f_3} \int (y_3^2 + \omega u_3^2) f(y_3, u_3|D(2)) d(y_3, u_3) \right\} \times \\ \times f(y_2, u_2, y_1, u_1|D(0)) d(y_2, u_2, y_1, u_1).$$

Integrál v druhém sčítanci ve složené závorce označíme J_3 , dostáváme

$$J_3 = \int (y_3^2 + \omega u_3^2) f(y_3, u_3|D(2)) d(y_3, u_3) = \\ = \int \int (y_3^2 + \omega u_3^2) f(y_3|u_3, D(2)) f(u_3|D(2)) dy_3 du_3 = \\ = \int E [y_3^2 + \omega u_3^2|u_3, D(2)] f(u_3|D(2)) du_3,$$

kde $E[\cdot|u_3, D(2)]$ je podmíněná střední hodnota, která je funkcí dat z podmínky, tedy u_3 a $D(2)$, tj. veličin $u_3, y_2, u_2, y_1, u_1, y_0, u_0, \dots$.

Poznámka: Zmíněný integrál představuje střední hodnotu součtu $y_3^2 + \omega u_3^2$ vzhledem k náhodné veličině y_3 . Pro hp $f(y_3|u_3, D(2))$, definovanou regresním modelem $y_3 = \varphi_3^T \theta + e_3$ nebo jeho speciálním tvarem prvního řádu (2.7), který jsme uvažovali ve 2. kapitole, bude výsledek středování shodný (až na rozptyl šumu) s dosazením rovnice deterministického modelu za náhodnou veličinu y_3 . Tento obrat jsme používali ve 2. kapitole k tomu, abychom veličinu y_3 předpověděli (zbavili se jí jako neznámé veličiny), protože byla v daném okamžiku neznámá. V tomto bodě se tedy setkává deterministická i stochastická optimalizace a pojmy v nich užívané: "dosazení modelu" a "středování". Oba dva pojmy představují **předpověď**, na níž je řízení (podobně jako odhadování) založeno.

Po provedeném středování se dostáváme k druhému kroku optimalizace v čase $t = 3$ a tím je minimalizace. Máme určit hp $f(u_3|D(2))$ tak, abychom minimalizovali kritérium J . Protože z celého kritéria na $f(u_3|D(2))$ závisí pouze člen J_3 , hledáme minimum tohoto členu. Toho dosáhneme tehdy, když nalezneme minimum střední hodnoty v integrandu a do něho soustředíme celou hp $f(u_3|D(2))$ ve formě Diracova impulsu (definovaného vztahem $\delta(x) = 0$ pro $x \neq 0$ a $\int \delta(x) dx = 1$), tedy

$$f(u_3|D(2)) = \delta(u_3 - u_3^*),$$

kde

$$u_3^* = \arg \min_{u_3} E [y_3^2 + \omega u_3^2|u_3, D(2)].$$

Dosadíme-li za hp $f(u_3|D(2))$ Diracovu funkci $\delta(u_3 - u_3^*)$, představuje integrace ve výrazu J_3 pouze dosazení hodnoty u_3^* za u_3 do střední hodnoty v integrandu. Výsledek minimalizace je tedy

$$\min_{u_3} J_3 = E [y_3^2 + \omega (u_3^*)^2|u_3^*, D(2)] = J_3^*(D(2)).$$

Tímto je ukončena minimalizace v čase $t = 3$. Výraz J_3^* je zbytek po této minimalizaci a je funkcí dat do času 2. Dosadíme jej zpět do kritéria J a budeme pokračovat minimalizací v čase $t = 2$

$$J = \min_{f_1, f_2} \int \left(\sum_{\tau=1}^2 (y_\tau^2 + \omega u_\tau^2) + J_3^* \right) f(y_2, u_2, y_1, u_1|D(0)) d(y_2, u_2, y_1, u_1).$$

Postup minimalizace kritéria J pro čas 2 bude obdobný, jako pro čas 3

$$\begin{aligned} J &= \min_{f_1, f_2} \int \int \left[\sum_{\tau=1}^1 (y_\tau^2 + \omega u_\tau^2) + (y_2^2 + \omega u_2^2) + J_3^* \right] \times \\ &\quad \times f(y_2, u_2|D(1)) f(y_1, u_1|D(0)) d(y_2, u_2) d(y_1, u_1) = \\ &= \min_{f_1} \int \left\{ y_1^2 + \omega u_1^2 + \min_{f_2} \int (y_2^2 + \omega u_2^2 + J_3^*) f(y_2, u_2|D(1)) d(y_2, u_2) \right\} \times \\ &\quad \times f(y_1, u_1|D(0)) d(y_1, u_1). \end{aligned}$$

Druhý sčítanec v integrandu upravíme analogicky k předchozímu kroku

$$\begin{aligned} J_2 &= \int \int (y_2^2 + \omega u_2^2 + J_3^*) f(y_2|u_2, D(1)) f(u_2|D(1)) dy_2 du_2 = \\ &= \int E [y_2^2 + \omega u_2^2 + J_3^*|u_2, D(1)] f(u_2|D(1)) du_2. \end{aligned}$$

Minima dosáhneme opět volbou $f(u_2|D(1)) = \delta(u_2 - u_2^*)$ a minimalizací střední hodnoty. Opět platí

$$\begin{aligned} u_2^* &= \arg \min_{u_2} E [y_2^2 + \omega u_2^2 + J_3^*|u_2, D(1)] \\ \min_{u_2} J_2 &= E [y_2^2 + \omega (u_2^*)^2 + J_3^*(D^*(2))|u_2^*, D(1)] = J_2^*(D(1)). \end{aligned}$$

Zde $D^*(2)$ označuje data do času 2, kde za akční veličinu u_2 je dosazena její optimální hodnota u_2^* .

Po zpětném dosazení do kritéria J obdržíme

$$J = \min_{f_1} \int \left(\sum_{\tau=1}^1 (y_\tau^2 + \omega u_\tau^2) + J_2^* \right) f(y_1, u_1|D(0)) d(y_1, u_1),$$

což je formálně stejný tvar jako na začátku předchozího kroku postupné minimalizace, jen s tím rozdílem, že indexy jsou o jedničku posunuty dolů. Stejnými úpravami bychom tedy dostali

$$\begin{aligned} J_1 &= \int E [y_1^2 + \omega u_1^2 + J_2^*|u_1, D(0)] f(u_1|D(0)) du_1, \\ f(u_1|D(0)) &= \delta(u_1 - u_1^*), \quad \text{kde } u_1^* = \arg \min_{u_1} E [y_1^2 + \omega u_1^2 + J_2^*|u_1, D(0)], \\ \min_{u_1} J_1 &= E [y_1^2 + \omega (u_1^*)^2 + J_2^*(D^*(1))|u_1, D(0)] = J_1^*(D(0)), \end{aligned}$$

což je hledaná výsledná minimální hodnota kritéria J .

Nyní popíšeme obecný algoritmus minimalizace kritéria pro optimální řízení na horizontě délky N . Tento algoritmus obsahuje obdobné kroky jako levá část algoritmu řízení ze str. 23. Formální období však dosáhneme až po specifikaci modelu pro normální rozdělení.

ALGORITMUS

SYNTÉZA OPTIMÁLNÍHO ŘÍZENÍ

Start počáteční hodnota kritéria

$$J_{N+1}^* = 0$$

for $\tau = N : -1 : 1$

středování

$$\begin{aligned} E_\tau &= E \left[y_\tau^2 + \omega u_\tau^2 + J_{\tau+1}^* | u_\tau, D(\tau - 1) \right] = \\ &= \int \left(y_\tau^2 + \omega u_\tau^2 + J_{\tau+1}^* \right) f(y_\tau | u_\tau, D(\tau - 1)) dy_\tau \end{aligned}$$

minimalizace

$$J_\tau^* = \min_{u_\tau} E_\tau, \quad u_\tau^* = \arg \min_{u_\tau} E_\tau$$

end %for

Optimální hodnoty řízení $u_1^*, u_2^*, \dots, u_N^*$

Minimální hodnota kritéria $J = J_1^*$

B. Syntéza řízení pro statistiky

Předchozí algoritmus ukazuje rekurzivní postup minimalizace kritéria pro hp. Konkrétní tvar těchto hustot zatím nebyl zadán ani vyžadován. V každém kroku se přitom počítá střední hodnota

$$E \left[y_\tau^2 + \omega u_\tau^2 + J_{\tau+1}^* | u_\tau, D(\tau - 1) \right]$$

a její minimalizace podle u_τ . Jak jsme již uvedli, tento algoritmus nelze přímo realizovat, neboť jako proměnné se v něm vyskytují funkce (hp náhodných veličin, pro které se počítají střední hodnoty). Teprve když pro zmíněné hp zvolíme konkrétní tvar rozdělení, který tyto hustoty popisuje, můžeme převést uvedený algoritmus pro hp na algoritmus pro statistiky. Ten již místo s funkcemi (hp) pracuje s číselnými proměnnými (statistikami hp).

Použijeme-li pro definici potřebné hp $f(y_t | u_t, D(t - 1))$ již zavedený regresní model $y_t = \varphi_t^T \theta + e_t$ (4.7), kde $e_t \sim N(0, \sigma^2)$, můžeme spočítat střední hodnotu kvadrátu výstupu y_t (4.10)

$$\begin{aligned} E \left[y_t^2 | u_t, D(t - 1) \right] &= E \left[(\theta^T \varphi_t + e_t)^2 | u_t, D(t - 1) \right] = \\ &= E \left[(\theta^T \varphi_t)^2 + 2\theta^T \varphi_t e_t + e_t^2 | u_t, D(t - 1) \right] = \left(\theta^T \varphi_t \right)^2 + \sigma^2, \end{aligned}$$

neboť $E[e_t | u_t, D(t - 1)] = 0$ a $E[e_t^2 | u_t, D(t - 1)] = \sigma^2$.

Z předchozího závěru je vidět, že výsledek je stejný jako při dosažení rovnice deterministického modelu $y_t = \theta^T \varphi_t$ za výstup y_t . Oba výsledky se liší pouze rozptylem šumu σ^2 , který ale výběr optimální strategie neovlivní. Pouze zvětší hodnotu minima kritéria na horizontě délky N o $N\sigma^2$.

Algoritmus minimalizace kritéria pro statistiky hp je tedy formálně shodný s tím, který jsme odvodili na str. 23 pro deterministický model.

6.2 Odvození v operátorovém vyjádření

Algoritmus pro hp, který jsme uvedli na str. 97, lze celý vyjádřit pomocí operátoru podmíněné střední hodnoty. Toto vyjádření je stručné a přehledné, není ale tak průhledné jako při použití integrálního vyjádření střední hodnoty.

Nejprve ukážeme, jak se v operátorovém tvaru projeví podmíněnost střední hodnoty, když rozvedeme sdružené hp (3.1) pomocí řetězového pravidla (3.2). Uvažujme střední hodnotu funkce $h(u, v)$ dvou náhodných veličin u a v se sdruženou hp $f(u, v)$

$$\begin{aligned} E[h(u, v)] &= \int \int h(u, v) f(u, v) du dv = \\ &= \int \left(\int h(u, v) f(u|v) du \right) f(v) dv = E [E[h(u, v)|v]]. \end{aligned}$$

S použitím tohoto pravidla vyjádříme jeden krok minimalizace kritéria. Budeme sledovat minimalizaci pro čas 2 z předchozího odvození na str. 95, protože tento krok je možno již považovat za obecný krok rekurzivní minimalizace (je v něm již obsažen zbytek z předchozího kroku). Abychom tuto obecnost vyjádřili, budeme místo času 2 psát obecný čas t . Zároveň budeme hp akčních zásahů uvažovat ve formě Diracových impulzů (deterministické akční zásahy) tak, jak jsme to již odvodili.

$$\begin{aligned} J &= \min_{u_1, u_2, \dots, u_t} E \left[\sum_{\tau=1}^t (y_\tau^2 + \omega u_\tau^2) + J_{t+1}^* | D(0) \right] = \\ &= \min_{u_1, u_2, \dots, u_t} E \left[E \left[\sum_{\tau=1}^{t-1} (y_\tau^2 + \omega u_\tau^2) + y_t^2 + \omega u_t^2 + J_{t+1}^* | D(1, \dots, t-1) \right] | D(0) \right] = \\ &= \min_{u_1, u_2, \dots, u_{t-1}} E \left[\sum_{\tau=1}^{t-1} (y_\tau^2 + \omega u_\tau^2) + \min_{u_t} E \left[y_t^2 + \omega u_t^2 + J_{t+1}^* | u_t, D(1, \dots, t-1) \right] | D(0) \right]. \end{aligned}$$

Označíme

$$J_t = E \left[y_t^2 + \omega u_t^2 + J_{t+1}^* | u_t, D(t-1) \right]$$

a hledáme minimum J_t přes u_t

$$J_t^* = \min_{u_t} J_t, \quad u_t^* = \arg \min_{u_t} J_t.$$

Tvar kritéria se po dosazení minima J_t reprodukuje

$$J = \min_{u_1, u_2, \dots, u_{t-1}} E \left[\sum_{\tau=1}^{t-1} (y_\tau^2 + \omega u_\tau^2) + J_t^* | D(0) \right].$$

Výsledek, který jsme dostali, je shodný s tím, který jsme odvodili pomocí integrálů a hp.

Kapitola 7

Řízení při neznámých parametrech modelu

Dosud jsme se zabývali úlohami odhadování a řízení více méně odděleně. Odvodili jsme algoritmy pro odhad parametrů modelu řízené soustavy (str. 29, nebo 71 a 74). Odhady parametrů jsme vyjádřili pomocí podmíněné hustoty parametrů (aposteriorní hp). Dále můžeme pomocí aposteriorní hp vypočítat bodové odhady parametrů. Jejich tvar závisí na kritériu, podle kterého je určujeme. Pro nejčastěji uvažované kvadratické kritérium je bodovým odhadem parametrů jejich podmíněná střední hodnota (5.11).

- ♣ Odhady parametrů můžeme vyjádřit buď pomocí aposteriorní hp parametrů nebo jejich číselnou hodnotou jako bodové odhady.

Pro syntézu řízení jsme uvažovali parametry známé. Optimální řízení, minimalizující kvadratické kritérium, jsme v každém kroku řízení obdrželi minimalizací podmíněné střední hodnoty součtu příslušné části kritéria a zbytku z minimalizace v předchozím kroku (str. 23 nebo 97).

- ♣ Syntéza řízení je odvozena pro známé parametry.

Jsou-li parametry soustavy neznámé, je třeba provádět úlohu odhadu i řízení současně. Protože se ale řízení počítá vždy od aktuálního času pro celý horizont dopředu, vyžaduje odhady parametrů (což je obecně aposteriorní hp z Bayesova vztahu (5.18)), které nejsou k dispozici ale jsou vyjádřeny jako funkce ještě neznámých dat. Takový výpočet je prakticky nerealizovatelný – místo s hodnotami veličin pracuje s celými funkcemi. To současná výpočetní technika zatím nedovoluje, a numerická realizace těchto výpočtů je neúnosně dlouhá. Řízení, při kterém bychom optimálně odhadovali i řídili se nazývá duální.

- ♣ V případě optimálního řízení s modelem soustavy, jehož parametry nejsou známy a mají být průběžně odhadovány, je třeba provést aproximaci, která odhadování a řízení oddělí.

Nyní ukážeme dva nejčastěji používané postupy, které umožní provést prakticky realizovatelnou syntézu řízení – viz opět [7]. V obou případech budeme při získávání odhadu parametrů pro řízení a při vlastní syntéze řízení postupovat tzv. **metodou klouzavého horizontu**. Stanovíme určitý časový horizont, tj. počet kroků odhadování a řízení, které směřují od současného okamžiku dopředu. Pro jednotlivé okamžiky tohoto intervalu počítáme od konce intervalu předpis

pro výpočet vstupu, tzv. **řídící zákon**, tj. předpis pro generování vstupu. Tento předpis se postupně ustaluje (jeho koeficienty se mění stále méně). Vezmeme poslední předpis a použijeme jej pro jeden krok řízení. Parametry modelu soustavy se odhadují průběžně, tj. s použitím všech dat, která jsme dosud naměřili. Pro konstrukci řídicího zákona na celém intervalu použijeme jejich aktuální reprezentaci (bodovou nebo jejich hp). Potom horizont posuneme o jeden krok dopředu a celou proceduru opakujeme.

Poznámka: Tento postup je zvláště výhodný pro statické modely. Ukázali jsme, že pro ně nevznikají při postupné minimalizaci kritéria zbytky, které by přecházely do dalšího kroku minimalizace. V důsledku toho je řídicí zákon ustálen již při prvním kroku a postačí zvolit horizont délky jedna. Tím se vyhneme nepříjemným iteracím na horizontě.

7.1 Důvěřivá strategie řízení

Nejjednodušší přístup k syntéze řízení s neznámými parametry spočívá v nahrazení skutečných parametrů jejich bodovými odhady. Potom celá syntéza zůstává tak, jak byla prezentována v předešlé kapitole pro známé parametry. Při realizaci řídicího zákona se parametry nahradí jejich aktuálními bodovými odhady (5.11).

- ♣ Důvěřivá strategie znamená bezprostřední kombinaci odhadování a řízení se známými parametry, kdy za parametry modelu pro řízení použijeme bodové odhady, získané odhadování parametrů.

Tuto metodu stručně ilustrujeme na příkladě.

Řízení s důvěřivou strategií

Pro ilustraci zvolíme nejjednodušší možný model, a to statický (4.12). Stejný model použijeme i pro další, tzv. opatrnou strategii. Ta je totiž komplikovanější a při použití složitějšího modelu by se úvahy staly nepřehlednými.

Model

Model soustavy bude podle (4.12)

$$y_t = \theta u_t + e_t, \quad (7.1)$$

kde $e_t \sim N(0, \sigma^2)$, σ^2 je známá konstanta pro všechna t . Za kritérium řízení volíme (srv. (6.2))

$$J = E \left[\sum_{t=1}^N \{ (y_t - w_t)^2 + \omega u_t^2 \} \mid D(0) \right], \quad (7.2)$$

kde ω je penalizace akční veličiny a w_t je zadaná žádaná hodnota výstupu, to je hodnota, kterou má nabýt řízený výstup soustavy. (Řízení je nejlepší, je-li argument kvadrátu nula, pak je $y_t = w_t$.)

Bodové odhady parametru

Obecný vzorec pro aktualizaci aposteriorní hp je uveden v 5. kapitole v poznámce ke vztahu (5.25). Jejím rekurentním výpočtem a maximalizací jsou dány bodové odhady parametru (5.31).

Algoritmus výpočtu bodových odhadů lze vyjádřit pomocí dvou následujících algoritmů. První algoritmus, který se opírá přímo o informační matici, je jednodušší a průhlednější, druhý, ve kterém se přepočítávají přímo bodové odhady a jejich charakteristiky, má lepší výpočetní vlastnosti.

Algoritmus I: Vychází z rekurzivního přepočtu informační matice (5.26) a její struktury (5.28). Při něm se přepočítávají charakteristiky měřených dat a v každém kroku počítají odhady parametrů, případně další charakteristiky.

Přepočet submatic informační matice:

$$\begin{aligned} S_y^2(t) &= S_y^2(t-1) + y_t^2, \\ S_{yu}^2(t) &= S_{yu}^2(t-1) + y_t u_t, \\ S_u^2(t) &= S_u^2(t-1) + u_t^2. \end{aligned} \quad (7.3)$$

s počáteční podmínkou $S_y^2(0), S_{yu}(0), S_u^2(0)$ danými volbou apriorní hp.

Výpočet odhadu parametrů:

$$\hat{\theta}_t = \frac{S_{yu}(t)}{S_u^2(t)}. \quad (7.4)$$

Kovariance odhadu parametrů:

$$C_t = \frac{1}{S_u^2(t)}. \quad (7.5)$$

Algoritmus II: Zde se rekurzivně přepočítávají přímo odhady parametrů a statistiky k tomu potřebné. Výhodou tohoto algoritmu je, že provádí přímý přepočet kovarianční matice odhadu parametrů a nemusí ji v každém kroku znovu počítat jako inverzi statistiky S_u^2 . To se projeví zejména u dynamických modelů, kdy S_u^2 je matice. Označíme.

$$C_{t-1}u_t^2 = \zeta_t.$$

Potom rekurze má podle (11.10) a (11.9) tvar

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{1 + \zeta_t} C_{t-1} u_t (y_t - \hat{\theta}_{t-1} u_t), \quad (7.6)$$

$$C_t = C_{t-1} - \frac{1}{1 + \zeta_t} (C_{t-1} u_t)^2. \quad (7.7)$$

Poznámka: Odvození vzorců Algoritmu II z Algoritmu I je uvedeno v (11.9) až (11.11). V obecném případě vyžaduje odvození těchto vzorců jemnější úvahy z maticového počtu, ale postup je obdobný.

Syntéza řízení

Pro určení vstupu u_t máme minimalizovat podmíněnou střední hodnotu – viz algoritmus na str. 97

$$\min_{u_t} E[(y_t - w_t)^2 + \omega u_t^2 + J_{t+1}^* | u_t, D(t-1)] \quad (7.8)$$

s koncovou podmínkou $J_{N+1}^* = 0$.

Podmíněná hp, kterou pro výpočet střední hodnoty potřebujeme, je dána modelem, v němž za neznámé parametry dosadíme jejich bodové odhady (5.30)

$$f(y_t | u_t, D_t) \sim f(y_t | u_t, D_t, \Theta = \hat{\Theta}_{t-1}) = N(\hat{\theta}_{t-1} u_t; \sigma^2).$$

V minimalizovaném kritériu provedeme naznačené umocnění a s použitím (4.9), (4.10) dostaneme

$$\min_{u_t} \{ \hat{\theta}_{t-1}^2 u_t^2 + \sigma^2 - 2\hat{\theta}_{t-1} u_t w_t + w_t^2 + \omega u_t^2 + E[J_{t+1}^* | u_t, D(t-1)] \}.$$

Hledáme minimum, např. pomocí derivace. Protože lze obecně ukázat, že pro statický model je zbytek kritéria J_{t+1}^* pro všechny časové okamžiky t nezávislý na u_t , bude derivace jeho střední hodnoty podle u_t rovna nule. Optimální řízení u_t^* pak musí splňovat rovnici

$$\hat{\theta}_{t-1}^2 u_t^* - \hat{\theta}_{t-1} w_t + \omega u_t^* = 0,$$

odkud plyne

$$u_t^* = \frac{\hat{\theta}_{t-1}}{\omega + \hat{\theta}_{t-1}^2} w_t, \quad (7.9)$$

pro $t = 1, 2, \dots, N$.

Poznámka: Odvozené závěry můžeme srovnat s výsledky z první části 2. kapitoly, str. 21 až 23. Tam jsme uvažovali vícekrokové řízení na určitém horizontu, tady je horizont roven jedné. Žádaná hodnota v kapitole 2 byla identicky nulová.

Poznámka: Z výsledku je patrné, že pro apriorní odhad parametru $\hat{\theta}_0 = 0$ dojde nutně ke "katastrofě". Hodnota vypočteného řízení bude nula, soustava nebude vybudena, apriorní odhad se tedy nezmění a celý systém odhadování a řízení se z tohoto stavu nemůže "vzpamatovat".

7.2 Opatrná strategie řízení

Nepřekvapuje, že jednoduchost předešlé strategie je na úkor její obecnosti. Skutečně, tato strategie používá bodové odhady parametrů a "spoléhá na ně", jako by to byly opravdu správné parametry (srv. bodové a intervalové odhady). Odtud také plyne její název "důvěřivá". Tyto parametry jsou ale ve skutečnosti jen odhady. Nemusí, a také jistě nebudou přesně souhlasit se skutečnou hodnotou parametru. Proto je lépe reprezentovat odhady parametrů jejich aposteriorními hp. V nich je zahrnut nejen bodový odhad parametru, ale také jeho rozptyl. Syntéza potom počítá s tím, že odhady jsou určitá konkrétní čísla, ale jsou to jen bodové odhady s neurčitostí, která je dána jejich rozptylem. Taková syntéza je při generování řízení "opatrnější". Odtud její název.

Syntéza řízení, dávající opatrnou strategii, je složitější. Není to jen prostá kombinace bodových odhadů z úlohy odhadování a algoritmu řízení se známými parametry. Naznačíme její odvození.

Základním krokem při postupné minimalizaci kritéria je výpočet střední hodnoty

$$\begin{aligned} J_t^* &= \min_{u_t} E[(y_t^2 + \omega u_t^2 + J_{t+1}^*) | u_t, D(t-1)] = \\ &= \min_{u_t} \int (y_t^2 + \omega u_t^2 + J_{t+1}^*) f(y_t | u_t, D(t-1)) dy_t. \end{aligned} \quad (7.10)$$

Pro výpočet střední hodnoty potřebujeme hp předpovědi $f(y_t | u_t, D(t-1))$. Tu však přímo známe jen pro model se známými parametry. Tam se totiž hp předpovědi přímo rovná hp definované modelem, neboť v ní parametry jako proměnné nevystupují. Pro model (7.1) s neznámými parametry θ (σ^2 uvažujeme známé) musíme hp předpovědi dopočítat – viz (5.15)

$$f(y_t | u_t, D(t-1)) = \int f(y_t | \varphi_t, \Theta) f(\Theta | D(t-1)) d\Theta, \quad (7.11)$$

kde první hp na pravé straně je dána modelem a druhá je hp parametrů, která je vyvíjena v odhadování (5.17).

Pro jednorázový výpočet hp předpovědi je výhodné použít její vyjádření jako podíl Bayesových integrálů stejně jako v (5.16)

$$f(y_t|u_t, D(t-1)) = \frac{I_t}{I_{t-1}}. \quad (7.12)$$

Bayesův integrál je

$$I_t = \int L_t(\Theta) f(\Theta|D(0)) d\Theta, \quad (7.13)$$

kde integrujeme přes součin věrohodnostní funkce (5.5) a apriorní hp (5.6). Odvození vztahu pro jednorázový výpočet hp předpovědi je uvedeno v dodatku (11.4).

Poznámka: Vztah (7.11) pro hp předpovědi je přesně tím místem, kde se setkávají odhad a řízení a vytvářejí základy adaptivního řízení. První hp pod integrálem představuje model se známými parametry, se kterým bychom mohli řídit, pokud bychom parametry znali. Druhá hp přináší informace o neznámých parametrech.

Protože odvození opatrné řídicí strategie je v obecném případě dosti obtížné, budeme postup opět demonstrovat na příkladě. Zvolíme ten nejjednodušší možný příklad, abychom obešli formálně komplikovaný maticový počet. I navzdory značné jednoduchosti použitého modelu budou jak struktura odvození, tak i jednotlivé vzorce dobře naznačovat obecný postup.

Řízení s opatrnou strategií

Model

Volíme jej stejný jako pro důvěřivou strategii (7.1)

$$y_t = \theta u_t + e_t,$$

kde opět $e_t \sim N(0, \sigma^2)$, σ^2 je známá konstanta. Volíme rovněž stejné kritérium řízení (7.2)

$$J = \sum_{t=1}^N (y_t - w_t)^2 + \omega u_t^2,$$

kde opět w_t je žádaná hodnota výstupu a ω je penalizace vstupu.

Hp předpovědi

Pro konstrukci této hp použijeme jednorázový výpočet (7.12), jehož základem je Bayesův integrál (7.13). Ten budeme počítat podle odstavce 5.4.2, kde jsme pro podobný typ modelu uvedli tvar věrohodnostní funkce (tj. součinu modelů) a pro odpovídající tvar apriorní hp. Rozdíl mezi oběma modely, tj. (5.36) a nyníjším modelem, je jen v řízení u_t , které nyníjší model obsahuje.

Podle definice věrohodnostní funkce (5.5) a vztahu (5.37), který upravíme pro náš model a doplníme na čtverec (11.1), zjistíme, že věrohodnostní funkce bude

$$L_t(\Theta) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right)^t \exp \left\{ -\frac{1}{2\sigma^2} [S_u^2(t)(\theta - \hat{\theta}_t)^2 + \Lambda_t] \right\}, \quad (7.14)$$

kde

$$\hat{\theta}_t = \frac{S_{yu}(t)}{S_u^2(t)} \quad \text{a} \quad \Lambda_t = S_y^2(t) - \frac{(S_{yu}(t))^2}{S_u^2(t)}$$

a

$$S_y^2(t) = \sum_{\tau=1}^{\tau} y_{\tau}^2, \quad S_{yu}(t) = \sum_{\tau=1}^{\tau} y_{\tau} u_{\tau}, \quad S_u^2(t) = \sum_{\tau=1}^{\tau} u_{\tau}^2.$$

Apriorní hp zvolíme pro jednoduchost jako rovnoměrnou (což odpovídá nulové apriorní informaci). Bayesův integrál je potom úměrný přímo integrálu z věrohodnostní funkce

$$\begin{aligned} I_t &\propto \int \frac{1}{(2\pi\sigma^2)^{\frac{t}{2}}} \exp\left\{-\frac{1}{2\sigma^2}[S_u^2(t)(\theta - \hat{\theta}_t)^2 + \Lambda_t]\right\} d\theta = \\ &\propto (2\pi\sigma^2)^{-\frac{t}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_t\right\} \int \exp\left\{-\frac{1}{2\sigma^2/S_u^2(t)}(\theta - \hat{\theta}_t)^2\right\} d\theta = \\ &\propto (2\pi\sigma^2)^{-\frac{t}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_t\right\} \sqrt{2\pi\sigma^2/S_u^2(t)} = \\ &\propto (2\pi\sigma^2)^{-\frac{t-1}{2}} \exp\left\{\frac{-1}{2\sigma^2}\Lambda_{t-1}\right\} (S_u(t))^{-1} \exp\left\{\frac{-1}{2\sigma^2} \frac{S_u^2(t-1)}{S_u^2(t)} (y_t - \hat{\theta}_{t-1}u_t)^2\right\}, \end{aligned} \quad (7.15)$$

kde při poslední úpravě jsme využili rekurentního vyjádření vzorce pro Λ_t

$$\Lambda_t = \Lambda_{t-1} + \frac{S_u^2(t-1)}{S_u^2(t)} (y_t - \hat{\theta}_{t-1}u_t)^2,$$

odvozeného v (11.9) až (11.11) .

Podle (7.12) je hp předpovědi dána podílem Bayesových integrálů (k je konstanta úměrnosti)

$$\begin{aligned} f(y_t|u_t, D(t-1)) &= \frac{I_t}{I_{t-1}} = \\ &= \frac{k(2\pi\sigma^2)^{-\frac{t-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_{t-1}\right\} (S_u(t))^{-1} \exp\left\{-\frac{1}{2\sigma^2} \frac{S_u^2(t-1)}{S_u^2(t)} (y_t - \hat{\theta}_{t-1}u_t)^2\right\}}{k(2\pi\sigma^2)^{-\frac{t-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_{t-1}\right\} (2\pi\sigma^2)^{\frac{1}{2}} (S_u(t-1))^{-1}} = \\ &= \frac{1}{\sqrt{2\pi\sigma^2 S_u^2(t)/S_u^2(t-1)}} \exp\left\{-\frac{1}{2\sigma^2 S_u^2(t)/S_u^2(t-1)} (y_t - \hat{\theta}_{t-1}u_t)^2\right\}, \end{aligned} \quad (7.16)$$

kde I_{t-1} neobsahuje y_t , a tedy tvoří jen normalizační konstantu. Dostaneme jej ve vhodném tvaru jako integrál z I_t přes y_t .

Dostali jsme hp s normálním rozdělením se střední hodnotou $\hat{y}_t = \hat{\theta}_{t-1}u_t$, což je bodová předpověď výstupu, a rozptylem $\sigma^2 S_u^2(t)/S_u^2(t-1)$. Bodová předpověď výstupu \hat{y}_t , určená podle kvadratického kritéria, se rovná podmíněné střední hodnotě, a tedy bude

$$\hat{y}_t = \hat{\theta}_{t-1}u_t. \quad (7.17)$$

Je to vlastně rovnice modelu, kde za šum je dosazena jeho střední hodnota, což je nula, a za neznámé parametry jejich bodové odhady. Bodová předpověď je tedy stejná jako výpočet budoucí hodnoty výstupu z deterministického modelu s náhradou $\hat{\theta}$ za θ .

Poznámka: Bodovou předpověď výstupu jsme opět uvedli pouze pro úplnost. K odvození řídicího zákona podle opatrné strategie potřebujeme jen hp předpovědi.

Syntéza řízení

Pro určení akčních veličin použijeme opět obecné kritérium (7.8) jako v případě důvěřivé strategie

$$\min_{u_t} E[(y_t - w_t)^2 + \omega u_t^2 + J_{t+1}^* | u_t, D(t-1)]$$

s koncovou podmínkou $J_{N+1}^* = 0$. Podmíněná hp, kterou potřebujeme pro výpočet střední hodnoty, je nyní hp předpovědi (7.16). Rekurzivní výpočet statistik (7.3) a vztah (7.5) dávají rovnost $S_u^2(t)/S_u^2(t-1) = 1 + C_{t-1}u_t^2$ se kterou dostaneme

$$f(y_t | u_t, D_t) = N(\hat{\theta}_{t-1}u_t; \sigma^2(1 + C_{t-1}u_t^2)).$$

Podobným postupem jako při důvěřivé strategii nalezneme minimum

$$\min_{u_t} \{ \hat{\theta}_{t-1}^2 u_t^2 + \sigma^2 + \sigma^2 C_{t-1} u_t^2 - 2\hat{\theta}_{t-1} u_t w_t + w_t^2 + \omega u_t^2 + E[J_{t+1}^* | u_t, D(t-1)] \}.$$

Optimální řízení u_t^* je nyní dáno vztahem

$$\hat{\theta}_{t-1}^2 u_t^* + \sigma^2 C_{t-1} - \hat{\theta}_{t-1} w_t + \omega u_t^* = 0,$$

odkud plyne

$$u_t^* = \frac{\hat{\theta}_{t-1}}{\omega + \hat{\theta}_{t-1}^2 + \sigma^2 C_{t-1}} w_t$$

pro $t = 1, 2, \dots, N$.

Výsledek je podobný vztahu (7.9), který jsme obdrželi pro důvěřivou strategii. Jediný člen, který je tady navíc, je $\sigma^2 C_{t-1}$. Ten vnáší do řízení vliv rozptylu odhadu parametrů, tedy jejich nepřesnosti.

Poznámka: Budou-li odhady parametrů přesně rovny skutečným parametrům (tak, jak se tváříme při důvěřivé strategii), tj. bude-li $C_{t-1} = 0$, přejde odvozené řízení na to, které jsme dostali při důvěřivé strategii.

Poznámka: Jak pro důvěřivou tak i pro opatrnou strategii jsme obdrželi tzv. jednokrokové řízení (horizont, který jsme potřebovali pro ustálení řídicího zákona v každém kroku byl jedna). Je to dáno použitým modelem, který je statický. Tato varianta byla zvolena pro jednoduchost jinak dosti složité úlohy, abychom ji mohli přehledně demonstrovat na příkladě. Při použití dynamického modelu bychom v každém kroku minimalizace obdrželi zbytek závisející na předchozích akčních veličinách. Ten by vstupoval do dalšího kroku středování a minimalizace a vnášel by dynamiku do výpočtu řídicího zákona. K jeho ustálení bychom museli použít delší horizont, což úlohu syntézy značně komplikuje. Takovou syntézu jsme ukázali v 6. kapitole.

Kapitola 8

Odhadování multimodelů

Na rozdíl od předcházejících kapitol, kdy jsme pro popis sledované soustavy uvažovali jediný regresní model, budeme nyní jako základ popisu systémů uvažovat celou množinu různých regresních modelů, mezi kterými budeme přepínat. V různých okamžicích mohou tedy modelovanou veličinu popisovat různé modely. Tak lze kombinovat modely se stejnou strukturou, lišící se pouze v parametrech, ale také modely zcela odlišné [8, 9]. Zmíněnému modelu budeme říkat **multimodel** a jeho odhadem se budeme zabývat v této kapitole.

Význam multimodelů se ukazuje v popisu soustav, jejichž dynamická struktura se během odhadu skokově mění, nebo jako přiblížení modelům s proměnnými parametry, kdy místo spojitě změny parametrů uvažujeme skokově se měnící parametry s konečným počtem různých hodnot.

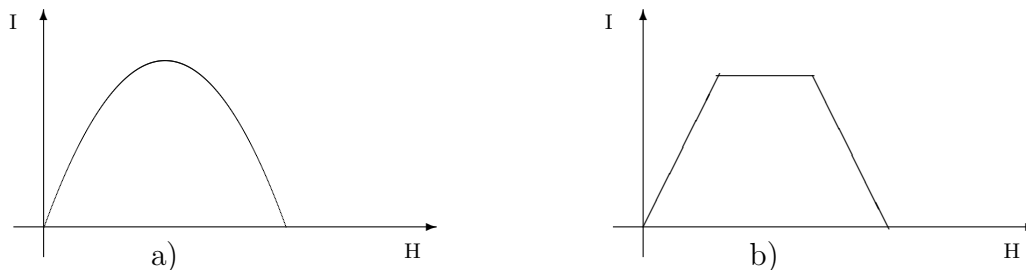
V této kapitole se budeme zabývat výhradně odhadem. Budeme proto uvažovat neřízený model, ve kterém místo dříve uvažovaných signálů výstupu y_t a vstupu u_t vystupují data d_t , $t = 1, 2, \dots, N$.

8.1 Multimodely a jejich popis

Motivace pro použití multimodelů

Mezi reálnými soustavami tvoří významnou třídu soustavy nelineární, jejichž chování lze přibližně popsat pomocí skupiny lineárních modelů tak, že podle polohy pracovního bodu nelineární soustavy je v daném okamžiku aktivní právě jeden z uvažovaných lineárních modelů.

Jako příklad uveďme modelování závislosti intenzity (I) dopravního toku na jeho hustotě (H) v jednom místě dopravní komunikace. Tato závislost má tvar známé křivky podobné parabole, jak je znázorněno na obrázku 8.1 a).



Obrázek 8.1: Nelineární model a jeho linearizace

Takový model je nelineární. Můžeme jej ale lokálně linearizovat tak, že jej nahradíme skupinou lineárních modelů (úseček) např. tak, jak ukazuje obrázek 8.1 b), kde

- rostoucí úsečka představuje oblast volného provozu,
- konstantní úsečka popisuje oblast nasycení a
- klesající úsečka je obrazem kolabující dopravní situace.

V každém okamžiku leží ideální pracovní bod $[H, I]$ na některé úsečce, a tedy, stav je v každém okamžiku popsán některým lineárním modelem.

Z uvedeného příkladu je patrné, že multimodel je tvořen celou množinou modelů, tzv. **komponent multimodelu**. V následujícím odstavci popíšeme modely jednotlivých komponent, v dalším pak veličinu, která modely svazuje a vytváří z nich vlastní multimodel.

Modely komponent

Jednotlivé komponenty multimodelu lze vyjádřit jako běžné regresní modely (4.3) ve formě podmíněných hustot pravděpodobnosti (hp) indexovaných pořadovým číslem komponenty

$$\mathcal{K}_c : f(d_t | D(t-1), c, \Theta_c) = f_c(d_t | \varphi_t, \Theta_c), \quad c \in \{1, 2, \dots, n\} \quad (8.1)$$

kde

c označuje pořadové číslo, kterým je komponenta indexována; n značí počet komponent multimodelu,

d_t je modelovaná veličina v čase t ,

φ_t je regresní vektor, obsahující data v okamžiku t známá, na kterých závisí modelovaná veličina d_t a který pro jednoduchost uvažujeme společný pro všechny komponenty,

Θ značí vektor, obsahující parametry všech komponent,

Θ_c je subvektor vektoru Θ , který obsahuje parametry příslušející c -té komponentě.

V konečném tvaru modelu komponenty je vyjádřena podmínka nezávislosti modelovaných dat d_t na datech starších než je řád modelu a skutečnost, že každá komponenta c závisí pouze na svých parametrech Θ_c .

Pro další práci s modely komponent budeme předpokládat, že patří do **exponenciální třídy**, a to obecně, zatím bez další konkretizace. Je tedy

$$\mathcal{K}_c : f_c(d_t | \varphi_t, \Theta_c) = \exp\{q_c^T(\Theta) s_c(\phi_t)\}, \quad c = 1, 2, \dots, n, \quad (8.2)$$

kde každá komponenta má svou vlastní parametrickou funkci q_c a obecně také svou vlastní datovou funkci s_c , závisející na tzv. **rozšířeném regresním vektoru** $\phi_t = [d_t, \varphi_t^T]^T$. Parametrická funkce q a datová funkce s jsou n rozměrné sloupcové vektory, kde n je dimenze statistiky. Datová funkce s bývá často společná pro všechny komponenty.

Poznámka: Exponenciální třída hp [2, 6] je poměrně široká. Kromě rovnoměrného sem patří všechna rozdělení, kterými jsme se zde zabývali. Vyjádření hp v uvedeném tvaru exponenciální třídy umožní převod funkcionálního algoritmu ze str. 71 na algoritmus pro statistiky, str. 74, tedy na algoritmus prakticky realizovatelný.

Snad trochu neobvykle, ale v souladu s dalším postupem a formálně velmi výhodně [10], vyjádříme model jedné vybrané komponenty (ve shodě s dalším ji označíme symbolem c_t) v následujícím součinovém tvaru

$$\mathcal{K}_c : f_{c_t}(d_t | \varphi_t, \Theta_{c_t}) = \prod_{c=1}^n f_c(d_t | \varphi_t, \Theta_c)^{\delta(c, c_t)} = \exp \left\{ \sum_{c=1}^n q_c^T(\Theta) s_c(\phi_t) \delta(c, c_t) \right\}. \quad (8.3)$$

Zde c je opět pouhý index, kterým jsou komponenty označovány a

$\delta(c, c_t)$ je Kroneckerova funkce (5.65), definována předpisem

$$\delta(c, c_t) = \begin{cases} 1 & \text{pro } c = c_t, \\ 0 & \text{pro } c \neq c_t. \end{cases}, \quad c \in c^* \quad (8.4)$$

Tuto funkci budeme používat i v dalších modelech a v této kapitole ji budeme nazývat **ukazatel**, protože index c , pro který má tato funkce hodnotu jedna, ukazuje na vybranou komponentu.

Model přepínače

Aby uvedená množina modelů (komponent) tvořila multimodel, je třeba definovat ještě další veličinu. Její hodnota ukazuje na komponentu, která je v daném časovém okamžiku aktivní. Tuto náhodnou veličinu označíme C_t a budeme ji nazývat **přepínač**. Její hodnotu v okamžiku t pro $t = 1, 2, \dots, N$ budeme značit c_t . Číslo c_t udává pořadové číslo aktivní komponenty. Stejně jako pro obyčejný index komponenty c , platí i pro hodnoty přepínače

$$c_t \in \{1, 2, \dots, n\}, \quad \text{kde } n \text{ je počet komponent multimodelu.}$$

Poznámka: Je třeba dobře rozlišovat mezi indexem komponent c a přepínačem C_t s hodnotami c_t . Index c je obyčejné přirozené číslo z množiny $\{1, 2, \dots, n\}$. Řekneme-li "pro komponentu c platí", říkáme "pro libovolnou komponentu platí". Hodnota přepínače c_t se za prvé vztahuje jen k časovému okamžiku t , za druhé označuje právě tu jedinou komponentu, která je v čase t aktivní. Její hodnoty mohou být ze stejného intervalu jako hodnoty indexu c , ale nemá smysl psát $c_t = 1, 2, \dots, n$. Komponentu c známe, známe-li parametry komponent Θ ; komponentu c_t neznáme, ani při znalosti Θ , pokud nemáme další informaci o tom, která komponenta byla skutečně v čase t aktivní.

Ve výjimečných případech, kdy lze aktivity jednotlivých komponent měřit, je tato veličina známá diskrétní funkce – konečná posloupnost.

Ve většině případů však nevíme, která komponenta je aktivní, a C_t je posloupností náhodných veličin s realizacemi c_t . Pro tyto náhodné veličiny volíme následující popis

$$\mathcal{P} : f(c_t | D(t-1), \alpha) = f(c_t | \alpha_{c_t}) = \alpha_{c_t}, \quad (8.5)$$

kde α_{c_t} je parametr, udávající pravděpodobnost, s níž přepínač nabude hodnotu c_t .

Parametry α_c , $c = 1, 2, \dots, n$, tvoří vektor α s n složkami. Tento vektor můžeme interpretovat přímo jako vektor pravděpodobností toho, že v čase t je aktivní c -tá komponenta multimodelu (za podmínky, že známe pouze starší data $D(t-1)$ a nikoliv současnou hodnotu data d_t). Proto musí platit

$$\alpha_c \in \langle 0, 1 \rangle, \quad \text{pro } c = 1, 2, \dots, n \quad \text{a} \quad \sum_{c=1}^n \alpha_c = 1.$$

V modelu přepínače je vyjádřena podmínka nezávislosti náhodné veličiny C_t na starších datech $D(t-1)$.

Součinný tvar modelu přepínače je

$$\mathcal{P} : f(c_t | \alpha_{c_t}) = \prod_{c=1}^n \alpha_c^{\delta(c, c_t)}, \quad (8.6)$$

kde $\delta(c, c_t)$ je Kroneckerova funkce (8.4).

Popis multimodelu

V případě, kdy neznáme aktivity jednotlivých komponent, je multimodel popsán podmíněnými hp obou náhodných veličin d_t a c_t . Ty lze použitím řetězového pravidla dále rozložit a při respektování předpokladů v (8.1) a (8.5) dostaneme

$$\begin{aligned} \mathcal{M} : f(d_t, c_t | D(t-1), \Theta_{c_t}, \alpha_{c_t}) &= f(d_t | D(t-1), c_t, \Theta_{c_t}) f(c_t | D(t-1), \alpha_{c_t}) = \\ &= f_{c_t}(d_t | \varphi_t, \Theta_{c_t}) \alpha_{c_t}. \end{aligned} \quad (8.7)$$

První hp v rozkladu na pravé straně je model komponenty (8.1) a druhá model přepínače (8.5).

Multimodel, popisující aktivní komponentu v čase t , lze při použití vztahů (8.3), (8.6) také vyjádřit v součinném tvaru

$$\begin{aligned} \mathcal{M} : f(d_t, c_t | D(t-1), \Theta_{c_t}, \alpha_{c_t}) &= \prod_{c=1}^n f(d_t, c | D(t-1), \Theta_c, \alpha_c)^{\delta(c, c_t)} = \\ &= \prod_{c=1}^n \exp \left\{ q_c^T(\Theta) s_c(\phi_t) \delta(c, c_t) \right\} \alpha_c^{\delta(c, c_t)}. \end{aligned} \quad (8.8)$$

Poznámka: Vyjádření modelu v součinném tvaru je formální krok. Přesto má význam. Hp na pravé straně neobsahují neznámou funkci c_t . Obsahují jen indexovou proměnnou c za kterou postupně dosazujeme hodnoty $1, 2, \dots, n$. Funkce c_t je obsažena jen v ukazateli $\delta(c, c_t)$.

Poznámka: Ukazatel $\delta(c, c_t)$ můžeme chápat jako hp veličiny c_t jestliže její hodnotu známe, tj. víme, která komponenta je aktivní.

Popis předpovědi dat

Jako u obyčejných modelů (5.8), i u multimodelů je jejich úkolem předpovídat data d_t bez ohledu na to, která komponenta je zrovna popisuje. Jinými slovy, potřebujeme z popisu multimodelu eliminovat veličinu c_t . To lze provést výpočtem marginální hp pomocí (8.7) podle vzorce

$$\begin{aligned} f(d_t | D(t-1), \Theta, \alpha) &= \sum_{c_t=1}^n f(d_t, c_t | D(t-1), \Theta, \alpha) = \\ &= \sum_{c_t=1}^n f_{c_t}(d_t | \varphi_t, \Theta_{c_t}) \alpha_{c_t}. \end{aligned} \quad (8.9)$$

Vidíme, že při volbě rozdělení pro model přepínače podle (8.5) dostáváme hp předpovědi ve tvaru lineární kombinace komponent, kde vahami kombinace jsou pravděpodobnosti jednotlivých komponent. Takováto lineární kombinace se nazývá **směs hustot pravděpodobnosti** (více o směsích viz např. [8]).

8.2 Odhad parametrů multimodelu**Bayesův vztah**

Dosud jsme se zabývali matematickým popisem multimodelu. Ten je na obecné úrovni dán podmíněnou hp (8.7), závislou na neznámých parametrech Θ, α . Základem pro odhad těchto parametrů je, podobně jako u obyčejných modelů, Bayesův vztah (5.17), který za pomoci měřených

dat d_t přepočítává apriorní hp $f(\Theta, \alpha|D(t-1))$ na posteriorní hp $f(\Theta, \alpha|D(t))$. Bayesův vzorec má v tomto případě tvar

$$\mathcal{B}: f(\Theta, \alpha|D(t)) \propto f(d_t|D(t-1), \Theta, \alpha)f(\Theta, \alpha|D(t-1)), \quad (8.10)$$

který je rekurzivně počítán pro $t = 1, 2, \dots, N$, a který startuje s apriorní hp $f(\Theta, \alpha|D(0))$ vybudovanou na základě apriorní znalosti.

Aproximace Bayesova vztahu

Princip aproximace: Z předchozího vyjádření Bayesova vztahu (8.10) je patrné, že pro odhad je zapotřebí nikoliv multimodel, ale jeho marginální tvar (8.9), který jsme nazvali předpovědí a který lze vyjádřit ve tvaru směsi hp. Po dosazení dostaneme

$$\mathcal{B}: f(\Theta, \alpha|D(t)) \propto \left[\sum_{c=1}^n f(d_t, c|D(t-1), \Theta, \alpha) \right] f(\Theta, \alpha|D(t-1)). \quad (8.11)$$

Protože tento vztah se počítá rekurzivně pro $t = 1, 2, \dots, N$, dostáváme posteriorní hp jako opakované součiny součtů v neuzavřeném tvaru (počet sčítanců neustále narůstá). Tento tvar je pro další počítání zcela nevhodný a je třeba jej aproximovat.

Navrženou aproximaci lze popsat v následujících dvou krocích:

- A. Marginální hp (8.9) nahradíme jedinou ("nejlepší") podmíněnou hp tak, jako kdybychom znali index aktivní komponenty $c = c_t$ a místo marginalizace bychom hodnotu tohoto indexu do podmíněné hp prostě dosadili.
- B. Výběr "nejlepší" hp provedeme tak, že ukazatel aproximujeme jeho optimálním bodovým odhadem. Pro kvadratické kritérium je jím střední hodnota, podmíněná dosud pozorovanými daty (5.11).

Schéma aproximace: Kroky, naznačené v předchozím odstavci, popíšeme podrobněji:

- A. V Bayesově vzorci zaměníme směs hp jednou s indexem c_t , o které prohlásíme, že "to je ta nejlepší", která popisuje aktivní komponentu, a vyjádříme ji v součinném tvaru (8.3)

$$\sum_{c=1}^n f(d_t, c|D(t-1), \Theta, \alpha) \rightarrow f(d_t, c_t|D(t-1), \Theta, \alpha) = \prod_{c=1}^n f(d_t, c|D(t-1), \Theta, \alpha)^{\delta(c, c_t)}.$$

- B. Ukazatel, obsahující ve skutečnosti neznámou funkci c_t , nahradíme bodovým odhadem optimálním ve smyslu středně kvadratické odchylky, kterým je jeho střední hodnota

$$\delta(c, c_t) \rightarrow E[\delta(c, c_t)|D(t)] = \sum_{c=1}^n \delta(c, c_t)f(c_t|D(t)) = f(c_t = c|D(t)), \quad c = 1, 2, \dots, n,$$

podmíněná všemi dosud známými daty, včetně současných. Zde, ale i dále, používáme následující značení: $f(c_t = c|D(t)) = P(c_t = c|D(t))$, což je hp náhodné veličiny c_t v bodě c . $P(\cdot|\cdot)$ označuje podmíněnou pravděpodobnost.

Střední hodnotou ukazatele je podmíněná hp $f(c_t = c|D(t))$. Její hodnoty pro $c = 1, 2, \dots, n$ v čase t označíme $w_{c,t}$ a vypočteme je takto

$$w_{c,t} = f(c_t = c|D(t)) = \int \int f(c_t = c, \Theta, \alpha|d_t, D(t-1))d\Theta d\alpha \propto$$

$$\begin{aligned}
& \propto \int \int f(d_t|D(t-1), c, \Theta, \alpha) f(c_t = c|D(t-1)) f(\Theta, \alpha|D(t-1)) d\Theta d\alpha = \\
& = \int \int f_c(d_t|\varphi_t, \Theta_c) \alpha_c f(\Theta|D(t-1)) f(\alpha|D(t-1)) d\Theta d\alpha \\
& = \int f_c(d_t|\varphi_t, \Theta_c) f(\Theta|D(t-1)) d\Theta \int \alpha_c f(\alpha|D(t-1)) d\alpha \quad (8.12)
\end{aligned}$$

pro $c = 1, 2, \dots, n$.

V předchozím odvození jsme v prvním řádku přidali a ihned vyintegrovali parametry Θ a α ; při přechodu na druhý řádek jsme aplikovali Bayesův vzorec (3.5) a řetězové pravidlo (3.2) a ve třetím řádku již jen použili zavedené podmínky a značení pro jednotlivé hp (8.1) a (8.5). Výsledkem integrace pro jednu hodnotu c je číslo – pravděpodobnost, že $c_t = c$, tj., že v čase t je aktivní právě c -tá komponenta. Integrace provedeme pro všechny hodnoty veličiny $c = 1, 2, \dots, n$ a výsledkem bude vektor pravděpodobností aktivity jednotlivých komponent. Tento vektor "aktuálních pravděpodobností" jednotlivých komponent jsme označili $w_{c,t}$.

Poznámka: Všimněme si, že výsledný tvar výrazu pro výpočet hodnoty $w_{c,t}$ je dán součinem dvou nám již známých výrazů. Jedná se o hp předpovědi pro regresní model s normálním šumem (5.33) a stejnou hp ale pro model s diskretním šumem (5.72).

Výsledek aproximace: V předchozím odstavci jsme použili vyjádření modelu komponenty v součinném tvaru a nahradili ukazatel $\delta(c, c_t)$ jeho aproximací $w_{c,t}$. Podobně, v součinném tvaru (tzv. Dirichletův tvar) vyjádříme také hp parametrů

$$f(\Theta, \alpha|D(t)) = \prod_{c=1}^n f(\Theta_c, \alpha_c|D(t)).$$

Označíme-li společně symbolem $U_{c,t}$ ukazatel $\delta(c, c_t)$ nebo aproximovaný ukazatel $w_{c,t}$, lze Bayesův vztah, bez aproximace i aproximovaný, psát jednotně takto

$$\mathcal{B} : \prod_{c=1}^n f(\Theta_c, \alpha_c|D(t)) \propto \prod_{c=1}^n [f_c(d_t, c|D(t-1), \Theta_c, \alpha_c)^{U_{c,t}} f(\Theta_c, \alpha_c|D(t-1))]. \quad (8.13)$$

Tento vzorec převedeme později na rekuzi pro statistiky. Prozatím poznamenejme jen to, že aproximací jsme dosáhli svého cíle. Zbavili jsme se součtového vyjádření modelu komponenty, který způsoboval neuzavřenost Bayesova vztahu.

Tvar Bayesova vorce (8.13) rovněž ukazuje, že pro vypočtené váhy $w_{c,t}$ je jeho výpočet možno rozdělit a provádět pro každou komponentu $c = 1, 2, \dots, n$ zvlášť. Jen při výpočtu vah se komponenty kombinují.

Bayesův vztah pro exponenciální třídu komponent

Bayesův vztah (8.13), tak jsme jej uvedli v předešlém odstavci, představuje funkcionální rekuzi. Ukazuje přepočtení funkce $f(\Theta_c, \alpha_c|D(t-1))$ na funkci $f(\Theta_c, \alpha_c|D(t))$. Protože takové výpočty nelze na současné výpočetní technice provádět, potřebujeme tuto funkcionální rekuzi převést na číselnou. To uděláme podobně jako v případě obyčejných modelů (viz algoritmy na str. 71 a 74) tak, že vyjádříme modely v konkrétním tvaru – přepínač s rozdělením (8.5) a komponenty obecně v exponenciálním tvaru podle (8.2).

Odpovídající (konjugovaný) tvar hp parametrů, viz postup od modelu k aposteriori hp na str. 66 a dále, je

$$f(\Theta_c, \alpha_c|D(t)) = \exp\{q_c^T(\Theta) S_c(t)\} \alpha_c^{t_c(t)-1}, \quad c = 1, 2, \dots, n,$$

kde q_c je parametrická funkce z exponenciálního vyjádření modelu a $S_c(t)$ je statistika pro odhadování parametrů modelu c -té komponenty a $\iota_c(t)$ je statistika pro odhad parametrů přepínače.

Po dosažení exponenciálního tvaru modelu komponent a jemu konjugovaného tvaru hp parametrů a po vypuštění symbolu pro součin má Bayesův vzorec (8.13) následující tvar (pro jednotlivé komponenty)

$$\begin{aligned} \mathcal{B} : \exp\{q_c^T(\Theta)S_c(t)\alpha^{\iota_c(t)} &\propto \left[\exp\{q_c^T(\Theta)s_c(\phi_t)\}\alpha_c\right]^{U_{c,t}} \exp\{q_c^T(\Theta)S_c(t-1)\}\alpha^{\iota_c(t-1)} = \\ &= \exp\{q_c^T(\Theta)[s_c(\phi_t)U_{c,t} + S_c(t-1)]\}\alpha_c^{U_{c,t} + \iota_c(t-1)}, \quad c = 1, 2, \dots, n \end{aligned} \quad (8.14)$$

Porovnáním odpovídajících si statistik na obou stranách Bayesova vzorce dostáváme následující rekurze, srv. (5.25)

$$S_c(t) = S_c(t-1) + s_c(\phi_t)U_{c,t} \quad (8.15)$$

$$\iota_c(t) = \iota_c(t-1) + U_{c,t} \quad (8.16)$$

pro $c = 1, 2, \dots, n$.

Vlastní algoritmus pro odhadování parametrů Θ a α multimodelu můžeme shrnout takto:

Příprava odhadu:

sestav apriorní statistiky S_0, ι_0 .

Vlastní odhad:

for $t = 1 : N$

změř nová data,

sestav rozšířený regresní vektor, viz (8.2),

přepočti statistiky podle (8.15) a (8.16).

end %for

Závěr odhadu:

z napočtených statistik sestav aposteriorní hp,

případně vypočti bodové odhady parametrů.

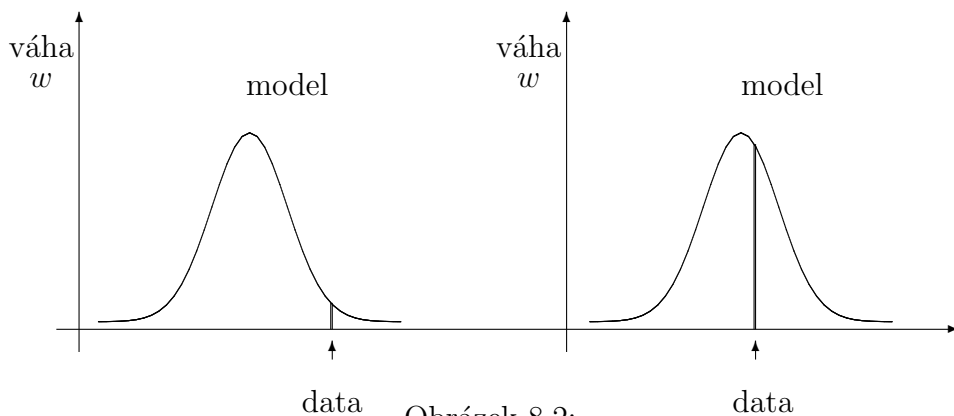
Struktura prostřední části uvedeného algoritmu počítaného v cyklu času je následující:

- změř data,
- udělej "pokusný" přepočet statistik (8.15) a (8.16) pro $U_{c,t} = 1, \quad c = 1, 2, \dots, n$
- vypočti aktuální váhy komponent $w_{c,t}$ (8.12) s právě přepočtenými statistikami,
- udělej "skutečný" přepočet statistik (8.15) a (8.16) pro $U_{c,t} = w_{c,t}, \quad c = 1, 2, \dots, n$.

Komentář k algoritmu:

První přepočet statistik je třeba provést pro výpočet aktuálních vah (pravděpodobností aktivit) jednotlivých komponent. Při něm se stavíme do situace **A.**, kdy předpokládáme znalost aktivní komponenty. Přitom každou komponentu $c = 1, 2, \dots, n$ pokládáme vždy za známou a zkusíme, s jakou je pravděpodobností naměřená data odpovídají této komponentě. To jsou

právě hledané váhy $w_{c,t}$. Korespondenci dat a komponenty je možno dobře ilustrovat na statickém modelu, kde největší pravděpodobnost dat je přímo pod vrcholem hp.



Obrázek 8.2:

V levé části obrázku 8.2 leží data daleko od středu, pořadnice pod hp (odrážející pravděpodobnost, že data pochází z tohoto modelu) je malá. Na pravém obrázku leží data blíže středu, pořadnice, a tedy i pravděpodobnost příslušnosti dat k modelu, je větší.

Pravděpodobnosti, které určíme po prvním přepočtu statistik, použijeme jako váhy $w_{c,t}$, se kterými přidáváme informaci ze současných dat do statistik jednotlivých komponent. Čím více data komponentě odpovídají, s tím větší vahou jsou zabudována do jejích statistik.

Bezprostřední důsledek aproximace je následující: jestliže víme přesně, která z komponent je aktivní, přepočteme její statistiky s jednotkovou vahou a ostatní ponecháme nedotčené. Pokud si ale nejsme aktivní komponentou jisti (což je důvodem k aproximaci), přidělujeme data do všech komponent, do každé s takovou vahou, která ji vzhledem k naměřeným datům náleží. Původní ukazatel na aktivní komponentu se tak "rozštěpí" a změní se na hustotu pravděpodobnosti pro jednotlivé komponenty. Tato situace je naznačena na obrázku 8.3

8.3 Příklad pro statický multimodel

Jako příklad ilustrující teorii uvedeme odhad multimodelu s dvěma statickými komponentami a normálním šumem. Obyčejným statickým modelem jsme se zabývali v odstavci 4.1.2 a 5.4.2. Nebudeme proto již opakovat odvození odhadu pro jednotlivé komponenty našeho multimodelu, ale využijeme již odvozených výsledků.

Modely komponent

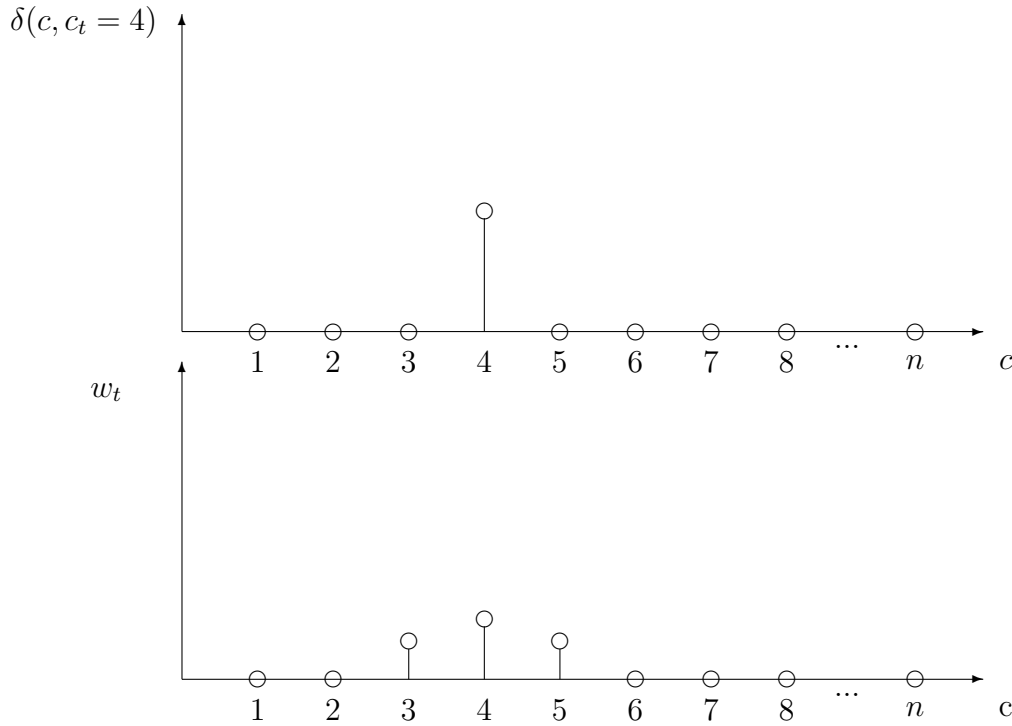
Uvažujeme dvě statické komponenty podle (4.15)

$$\mathcal{K}_c : f(d_t | \Theta_c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp \left\{ -\frac{1}{2\sigma_c^2} (d_t^2 - 2\theta_c d_t + \theta_c^2) \right\}, \quad c = 1, 2.$$

s parametry θ_c a σ_c^2 označené indexem komponent c a s vektorem parametrů

$$\Theta^T = [\Theta_1^T, \Theta_2^T], \quad \Theta_1^T = [\theta_1, \sigma_1^2], \quad \Theta_2^T = [\theta_2, \sigma_2^2],$$

společným regresním vektorem $\varphi_t = 1$ a rozšířeným regresním vektorem $\phi_t^T = [d_t, 1]$.



Obrázek 8.3:

Komponenty tohoto modelu vyjádříme ve tvaru exponenciální třídy (8.2)

$$\begin{aligned}
 f(d_t | \Theta_c) &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{d_t^2}{2\sigma_c^2} + \frac{d_t \theta_c}{\sigma_c^2} - \frac{\theta_c^2}{2\sigma_c^2} - \log \sigma \right\} = \\
 &= \frac{1}{\sqrt{2\pi}} \exp \left\{ \left[\frac{-1}{2\sigma_c^2}, \frac{\theta_c}{\sigma_c^2}, \frac{-\theta_c^2}{2\sigma_c^2} - \log \sigma \right] [d_t^2, d_t, 1]^T \right\} = \\
 &= \frac{1}{\sqrt{2\pi}} \exp \left\{ q_c^T(\Theta_c) s(\phi_t) \right\},
 \end{aligned}$$

kde

$$q_c^T = [q_{1,c}, q_{2,c}, q_{3,c}], \quad s_c^T = [s_{1,c}, s_{2,c}, s_{3,c}], \quad c = 1, 2$$

a jednotlivé prvky těchto vektorů jsou dány předchozím vztahem.

Protože aposteriorní hp je, až na normování, dána součinem modelů, a protože apriorní hp musí být strukturálně stejná jako aposteriorní, lze podle rovnice modelu, který je kvadratickou funkcí parametrů navrhnout i strukturu aposteriorní (a podobně i apriorní) hp. Ta bude mít tvar

$$f(\Theta_c | D(t)) \propto \exp \left\{ -\frac{1}{2} \left(\kappa_c(t) \theta_c^2 - 2S_{d1,c}(t) \theta_c + S_{d,c}^2(t) \right) \right\},$$

kde κ_c , $S_{d1,c}$ a $S_{d,c}^2$ jsou statistiky pro odhad parametrů komponent – viz (5.39) a (5.40).

Model přepínače

Hp přepínače má vždy rozdělení (8.5), tj.

$$\mathcal{P} : f(c_t | \alpha) = \alpha_{c_t}, \quad c_t \in \{1, 2\},$$

kde

$$\alpha^T = [\alpha_1, \alpha_2], \quad \alpha_1, \alpha_2 \geq 0, \quad \alpha_1 + \alpha_2 = 1.$$

Aposteriorní (jako i apriorní) hp má stejně jako model přepínače pevnou strukturu

$$f(\alpha_c | D(t)) \propto \alpha_c^{\iota_c(t)-1}.$$

Přepočet statistik

Podle Bayesova vzorce (8.14) pro exponenciální třídu komponent, rozepsaného pro jednotlivé komponenty dostáváme přepočet statistik podle (8.15) a (8.16) ve tvaru

$$\begin{aligned} S_{d,c}^2(t) &= S_{d,c}^2(t-1) + d_t^2 U_{c,t}, \\ S_{d1,c}(t) &= S_{d1,c}(t-1) + d_t U_{c,t}, \\ \kappa_c(t) &= \kappa_c(t-1) + U_{c,t}, \\ \iota_c(t) &= \iota_c(t-1) + U_{c,t}, \end{aligned}$$

pro $c = 1, 2$ a $t = 1, 2, \dots, N$,

kde

$U_{c,t}$ je rovno 1 pro první přepočet statistik a rovná se $w_{c,t}$ pro druhý přepočet statistik,

$S_{d,c}^2(0)$, $S_{d1,c}(0)$, $\iota_c(0)$ a $\kappa_c(0)$ jsou apriorní statistiky,

N je počet kroků odhadování.

Výpočet vah $w_{c,t}$ se děje obecně podle vztahu (8.12), kde za první hp předpovědi spojitého modelu dosadíme z (5.54).

8.4 Programový systém MixTools

Na str. 113 jsme uvedli algoritmus pro odhadování multimodelu a v příkladu 8.3 jsme vzorce používané v tomto algoritmu specifikovali pro statický regresní model s normálními komponentami. Pro obecný případ dynamického modelu, případně s jiným než normálním rozdělením šumu, je situace mnohem složitější, a nebudeme ji zde podrobně popisovat. Místo toho zde stručně popíšeme základní prvky programového systému MixTools, vyvíjeného v ÚTIA AV ČR, který slouží, mimo jiné, k odhadování obecně mnohorozměrných dynamických multimodelů [11]. Učebnicovou verzi programového systému MixTools je možno zdarma získat na webové adrese (pod názvem)

http://www.utia.cas.cz/AS_dept/ACTIVITY/RECIAS/allexam.htm

(Basic tool for mixture estimation and prediction)

Pomocí tohoto toolboxu lze snadno definovat multimodel, na základě změřených dat určit počet komponent i jejich dynamickou strukturu, vytvořit apriorní statistiky a model odhadovat. Základními prvky, se kterými pracuje, jsou: **faktor**, **komponenta** a **směs**. Vztah směsi a komponent je zřejmý z teoretického pojednání. Hp popisující multimodel je lineární kombinace hp komponent s vahami kombinace rovnými pravděpodobnostem jednotlivých komponent. Souvislost komponenty a jejich faktorů je nejlépe patrná opět na hp komponenty. Ukážeme ji pro dvourozměrný případ.

Komponenta a její faktory

Budeme uvažovat dvourozměrnou komponentu 1. řádu s konstantou. Ta popisuje dvourozměrnou náhodnou veličinu $y_t = [y_{1;t}, y_{2;t}]^T$ pomocí podmíněné hp s vektorem parametrů θ . Tuto dvourozměrnou hp lze rozepsat podle řetězového pravidla (3.2) – index komponenty c pro přehlednost vynecháme

$$f(y_{1;t}, y_{2;t} | y_{1;t-1}, y_{2;t-1}, \theta) = f(y_{2;t} | y_{1;t}, y_{1;t-1}, y_{2;t-1}, \theta) f(y_{1;t} | y_{1;t-1}, y_{2;t-1}, \theta).$$

Jednorozměrné hp v součinu jsou faktory. Odtud je zřejmé, že model komponenty ve tvaru hustoty pravděpodobnosti lze vyjádřit jako součin jednorozměrných hustot pravděpodobností modelů jednotlivých faktorů.

Strukturu faktorů a přepočtení parametrů komponenty na parametry faktorů ukážeme na vyjádření modelu komponenty ve tvaru rovnice

$$\begin{bmatrix} y_{1;t} \\ y_{2;t} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & k_1 \\ a_{2,1} & a_{2,2} & k_2 \end{bmatrix} \begin{bmatrix} y_{1;t-1} \\ y_{2;t-1} \\ 1 \end{bmatrix} + \begin{bmatrix} r_{1,1} & r_{1,2} \\ 0 & r_{2,2} \end{bmatrix} \begin{bmatrix} e_{1;t} \\ e_{2;t} \end{bmatrix},$$

kde náhodnou část modelu jsme vyjádřili ve tvaru "směrodatná odchylka" (matice r) krát "normovaný šum" (vektor e). Zde směrodatná odchylka je Choleskyho odmocnina z kovarianční matice šumu a normovaný šum má rozptyl roven jedné.

Model komponenty dále upravíme tak, abychom dostali dva modely s oddělenými, tj. vzájemně nekorelovanými šумы. Tyto modely budeme nazývat faktory.

Z matice r vytkneme diagonální prvky a zbytek invertujeme

$$\begin{bmatrix} r_{1,1} & r_{1,2} \\ 0 & r_{2,2} \end{bmatrix} = \begin{bmatrix} r_{1,1} & 0 \\ 0 & r_{2,2} \end{bmatrix} \begin{bmatrix} 1 & r_{1,2}/r_{1,1} \\ 0 & 1 \end{bmatrix} \quad (\text{vytknutí}),$$

$$\begin{bmatrix} 1 & r_{1,2}/r_{1,1} \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & -r_{1,2}/r_{1,1} \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & g_{1,2} \\ 0 & 1 \end{bmatrix} \quad (\text{inverze}).$$

Inverzí g vynásobíme model komponenty a dostaneme

$$\begin{bmatrix} 1 & g_{1,2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} y_{1;t} \\ y_{2;t} \end{bmatrix} = \begin{bmatrix} 1 & g_{1,2} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{1,1} & a_{1,2} & k_1 \\ a_{2,1} & a_{2,2} & k_2 \end{bmatrix} \begin{bmatrix} y_{1;t-1} \\ y_{2;t-1} \\ 1 \end{bmatrix} + \begin{bmatrix} r_{1,1} & 0 \\ 0 & r_{2,2} \end{bmatrix} \begin{bmatrix} e_{1;t} \\ e_{2;t} \end{bmatrix}.$$

Po roznásobení matic a převedení členu $g_{1,2}y_{2;t}$ na pravou stranu rovnice lze model upravit do následujícího tvaru

$$\begin{bmatrix} y_{1;t} \\ y_{2;t} \end{bmatrix} = \begin{bmatrix} q_{1,0} & q_{1,1} & q_{1,2} & q_{1,3} \\ 0 & q_{2,1} & q_{2,2} & q_{2,3} \end{bmatrix} \cdot \begin{bmatrix} y_{2;t} \\ y_{1;t-1} \\ y_{2;t-1} \\ 1 \end{bmatrix} + \begin{bmatrix} r_{1,1}e_{1;t} \\ r_{2,2}e_{2;t} \end{bmatrix},$$

kde koeficienty q dostaneme z úprav, které jsme prováděli.

Výsledný model lze rozepsat na dva modely, každý se svým šumem. To jsou faktory výchozí dvourozměrné komponenty.

$$\begin{aligned} y_{1;t} &= q_{1,0}y_{2;t} + q_{1,1}y_{1;t-1} + q_{1,2}y_{2;t-1} + q_{1,3} + r_{1,1}e_{1;t} \\ y_{2;t} &= q_{2,1}y_{1;t-1} + q_{2,2}y_{2;t-1} + q_{2,3} + r_{2,2}e_{2;t}. \end{aligned}$$

Každý faktor modeluje jednu složku výstupu y , tzv. datový kanál.

Základní prvky programu

Globální proměnné

V MixTools jsou definovány 3 globální proměnné, které, po jejich deklaraci, lze kdekoliv použít. Jsou to proměnné

TIME – *dynamický čas* běhu programu (v teorii značený t),

DATA – *datová matice*, kde data jsou uspořádána po řádcích, tj. např. pro dvourozměrný výstup a TIME=1,2,3 je

$$\text{DATA} = \begin{bmatrix} y_{1;1} & y_{1;2} & y_{1;3} \\ y_{2;1} & y_{2;2} & y_{2;3} \end{bmatrix},$$

DEBUG – *úroveň dialogu* (DEBUG=1,2,3; vyšší číslo dá více informací při běhu programu).

Důležité proměnné

Dále jsou zavedena jména řady proměnných. Nejdůležitější jsou

nchn a ndat – *počet kanálů* (dimenze modelované veličiny n_y) a *počet datových vzorků* N .

Platí: `[nchn ndat]=size(DATA)`.

ychns – *vektor modelovaných kanálů*. Např. ychns=[1 2].

str – *struktura faktoru*, která určuje veličiny a jejich pořadí v regresním vektoru faktoru. Je to dvouřádková matice, kde v prvním řádku je uvedeno číslo modelovaného kanálu a na odpovídajícím místě druhého řádku je jeho zpoždění. Např. pro dvě modelované veličiny a faktor

$$y_{1;t} = a_{1,1}y_{1;t-1} + a_{1,2}y_{2;t-1} + a_{1,3}y_{1;t-2} + a_{1,4}y_{2;t-2} + k + e_{1;t}$$

je odpovídající struktura

$$\text{str} = \begin{bmatrix} 1 & 2 & 1 & 2 & 0 \\ 1 & 1 & 2 & 2 & 1 \end{bmatrix}$$

tj., regresní vektor bude obsahovat první kanál jednou zpožděný, druhý kanál jednou zpožděný, první a druhý kanál dvakrát zpožděný a jedničku pro modelování konstanty.

dfcs – *počet stupňů volnosti* komponent. Po normování na jednotkový součet dá pravděpodobnosti komponent.

ncom – *počet komponent* multimodelu.

Coms – *cell vektor ARX LS komponent*. Jeho složky jsou Coms $\{i\}$, $i=1,2,\dots,ncom$ a obsahují informace o komponentách.

Komponenta Coms $\{.\}$ obsahuje následující proměnné: (pro dvourozměrný model s třemi parametry)

proměnná	název	typ
ychns	<i>modelované kanály</i>	vektor 1x nchn
str	<i>struktura</i>	matice 2x3
dfm	<i>počet stupňů volnosti parametrů</i>	číslo
type	<i>typ komponenty</i>	číslo
cove	<i>kovarianční matice šumu</i>	matice 2x2
Eth	<i>regresní koeficienty</i>	matice 2x3
Cth	<i>kovariance regresních koeficientů</i>	matice 3x3

Sim – směs pro simulaci – typ *ARX LS*. Struktura je následující

proměnná	název	typ
Facs	<i>faktory</i>	cell vektor
coms	<i>faktory v komponentách</i>	matice (řád. komp., sl. faktory)
dfcs	<i>vektor stupňů volnosti</i>	vektor $1 \times n_{com}$
type	<i>typ směsi</i>	číslo
states	<i>stavy směsi</i>	matlabská struktura

Struktura jednoho faktoru **Facs**{.} je

proměnná	název	typ
ychn	<i>číslo modelovaného kanálu</i>	číslo
str	<i>struktura faktoru</i>	matice
dfm	<i>stupně volnosti</i>	číslo
type	<i>typ faktoru</i>	číslo
cove	<i>rozptyl šumu</i>	číslo
Eth	<i>regresní koeficienty</i>	vektor
Cth	<i>kovariance regresních, koeficientů</i>	matice

Mix – směs pro odhadování – typ *ARX*. Její struktura je téměř stejná jako u **Sim**, místo parametrů **cove**, **Eth** a **Cth** však obsahuje odmocniny statistik z rozšířené informační matice.

Struktura jednoho faktoru **Facs**{.} je

proměnná	název	typ
ychn	<i>číslo modelovaného kanálu</i>	číslo
str	<i>struktura faktoru</i>	matice
dfm	<i>stupně volnosti</i>	číslo
type	<i>typ faktoru</i>	číslo
LD	<i>odmocnina z informační matice</i>	matice

Poznámka: *Symboly značí:*

ARX - autoregresní model s externím signálem – příslušná struktura obsahuje matici **LD** (rozklad informační matice);

ARX LS totéž, ale v "least square podobě", tj. struktura obsahuje prvky **Eth**, **Cth**, **cove** - což jsou parametry vypočtené z informační matice.

Důležité procedury

Podrobnější popis lze získat v Matlabu zadáním: **help** a jméno procedury, výpis celé procedury (tj. komentáře i kódu) příkazem: **type** a jméno procedury.

Coms{i} = **comarxls**(**ychns**, **str**)
konstrukce *i* té komponenty.

Mix = **mixconst**(**Coms**, **dfcs**)
konstrukce směsi (simulátoru nebo estimátoru).

Mix = **genmixs**(**ychns**, **str**, **dfcs**, **Pt**)
konstrukce směsi pro simulaci (*i* se zadáním parametrů **Pt**).

```

Mix = genmixe(ncom,ychns,str,ndat,diagCth,diagcove,dfm,dfcs)
      konstrukce směsi pro odhad (i se zadáním apriorních statistik).

Mix1 = mix2mix(Mix, type)
      konverze směsi na jiný typ
      (důležité typy: 21=ARX, 22=ARX LS, 23=matr ARX, 24=matr ARX LS).

mixsimul(Sim, ndat)
      simulace směsi.

Mix = mixest(Mix, frg, ndat, niter, opt)
      odhad směsi.

Mix = mixstats(Mix, ndat)
      přepočítání statistik směsi (s finálními parametry).

```

Další programy slouží jako návod, jak stavět pomocí prostředků MixTools základní úlohy simulace a odhadování.

Jednoduchý příklad - statická verze

V následujících třech programech jsou demonstrovány tři nezákladnější úlohy z oblasti odhadu. Je to simulace (příprava dat), odhad (nalezení modelu odpovídajícího získaným datům) a předpověď (využití odhadnutého modelu pro předpovídání dat). Ve všech úlohách se předpokládá nejjednodušší statický model.

Simulace tří dvourozměrných statických komponent

```

prodini, echo off                                % zahájení programu
%
% Komponenta
% |y1(t)| = |Eth1| + |cove1 cove12|. |e1(t)|
% |y2(t)|   |Eth2|   |0   cove2 | |e2(t)|
%
ndat=1000;                                       % počet kroků simulace
ychns = [1 2];                                  % modelované kanály
str    = [0;1];                                  % struktura pro statické faktory
dfcs   = [10 15 30];                             % váhy komponent

% parametry simulované soustavy uložené ve struktuře Pt
sd = 1;
Pt{1}.Eth = [0; 0];
Pt{2}.Eth = [5; 0];
Pt{3}.Eth = [3; 8];
Pt{1}.cove = [.1 1; 1 .1]*sd;
Pt{2}.cove = [1 0; 0 1]*sd;
Pt{3}.cove = [.1 -1; -1 .1]*sd;

Sim = genmixs(ychns, str, dfcs, Pt); % vytvoření simulátoru
DATA = zeros(2,ndat);               % deklarace datového vzorku

```



```

% SIMULACE
mixsimul(Sim, ndat);           % simulace
save simDat Sim DATA        % uložení dat na disk

% ZOBRAZENÍ
plot(DATA(1,:),DATA(2,:),'.','MarkerSize',1)
title('Simulace dvourozměrné statické soustavy -- shluky dat [y1(t),y2(t)]')
axis([-5,10,-5,15])

```

Poznámka: *Bodové odhady regresních koeficientů srovnajte s hodnotami parametrů v simulaci.*

Doporučené experimenty

1. Změna regresních koeficientů `.Eth` znamená změnu polohy shluku. Generujte shluky se středy na diagonále obrázku.
2. Změna kovarianční matice šumu `.cove` znamená: zvětšení (zmenšené) `sd` zvětšuje (zmenšuje) šířku shluku – volte např. `sd=5.`; změna matice znamená změnu tvaru shluku – porovnejte kovarianční matice a tvary jednotlivých komponent, volte podobné změny.

Odhad tří dvourozměrných statických komponent

```

prodini, echo off           % zahájení programu
load simDat;                % natažení dat z předchozí simulace

% PŘÍPRAVA ODHADU
ncom = 3;                   % počet komponent v odhadované směsi
ychns = [1 2];              % modelované kanály
str = [0;1];                % struktura faktorů (statické)
ndat = 1000;                % počet kroků pro odhadování

% počáteční statistiky
dCt = 1e5; dco = 1e-3; dfm = 10;
dfs = .1*ndat*ones(1,ncom);

% vytvoření směsi pro odhad
Mix0 = genmixe(ncom, ychns, str, ndat, dCt, dco, dfm, dfs);

% ODHAD
frg = .99;                  % exponenciální zapomínání
Mix = mixest(Mix0, frg, ndat); % odhadování směsi
save simMix                 % uložení odhadnuté směsi

% VÝSLEDKY
M24 = mix2mix(Mix, 24);     % odhadnutá směs (v maticové formě)
disp('Bodové odhady regresních koeficientů')
for i=1:ncom
    Eth2 = M24.Coms{i}.Eth   % střední hodnoty komponent
end

```



```

disp('Koeficienty chyby predikce')
SE = sqrt(var(ep')+mean(ep').^2)./std(dd')    % koeficient chyby předpovědi
disp('Odhadnuté váhy komponent')
al                                           % průměrné váhy komponent

figure(1)
chn = 2;
plot(time, dd(chn,:), 'ob', 'MarkerSize', 3)
hold on
plot(time, yp(chn,:), 'xr', 'MarkerSize', 3)
hold off
title('Časové průběhy dat (o) a jejich předpovědi (x)')

if length(pchms)>1
    figure(2)
    subplot(121), plot(dd(1,:), dd(2,:), '.', 'MarkerSize', 5)
    axis([-5 10 -5 10]), title('Shluky dat');
    subplot(122), plot(yp(1,:), yp(2,:), '.', 'MarkerSize', 5)
    axis([-5 10 -5 10]), title('Předpovědi shluků dat');
end

```

Poznámka: Příkazem load simMix se natáhne z disku odhadnutá směs, která tam byla uložena při odhadu. Spolu s ní se ale natáhnou i další proměnné z programu pro odhad, které byly v okamžiku ukládání směsi definovány. Proto není třeba např. definovat proměnnou ndat, pokud chceme předpovídat všechna data. Samozřejmě, lze tuto proměnnou zavést a definovat jiný počet dat pro předpověď, ten ale nesmí překročit počet dat, existujících ze simulace.

Doporučené experimenty

1. Nejzajímavější na předpovědi je sledovat, jak daleko dokážeme slušně předpovídat. Počet kroků předpovědi lze měnit parametrem `nstep`, který je nastaven na 5 kroků. Zvětšujte počet kroků předpovědi a sledujte kvalitu předpovědi – koeficient SE.

Pozn.: V případě statického modelu, kdy předpovídáme konstantní střední hodnotu modelu, není zhoršení předpovědi příliš významné. Tato závislost je zajímavější v případě dynamických modelů, kdy předpovídaná hodnota závisí na hodnotách minulých dat.

2. V časové smyčce pro výpočet předpovědí je možno volit buď okamžité (`w`) nebo průměrné (`al`) váhy komponent. Porovnejte vliv těchto vah na výsledek předpovídání.

Pozn.: Zajímavější opět pro dynamický model.

Jednoduchý příklad - dynamická verze

V dalších třech programech jsou řešeny stejné úlohy jako v předchozích, ale pro dynamický model. Volbu řádu modelu je možno provést volbou struktury faktorů `str`.

Simulace tří dvourozměrných dynamických komponent

```

prodini, echo off                               % zahájení programu

```

```

ndat = 1000; % počet kroků simulace
ychns = [1 2]; % modelované kanály
str = [1 2 0;1 1 1]; % struktura pro 1. řád modelu
dfcs = [10 15 30]; % váhy komponent pro simulaci

% parametry simulované soustavy
sd = .1;
Pt{1}.Eth = [.9 0 0; 0 .5 0]; % regresní koeficienty
Pt{2}.Eth = [.7 0 5; 0 .8 0]; % (musí odpovídat struktuře "str")
Pt{3}.Eth = [.6 0 3; 0 .1 8];
Pt{1}.cove = sd; % kovariance šumu
Pt{2}.cove = sd; % (je-li zadán skalár, "genmixs"
Pt{3}.cove = sd; % si sám určí správné dimenze)

Sim = genmixs(ychns, str, dfcs, Pt); % vytvoření simulátoru
DATA = zeros(2,ndat); % deklarace datové matice

% SIMULACE
mixsimul(Sim, ndat); % simulace
save simDat3 Sim DATA % uložení dat na disk

% ZOBRAZENÍ
plot(DATA(1,:),DATA(2,:),'.','MarkerSize',1)
title('Shluky simulovaných dat')

```

Doporučené experimenty

1. Stejně experimenty jako pro statický model.
2. Změna řádu simulované soustavy prostřednictvím proměnné `stt` (pozor! se změnou struktury je třeba změnit odpovídajícím způsobem i parametry `.Eth`).

Odhad tří dvourozměrných dynamických komponent

```

prodini, echo off % zahájení programu
load simDat3; % natažení dat ze simulace

% PŘÍPRAVA ODHADU
ndat = 1000; % počet kroků pro odhadování
ncom = 3; % počet komponent odhadovaného modelu
ychns = [1 2]; % modelované kanály
str = [0 1 2;1 1 1]; % struktura pro faktory 1. řádu

% počáteční statistiky
dCt = 1e5; % kovariance parametrů
dco = 1e-3; % kovariance šumu
dfm = 10; % počítadlo komponent
dfs = .1*ndat*ones(1,ncom); % stupně volnosti komponent

% vytvoření modelu pro odhad

```

```

Mix0 = genmixe(ncom, ychns, str, ndat, dCt, dco, dfm, dfs);

% ODHAD
frg = .99; % exponenciální zapominání
Mix = mixest(Mix0, frg, ndat); % odhad směsi
save simMix3 % uložení odhadnuté směsi

% VÝSLEDKY
M24 = mix2mix(Mix, 24); % odhadnutá směs v maticové formě
disp('Bodové odhady regresních koeficientů')
for i=1:ncom
    Eth2 = M24.Coms{i}.Eth % střední hodnoty komponent
end

Data = DATA; % zapamatování simulovaných dat
DATA = zeros(size(DATA));
mixsimul(Mix, ndat); % generování dat z odhadnuté směsi

% Porovnání simulace a odhadování přes datové shluky
figure(1)
subplot(121),plot(Data(1,:),Data(2:,:),'bx')
title('Simulované datové shluky')
subplot(122),plot(DATA(1,:),DATA(2:,:),'rx')
title('Předpovídané datové shluky')

```

Doporučené experimenty

1. Stejné experimenty jako pro statický model. Zejména změna počátečních odhadů parametrů dCt a dco.
2. Odlišný řád modelu pro odhad a simulované soustavy. Odhad provádí aproximaci soustavy. Odhady parametrů musí vyjít jiné než parametry simulace. Kontrola se provede vyhodnocením předpovědi.

Předpověď tří dvourozměrných dynamických komponent

```

prodini, echo off % zahájení programu
load simMix; % natažení odhadnuté směsi

% PŘEDPOVĚĎ
nstep = 5; % počet kroků pro předpověď
pchns = [1 2]; % předpovídané kanály
cchns = []; % kanály v podmínce
pMix = mix2pro(Mix, pchns, cchns); % podmíněná směs
yp = []; time = []; ws = [];

% časová smyčka pro předpovědi
for TIME = Mix.states.maxtd+1:ndat-nstep
    [EthS, coves, al, w] = profixn(pMix, [], [], nstep); % předpovědi komponent
    th = GetTh(EthS);

```

```

    if 1, ypt = th*w'; % předpověď (okamžité váhy komponent)
    else ypt = th*al'; % předpověď (průměrné váhy komponent)
    end
    yp=[yp ypt]; time=[time TIME]; ws=[ws w']; % uchování spočtených hodnot
end

% VÝSLEDKY
dd = DATA(pchns,time+1); % předpovídaná data
yp = yp(pchns,:); % předpovědi
ep = dd-yp; % chyby předpovědi
SE = sqrt(var(ep')+mean(ep').^2)./std(dd') % koeficient chyby předpovědi
al % průměrné váhy komponent

figure(1)
chn = 2;
plot(time, dd(chn,:), 'ob', 'MarkerSize', 3)
hold on
plot(time, yp(chn,:), 'xr', 'MarkerSize', 3)
hold off
title('Časové průběhy dat a jejich předpovědi')

if length(pchns)>1
    figure(2)
    subplot(121), plot(dd(1,:), dd(2,:), '.', 'MarkerSize', 5)
    axis([-5 10 -5 10]), title('Shluky dat');
    subplot(122), plot(yp(1,:), yp(2,:), '.', 'MarkerSize', 5)
    axis([-5 10 -5 10]), title('Předpovědi shluků dat');
end

```

Doporučené experimenty

1. Totéž jako pro statický model, zejména vyhodnocení odhadu s různými apriorními parametry.
2. Vyhodnocení odhadu pro nestejnou strukturu simulace a odhadu.

Obecný příklad

Zde jsou uvedeny opět základní úlohy simulace, odhadování a předpovídání, ale pro obecný model směsi komponent. Zdává se počet komponent (*ncom*), dimenze modelovaných signálů (*nchn*) i řád komponent (prostřednictvím proměnné *str*). Pozn.: Strukturu *str* lze také generovat automaticky pomocí funkce *str = genstr(ord, nchn)*, kde *ord* je řád modelu a *nchn* počet modelovaných kanálů.

Simulace 'ncom', 'nchn'-rozměrných dynamických komponent

```

prodini, echo off % zahájení programu
ndat = 1000; % počet kroků simulace

```

```

ncom = 5; % počet komponent pro simulaci
nchn = 3; % počet simulovaných kanálů
ychns = 1:nchn; % simulované kanály
str = [1:nchn 0; 1*ones(1,nchn) 1]; % struktura prvního řádu soustavy
dfcs = 1+10*rand(1,ncom); % váhy komponent v simulaci

% Parametry simulované soustavy
sd = .1;
for i=1:ncom, Pt{i}.cove=sd; end
Pt{1}.Eth = [.9 0 0 0; 0 .5 0 0; 0 0 .6 0];
Pt{2}.Eth = [.7 0 0 5; 0 .8 0 2; 0 0 .4 1];
Pt{3}.Eth = [.6 0 0 3; 0 .1 0 8; 0 0 .7 6];
Pt{4}.Eth = [.3 0 0 2; 0 .9 0 6; 0 0 .7 3];
Pt{5}.Eth = [.4 0 0 8; 0 .7 0 2; 0 0 .5 1];

Sim = genmixs(ychns, str, dfcs, Pt); % vytvoření směsi pro simulaci
DATA = zeros(nchn,ndat);

% SIMULACE
mixsimul(Sim, ndat); % simulace dat
save simDat4 Sim DATA % uložení dat na disk

% ZOBRAZENÍ
ch1=1; ch2=2; % zobrazované kanály
figure(1)
plot(DATA(ch1,:),DATA(ch2,:),'.','MarkerSize',1)
title('Simulované shluky dat (pro vybrané kanály)')

```

Odhad 'ncom', 'nchn'-rozměrných dynamických komponent

```

prodini, echo off % zahájení programu
load simDat4; % natažení dat ze simulace

% PŘÍPRAVA ODHADU
ncom = 5; % počet komponent pro odhad
nchn = size(DATA,1); % počet modelovaných kanálů
ychns = 1:nchn; % modelované kanály pro odhad
str = [1:nchn 0; 1*ones(1,nchn) 1]; % struktura faktorů prvního řádu
ndat = size(DATA,2); % počet kroků odhadování

% počáteční statistiky pro odhad
dCt = 1e5; dco = 1e-3; dfm = 10;
dfs = .1*ndat*ones(1,ncom);

Mix0 = genmixe(ncom, ychns, str, ndat, dCt, dco, dfm, dfs);
% vytvoření estimátoru

% ODHAD
frg = .99; % exponenciální zapomínání
Mix = mixest(Mix0, frg, ndat); % odhad směsi
save simMix4 % uložení odhadnuté směsi

```

```

% VÝSLEDKY
M24 = mix2mix(Mix, 24); % odhadnutá směs v maticovém tvaru
disp('Odhadnuté regresní koeficienty')
for i=1:ncom, Eth1=M24.Coms{i}.Eth, end % odhady regresních koeficientů
Data = DATA; % pamatování simulovaných dat
DATA = zeros(size(DATA));
mixsimul(Mix, ndat); % simulace z odhadnuté směsi

% Porovnání simulovaných dat a jejich odhadů přes datové shluky
ch1=2; ch2=3;
figure(1)
d1 = [Data(ch1,:) DATA(ch1,:)];
d2 = [Data(ch2,:) DATA(ch2,:)];
mi1= min(d1); ma1=max(d1); mi2=min(d2); ma2=max(d2);
subplot(121),plot(Data(ch1,:),Data(ch2:),'bx')
axis([mi1 ma1 mi2 ma2])
title('Datové shluky ze simulované soustavy')
subplot(122),plot(DATA(ch1,:),DATA(ch2:),'rx')
axis([mi1 ma1 mi2 ma2])
title('Datové shluky simulované z odhadnuté směsi')

```

Předpověď 'ncom', 'nchn'-rozměrných statických komponent

```

prodini, echo off % zahájení programu
load simMix4 % natažení odhadnuté směsi

% PŘEDPOVĚĎ
nstep = 1; % počet kroků pro předpovědi
pchns = [1 2]; % předpovídané kanály
cchns = [3]; % kanály v podmínce
pMix = mix2pro(Mix, pchns, cchns); % podmíněná směs f(předp.|podm.)
yp = []; time = []; ws = [];

% časová smyčka pro předpovědi
for TIME = Mix.states.maxtd+2:ndat-nstep
    [Ethns, covns, al, w] = profixn(pMix, [], [], nstep);
    % předpovědi z jednotlivých komponent

    th = GetTh(Ethns);
    if 1, ypt = th*w'; % předpovědi (okamžité váhy)
    else ypt = th*al'; % předpovědi (průměrné váhy)
    end
    yp=[yp ypt]; time=[time TIME]; ws=[ws w']; % zaznamenání hodnot
end

% VÝSLEDKY
dd = DATA(pchns,time+1); % simulovaná data
yp = yp(pchns,:); % předpovědi
ep = dd-yp; % chyby předpovědí
disp('Koeficient chyby předpovědi')

```



```
SE = sqrt(var(ep')+mean(ep').^2)./std(dd') % koeficient chyby předpovědi
disp('Odhadnuté váhy komponent')
al % odhady vah komponent

% předpovědi časových průběhů dat
figure(1)
chn = 1; % zobrazovaný kanál
plot(time, dd(chn,:), 'ob', 'MarkerSize', 3)
hold on
plot(time, yp(chn,:), 'xr', 'MarkerSize', 3)
title('Vybraný kanál (o) a jeho předpověď (x)')
hold off

% předpovědi shluků
if length(pchns)>1
    figure(2)
    subplot(121), plot(dd(1,:), dd(2,:), '.', 'MarkerSize', 5)
    axis([-5 20 -5 10]), title('Datové shluky');
    subplot(122), plot(yp(1,:), yp(2,:), '.', 'MarkerSize', 5)
    axis([-5 20 -5 10]), title('Předpovídané shluky');
end
```

Poznámka: Obecný algoritmus simulace, odhadu a předpovídání má sloužit zejména jako výchozí "materiál" pro tvorbu vlastních programů.

Kapitola 9

Spline modely

Modelování dopravních veličin [12, 13] je důležitým prostředkem pro zjišťování okamžitého stavu dopravy v dané oblasti [14, 15], jeho předpověď [16] a jeho případné řízení [17, 18, 19]. Budeme se zabývat tzv. jednorozměrným modelováním dopravy, tj. modelováním jedné dopravní veličiny (např. intenzity provozu) na jednom místě. Měřením této veličiny ve velkém městě dostaneme funkci charakteristického tvaru, který je dán dopravní situací v městech. Přes noc a brzy ráno je provoz téměř nulový. Okolo 7 hod. prudce narůstá a téměř během celého dopoledne je v saturaci. Na poledne poněkud zeslabí a během odpoledne opět vzroste. Navečer kolem 19 hod. postupně slábne a po 21 hod. ustává. Sledovaná funkce má tendenci k periodicitě, avšak velmi zvláštní. Průběhy za všední dny jsou velmi podobné a tvoří jednu skupinu, víkendy jsou odlišné a tvoří druhou skupinu. Průběhy v jednotlivých skupinách se liší jen velmi málo, ale jejich rozdíly určitě netvoří pouhý šum. Průběhy v obou skupinách se dále ještě liší podle ročního období a dalších více nebo méně pravidelných faktorů jako prázdniny, svátky apod. Nepravidelnými faktory, které mohou tvořit řadu dalších samostatných skupin, jsou různé události (návštěvy ze zahraničí, demonstrace, uzavírky komunikací atd.).

Jeden z možných a také dosud nejpoužívanějších přístupů k jednorozměrnému modelování dopravy spočívá v použití statického modelu, opírajícího se o pevný průběh intenzity dopravy v daném místě. Tento průběh se získá prostým měřením v jednotlivých dnech týdne a ověřením jejich "stejnosti" pomocí korelačních metod. Na základě tohoto průběhu se potom provádí předpovědi a řízení dopravy (např. metoda TRANSYT – viz webové stránky <http://www.tdg.co.nz/transyt.htm>).

Předkládaný přístup vychází ze statického modelování a snaží se o jeho vylepšení ve dvou směrech. Za prvé: aproximací se bude průběžně určovat "ideální" průběh intenzity, za druhé: pomocí zapomínání bude možno sledovat pomalé změny ideálního průběhu intenzity během roku. Mezi jednotlivými skupinami funkcí je možno přepínat ručně (podle data), nebo je možno využít multimodelů (kapitola 8) a přepínat nebo míchat jednotlivé komponenty modelu automaticky.

9.1 Spline funkce

Definice spline funkce

Spline funkcí stupně ∂ s defektem $\partial - k$ nazveme funkci $p(x)$ reálného argumentu x , definovanou na systému intervalů $I_i = \langle x_{i-1}, x_i \rangle$, $i = 1, 2, \dots, m$, která splňuje následující dvě podmínky

1. na každém intervalu I_i , $i = 1, 2, \dots, m$, je tvořena polynomem $p_i(x)$ stupně ∂ ,

2. na svém definičním oboru je spojitá a má spojitě derivace až do řádu $k < \partial$.

Body x_i , $i = 0, 1, \dots, m$, se nazývají **uzly** spline-funkce – viz [20, 21].

Poznámka: Defekt je počet nespojitých derivací, který spline funkci "odlišuje od polynomu". Nespojitost nejvyšších derivací dává spline-funkci proti čistému polynomu "větší pružnost".

Poznámka: Prostým názvem spline funkce se většinou označuje spline-funkce s defektem 1, tedy tvořená polynomy stupně ∂ a mající $\partial - 1$ spojitých derivací.

Poznámka: Nejčastěji používaná je spline funkce prvního nebo třetího stupně. Spline funkce sudého stupně se nepoužívají, pokud k tomu není zvláštní důvod.

Vyjádření spline funkce pomocí báze

Spline funkce daného stupně tvoří vektorový prostor. Lze v něm tedy zvolit bázi. Pro naše účely je výhodná báze tzv. **fundamentálních spline funkcí** $q_i(x)$, $i = 0, 1, \dots, m$, která je definována požadavkem, aby i -tá bázová funkce měla hodnotu jedna v i -tém uzlu a ve všech ostatních uzlech byla nulová [20]. Tedy platí

$$q_i(x_i) = 1, \forall i \quad \text{a} \quad q_i(t_j) = 0, \forall i \neq j$$

V takové bázi jsou váhové koeficienty aproximované funkce $f(x)$ přímo rovny hodnotám $f(x_i)$ funkce $f(x)$ v jednotlivých uzlech, tj.

$$f(x) = \sum_{i=0}^m f(x_i)q_i(x) = \sum_{i=0}^m a_i q_i(x).$$

Použití fundamentálních bázových funkcí je z praktického hlediska nejvýznamnější u spline funkcí prvního řádu. Bázové funkce jsou zde totiž nenulové jen na dvou intervalech sousedících s příslušným uzlem. Pro spline funkce vyšších řádů se tato vlastnost lokality bázových funkcí nezachová. Uvedeme si proto příklad pro spline funkci prvního řádu.

Příklad [Báze fundamentálních spline funkcí]

V případě spline funkcí 1. řádu jsou bázové funkce $q_i(x)$, $i = 1, \dots, m + 1$, zadány předpisem

$$q_i(x) = \begin{cases} \frac{x-x_{i-1}}{x_i-x_{i-1}} & \text{pro } x \in I_{i-1} \text{ a } i > 1 \\ \frac{x_{i+1}-x}{x_{i+1}-x_i} & \text{pro } x \in I_i \text{ a } i \leq m \\ 0 & \text{jinde.} \end{cases}$$

Uvažujme konkrétní situaci: zvolme funkci $f(x)$ po částech lineární na intervalech

$$I_1 = \langle 0, 2 \rangle, I_2 = \langle 2, 6 \rangle \text{ a } I_3 = \langle 6, 7 \rangle$$

a hodnoty funkce $f(x)$ v uzlech

$$f(0) = a_0 = 1, \quad f(1) = a_1 = 5, \quad f(2) = a_2 = 2 \quad \text{a} \quad f(3) = a_3 = 3.$$

Bázové funkce budou čtyři

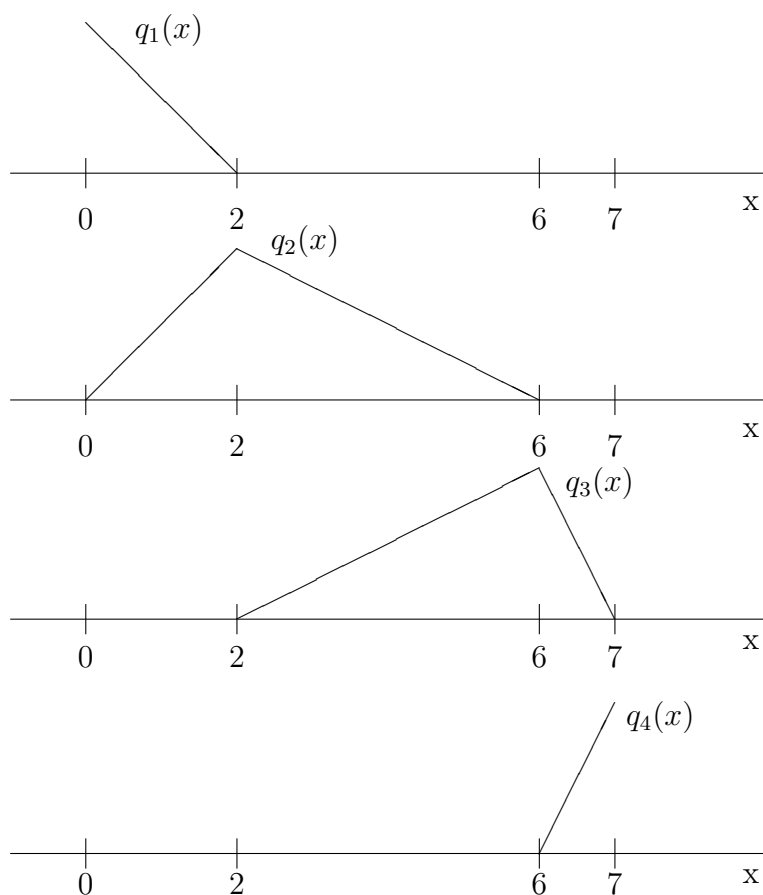
$$q_0(x) = \begin{cases} 1 - 0.5x & \text{pro } I_1 \\ 0 & \text{jinde} \end{cases}$$

$$q_1(x) = \begin{cases} 0.5x & \text{pro } I_1 \\ 1.5 - 0.25x & \text{pro } I_2 \\ 0 & \text{jinde} \end{cases}$$

$$q_2(x) = \begin{cases} 0.25x - 0.5 & \text{pro } I_2 \\ 7 - x & \text{pro } I_3 \\ 0 & \text{jinde} \end{cases}$$

$$q_3(x) = \begin{cases} x - 6 & \text{pro } I_3 \\ 0 & \text{jinde} \end{cases}$$

Jejich grafy jsou na obr. 9.1.



Obrázek 9.1: Fundamentální bázové funkce

Vyjádření spline funkce 1. řádu (tj. lineární lomené čáry s body zlomu v uzlech a hodnotami a_i v bodech zlomu) pomocí bázových funkcí je

$$f(x) = \sum_{i=0}^3 a_i q_i(x) = 1 \cdot q_0(x) + 5 \cdot q_1(x) + 2 \cdot q_2(x) + 3 \cdot q_3(x)$$

Poznámka: Správnost ověřte výpočtem.

(konec příkladu)

Interpolační spline funkce

O interpolační spline funkci hovoříme tehdy, jestliže požadujeme, aby spline funkce procházela předepsanými body. To lze samozřejmě jednoznačně splnit jen v případě, kdy počet předepsaných bodů přesně koresponduje s počtem uzlů, řádem a defektem spline funkce. Např. v případě defektu 1 musí být počet bodů roven počtu uzlů. Říkáme, že spline funkce interpoluje předepsané body.

Konstrukci interpolační spline funkce je nejlépe provést pomocí některé báze, např. pomocí již zmíněné báze fundamentálních spline funkcí. Pro účely modelování však tento druh spline funkcí není příliš zajímavý.

Aproximační spline funkce

Aproximační spline funkci dostaneme tehdy, jestliže počet bodů, které má křivka sledovat, je větší a průchody křivky jednotlivými body nelze splnit. Potom se snažíme, aby tvar křivky byl takový, aby se křivka co nejvíce blížila jednotlivým bodům, např. ve smyslu nejmenších čtverců. Řekneme, že spline funkce aproximuje předepsané body.

Tento druh spline funkcí bude předmětem našeho zájmu při konstrukci spline modelů.

Zobecněné spline funkce

V obou předchozích případech jsme uvažovali spline funkci jako po částech polynomiální funkci s podmínkami hladkosti v uzlech. Podobně, jako jsme při přechodu od interpolačních spline funkcí k aproximačním ustoupili od požadavku přesného průchodu křivky předepsanými body, můžeme nyní uvažovat i o uvolnění přesných požadavků na hladkost křivky v uzlech. Splnění těchto požadavků budeme požadovat jen přibližně, a to s určitou vahou, např. opět metodou nejmenších čtverců – viz (2.10) až (2.13). Přesnou spline funkci obdržíme s velkou vahou pro podmínky hladkosti, pro podmínky hladkosti s nulovou vahou dostáváme aproximaci nezávislými polynomy.

Vlastnosti zobecněných spline funkcí demonstruje obrázek 9.2.

9.2 Práce se spline funkcemi

Pro vyjádření spline-funkce na intervalu I_i je výhodné zavést normovaný argument $\xi_i(x)$ vztahem

$$\xi_i(x) = \frac{x - x_{i-1}}{x_i - x_{i-1}}. \quad (9.1)$$

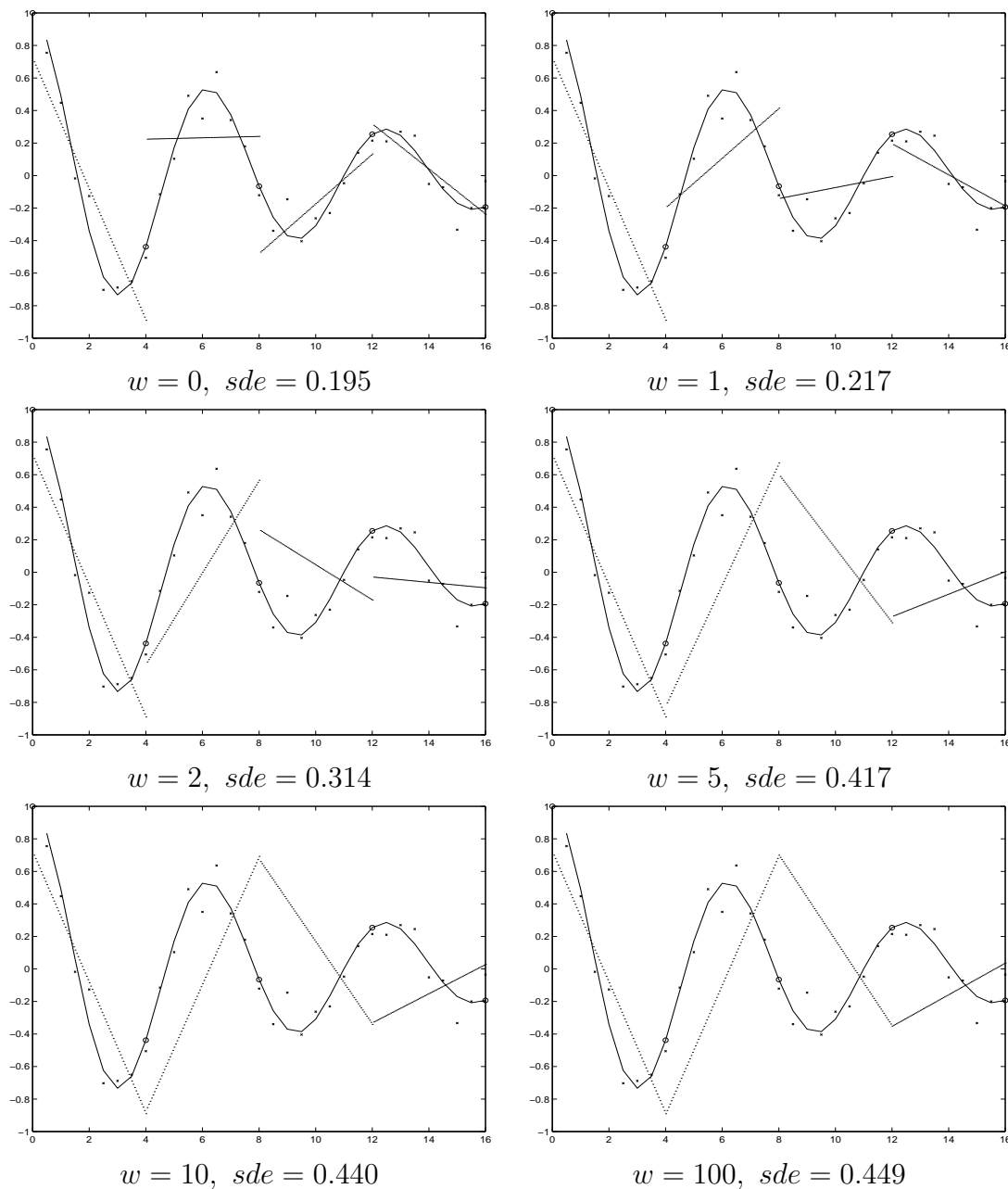
Tento normovaný argument na intervalu $[x_{i-1}, x_i]$ nabývá hodnot z intervalu $[0, 1]$. Jeho zavedením se vyhneme velkým hodnotám koeficientů použitých polynomů a dostaneme podmínky spojitosti v jednoduchém tvaru.

Celou spline funkci lze vyjádřit ve tvaru

$$f(x) = \sum_{i=1}^m p_i(\xi_i(x)) \chi_i(x),$$

Příklad [Normovaný argument]

Uvažujme polynom $p(x)$ pro $x \in (a, b)$ a jeho transformaci $T(x) = gx + h$, $g, h \in R$. Tato transformace představuje změnu měřítka (koeficient g) a posunutí (koeficient h). Vyjádříme-li transformovaný polynom v původní proměnné x , bude mít samozřejmě jiné koeficienty.



Plná čára značí původní funkce, křížky měřené body, tj. funkce a šum, kroužky označují uzly, tečkovaná čára představuje zobecněnou spline funkci. Váha podmínek hladkosti (zde spojitosti) je označena w . Kvalita aproximace je vyjádřena pomocí směrodatné odchylky sde chyby aproximace.

Z obrázků je zřejmé, že požadavek hladkosti jde na úkor kvality aproximace. Horší aproximace hladší funkcí však je většinou chápána jako filtrace zašuměné hladké křivky. Váha hladkosti pak určuje míru filtrace.

Obrázek 9.2: Aproximace pomocí zobecněné spline-funkce.

Například, když polynom $p(x) = 1 - 2x + x^2$ pro $x \in (0, 2)$ transformujeme posunutím deset jednotek vpravo, bude mít tvar

$$p(x - 10) = 1 - 2(x - 10)x + (x - 10)^2 = 121 - 22x + x^2.$$

Uvedenou transformaci je však možno provést nejen přepočtem koeficientů polynomu, ale také zavedením již zmíněného normovaného argumentu. Ten bude pro každou transformaci jiný podle vzorce (9.1) a bude záviset na hranicích transformovaného definičního oboru (a, b) . Pro takto zavedený normovaný argument bude navíc platit $\xi(a) = 0$ a $\xi(b) = 1$.

Pro uvedený číselný příklad bude normovaný argument dán výrazem

$$\xi(x) = \frac{x - 10}{12 - 10} = 0.5x - 5$$

a transformovaný (posunutý) polynom bude mít tvar

$$p(\xi(x)) = 1 + 4\xi + 4\xi^2,$$

s tím, že platí $\xi(10) = 0$ a $\xi(12) = 1$. Tento tvar bude pro všechny transformace stejný (samozřejmě vždy se svým normovaným argumentem) a normovaný argument bude vždy 0 na začátku intervalu a 1 na jeho konci. (konec příkladu)

Spline funkce 1. stupně

Na intervalu I_i je spline-funkce definována polynomem

$$p_i(\xi_i(x)) = a_i + b_i \xi_i(x). \tag{9.2}$$

Podmínky hladkosti se zde redukuje na podmínky spojitosti, které pro dva navazující intervaly I_i a I_{i+1} , zahrnující uzly x_{i-1} , x_i a x_{i+1} , představují polynomiální rovnici

$$p_i(\xi_i(x_i)) = p_{i+1}(\xi_{i+1}(x_i)), \quad \text{tj.} \quad p_i(1) = p_{i+1}(0).$$

Dosadíme-li do předchozího vztahu vyjádření polynomů podle (9.2), dostaneme rovnici, vyjadřující spojitý přechod v uzlu x_i

$$a_i + b_i = a_{i+1}. \tag{9.3}$$

Odtud je vidět, že když stanovíme jednu počáteční podmínku vlevo (nejspíše hodnotu, kde spline funkce začíná), zbývá pro každý interval jen jediný volný koeficient. Druhý je určen podmínkami spojitosti.

Chceme-li tyto podmínky vyjádřit přibližně, použijeme tvar regresního modelu

$$a_i + b_i = a_{i+1} + \varepsilon, \tag{9.4}$$

kde

$a_i + b_i$ je známé číslo, (zde představuje změřenou hodnotu),

a_{i+1} je nastavovaný koeficient (konstanta modelu) a

ε je obecné označení chyby v přesném splnění podmínek (šum modelu).

Poznámka: Podmínky lze použít buď bez šumu pro konstrukci "přesných" spline funkcí, nebo se šumem je přidat mezi podmínky aproximační. Výsledná spline funkce bude potom kompromisem mezi požadavky na aproximaci a spojitost.

Spline funkce 3. stupně

Na intervalu I_i je spline-funkce definována polynomem

$$p_i(\xi_i(x)) = a_i + b_i\xi_i(x) + c_i\xi_i^2(x) + d_i\xi_i^3(x). \quad (9.5)$$

Podmínky hladkosti jsou zde vyjádřeny podmínkami na spojitost a návaznost 1. a 2. derivace (značíme apostrofem ' a ''), tedy pro uzel x_i

$$p_i(1) = p_{i+1}(0), \quad p_i'(1) = p_{i+1}'(0), \quad p_i''(1) = p_{i+1}''(0).$$

Po dosazení (9.5) dostáváme podmínky pro koeficienty polynomů

$$a_i + b_i + c_i + d_i = a_{i+1}, \quad (9.6)$$

$$b_i + 2c_i + 3d_i = b_{i+1}, \quad (9.7)$$

$$c_i + 3d_i = c_{i+1}. \quad (9.8)$$

Podmínky hladkosti lze samozřejmě použít stejně jako v případě spline funkce 1. stupně.

9.3 Spline model intenzity dopravního toku

Jak jsme již uvedli, budeme modelovat jednu dopravní veličinu (intenzitu dopravního toku) na jednom místě. Budeme uvažovat neřízený statický model ve tvaru – srv. (4.14). Protože veličiny na veličiny v modelu pohlížíme jako na funkce času, budeme opět místo symbolu x v indexu používat symbol času t .

$$y_t = f_t + e_t, \quad (9.9)$$

kde

t je časový okamžik měření od určité referenční hodiny (např. 5.00 hod ráno).

y_t je modelovaná intenzita v časovém okamžiku t ,

f_t je hodnota typického průběhu intenzity v čase t ,

e_t je porucha, odchylka okamžité intenzity od její typické hodnoty.

Typickou hodnotu intenzity budeme modelovat pomocí zobecněné spline-funkce

$$f_t = \sum_{i=1}^m p_i(\xi_i(t))\chi_i(t), \quad (9.10)$$

kde p_i je polynom (9.2) nebo (9.5) na jehož koeficienty klademe "přibližné podmínky hladkosti".

Dosadíme-li předchozí vyjádření typické hodnoty průběhu intenzity do modelu (9.9), můžeme jej pro jeden den psát ve tvaru statického regresního modelu s proměnnou konstantou

$$y_t = \theta^T \varphi_t + e_t \quad (9.11)$$

kde

θ je vektor parametrů (koeficientů všech polynomů tvořících spline funkci)

$\theta = [p_1, p_2, \dots, p_m]^T$, kde p_i je vektor koeficientů i -tého polynomu, řazený od absolutního členu.

Pro 1. resp. 3. stupeň spline funkce je

$$p_i = [a_i, b_i] \quad \text{resp.} \quad p_i = [a_i, b_i, c_i, d_i], \quad i = 1, 2, \dots, m,$$

φ_t je regresní vektor

$$\varphi_t = [\phi_1\chi_1(t), \phi_2\chi_2(t), \dots, \phi_m\chi_m(t)].$$

Opět pro 1. resp. 3. stupeň spline funkce a $i = 1, 2, \dots, m$ platí

$$\phi_i = [1, \xi_i(t)], \text{ resp. } \phi_i = [1, \xi_i(t), \xi_i^2(t), \xi_i^3(t)].$$

Regresní vektor je tedy tvořen nulami, jen na místě, odpovídajícím právě aktuálnímu intervalu jsou umístěny normované argumenty příslušného polynomu. Takto konstruovaný regresní vektor představuje aproximační podmínky (tedy požadavek, aby spline funkce ležela blízko změřeným hodnotám).

Abychom zaručili požadovanou hladkost zobecněné spline-funkce, je třeba ještě zajistit "přibližné" splnění podmínek hladkosti, rovněž modelovaných ve tvaru regresního modelu (viz (9.4)), které pro 1. stupeň spline funkce jsou

$$a_i + b_i = a_{i+1} + \varepsilon \quad (9.12)$$

a pro 3. stupeň

$$a_i + b_i + c_i + d_i = a_{i+1} + \varepsilon \quad \text{spojitost} \quad (9.13)$$

$$b_i + 2c_i + 3d_i = b_{i+1} + \varepsilon \quad \text{1. derivace} \quad (9.14)$$

$$c_i + 3d_i = c_{i+1} + \varepsilon \quad \text{2. derivace} \quad (9.15)$$

Vezmeme-li v úvahu strukturu vektoru parametrů, zavedenou modelem (9.11), odpovídají těmto podmínkám hodnoty výstupu rovny levým stranám předchozích rovnic (určeným z koeficientů předchozího polynomu) a regresní vektory, tvořené nulami a obsahující jedničku na příslušném místě odpovídajícím koeficientu z pravé strany rovnice. Tak např. pro 3. stupeň spline funkce a druhý interval I_2 bude tzv. virtuální regresní vektor φ roven:

$$\varphi = [0, 0, 0, 1, 0, \dots] \quad \text{pro spojitost}$$

$$\varphi = [0, 0, 0, 0, 1, 0, \dots] \quad \text{pro 1. derivaci}$$

$$\varphi = [0, 0, 0, 0, 0, 1, 0, \dots] \quad \text{pro 2. derivaci.}$$

Hodnoty výstupu budou postupně: $a_1 + b_1 + c_1 + d_1$; $b_1 + 2c_1 + 3d_1$ a $c_1 + 3d_1$.

Podobné vztahy platí i pro další intervaly, hodnoty na prvním intervalu jsou konstruovány na základě okrajových podmínek.

Odhad parametrů probíhá např. metodou nejmenších čtverců, kam jako datový vektor dodáváme skutečný regresní vektor φ_t podle (9.11) a podmínky hladkosti respektujeme přidáním virtuálních regresních vektorů s příslušnými vahami. Velikost vah, s nimiž tyto regresní vektory použijeme určuje míru přesnosti splnění požadavků na hladkost aproximující zobecněné spline-funkce.

9.4 Program pro odhadování spline-modelu

Následující program generuje postupně zašuměné hodnoty funkce zadané procedurou `ym = func(t)` (zde je to tlumená funkce kosinus) a provádí jejich aproximaci zobecněnou spline funkcí prvního nebo třetího řádu určenou metodou nejmenších čtverců. Výsledek aproximace zobrazuje program ve formě grafu. Kromě předpisu pro aproximovanou funkci je v programu možno zadat

xx - vektor uzlů zobecněné spline funkce,
 ae - amplituda šumu měření při vzorkování aproximované funkce,
 h - (hustá) perioda vzorkování aproximované funkce,
 ord - řád spline funkce (1 nebo 3),
 w - váhy podmínek hladkosti (spojitost, první a druhá derivace),
 hs - perioda vzorkování pro zobrazení výsledků.

Hlavní program

```

% Aproximace pomocí zobecněné spline-funkce 1. nebo 3. řádu
% -----
clc, clear all, randn('seed',1234)
xx = 0:4:16; % uzly (i neekvidistantní)
ae = .1; % amplituda šumu měření
ff = func(xx); % konstrukce pro graf
h = .5; % perioda vzorkování
ord= 3; % řád spline funkce (1|3)
w = [1000 10 0]; % váhy pro 0., 1. a 2. derivaci
s = min(xx)+h:h:max(xx); % body vzorkování
no = ord+2;

% Měření hodnot a odhad
V = 1e-5*eye(no); % počáteční rozšířená informační matice
yy=[]; y1=[]; aa=[];

% cyklus pro čas
for t=s
    e = ae*randn; % šum měření
    ym = func(t); % výstup bez šumu
    y = ym+e; % výstup se šumem
    yy=[y y]; y1=[y1 ym]; % body funkce (vektor y)
    [ti, ii]=xi(t, xx); % norm. argument a interval pro y
    in=n_int(t+h, xx); % interval pro příští y
    psi = datv(y,ti,ord); % datový vektor pro 1. řád
    V = V+psi'*psi; % update informační matice
    if in>ii % přechod na nový interval
        a = inv(V(2:no,2:no))*V(2:no,1);% rozdělení inf. matice
        aa = [aa a]; % koef. polynomů (po sloupcích)
        psi= datv0(w,a,ord); % podmínky hladkosti
        V = psi'*psi; % inf. matice - podm. hladkosti
    end
end

% Výsledky
hs = .05; % krok vzorkování
ss = min(xx)+hs:hs:max(xx); % body vzorkování
ys = [];
for t=ss % generování spline funkce
    [ti, ii]=xi(t, xx); % norm. argument a č. intervalu

```

```

psi = datv([],ti,ord);           % datový vektor
yp = psi*aa(:,ii);             % hodnota spline funkce
ys = [ys yp];
end
ep = func(ss)-ys;              % chyba aproximace
sde = std(ep)                  % směr. odch. chyby aproximace
figure(1),plot(s,yy,'rx',s,y1,'r:',ss,ys,'c',xx,ff,'om','markersize',5)
title('Čistá funkce (..), funkce se šumem (x), aproximace (-), uzly (o)')

```

Doporučené experimenty

1. Bezprostřední experiment je změna aproximované funkce. Tu je možno zadat změnou podpůrné funkce `func`. Zkuste si, co aproximace "dokáže". Brzy ale zjistíte, že je třeba měnit i další parametry aproximace. O nich se zmíníme v dalším.
2. Pro stejnou funkci lze ještě měnit amplitudu šumu měření pomocí koeficientu `ae = .1`. Větší šum, např. `ae = 1` znamená samozřejmě obtížnější aproximaci.
3. Změna periody spline-aproximace. Provedeme ji změnou uzlů aproximace, které jsou zadány ihned na začátku programu příkazem `xx = 0:4:16`. Zadává se počátek : krok : konec. Zkrácení periody lze tedy provést např. příkazem `xx = 0:2:16`, prodloužení příkazem `xx = 0:8:16`. Zkuste s původní funkcí a porovnejte.
4. Spline funkce je po částech polynomiální funkce, kde polynomiální úseky jsou napojeny s určitou hladkostí. V našem pojetí lze dokonce určovat míru požadované hladkosti, a to prostřednictvím vah `w = [1000 10 0]`. První složka určuje míru hladkosti nulté derivace (tj. spojitost), druhá složka míru hladkosti první a třetí složka míru hladkosti druhé derivace. Čím je koeficient větší, tím vyšší míra hladkosti se požaduje.
5. Změna periody vzorkování zašuměné funkce se provede změnou parametru `h = .5`. Kratší perioda vzorkování znamená více bodů pro aproximaci. Perioda uzlů v aproximaci může být kratší, pokud je to potřeba.
6. Parametrem `ord = 3` lze změnit spline-funkci 3. řádu na řád 1. Obecně se doporučuje 3. řád, 1. řád je však velmi jednoduchý a mnohdy postačí. Zkuste pro různé funkce různé řády spline-aproximace.

Podpůrné funkce

Aproximovaná funkce

```

function y=func(x)
% y=func(x)
% aproximovaná funkce
%
% y   hodnota funkce
% x   argument funkce
%
y = cos(x).*exp(-.1*x);

```

Normovaný argument polynomu

Počítá hodnotu normovaného argumentu podle vzorce (9.1).

```
function [xx, ii]=xi(xt, x)
% xx = xi(xt, x)
% výpočet normovaného argumentu pro spline-aproximaci
%
% xx   normovaný argument
% ii   pořadí aktuálního intervalu
% xt   aktuální argument
% x    vektor spline-uzlů
%
ii = n_int(xt,x);
xx = (xt-x(ii))/(x(ii+1)-x(ii));
```

Pořadí aktuálního intervalu spline funkce

Podle zadané aktuální hodnoty argumentu x a vektoru uzlů spline funkce určí pořadí aktuálního intervalu spline funkce. Např. pro $x = 5$ a uzly $xi = [1, 3, 7, 10]$ je aktuální druhý interval.

```
function ii=n_int(xt, x)
% i = n_int(xt, x)
% pořadí aktuálního spline-intervalu
%
% ii   pořadí aktuálního intervalu
% xt   aktuální argument
% x    vector spline-uzlů
%
ii = sum(x<xt);
```

Konstrukce datového vektoru pro aproximaci

Datový vektor je tvořen změřenou hodnotou aproximované funkce a příslušnými mocninami normovaného argumentu. Pro spline funkce prvního řádu je to jen nultá a první mocnina (jednička a hodnota argumentu), pro třetí řád mocniny řádu nula až tři.

```
function dv=datv(y,xi,ord)
% dv = datv(y,xi,ord)
% konstrukce datového vektoru
%
% dv   datový vektor = [data, regresní vektor]
%           = [hodnota funkce, mocniny xi]
% y    změřená hodnota dat
% xi   normovaný argument
% ord  řád spline-funkce
%
if isempty(y), dv=1;
else          dv=[y 1];
end
for i=1:ord
    dv = [dv xi^i];
end
```

Konstrukce datového vektoru pro podmínky hladkosti

Virtuální regresní vektory se konstruují podle vztahů (9.3) pro první řád, nebo (9.6), (9.7), (9.8) pro třetí řád.

```
function dv=datv0(w,a,ord)
% dv = datv0(w,a,ord)
% konstrukce datového vektoru pro podmínky hladkosti
%
% dv   datový vektor
% w    váhy pro podmínky hladkosti
% a    parametry polynomu z předchozího intervalu
% ord  řád spline-funkce
%
dv=w(1)*[sum(a) 1 zeros(1,ord)]; % pro 1. řád
if ord==3 % pro 3. řád se přidá
    dv1 = w(2)*[a(2)+2*a(3)+3*a(4) 0 1 0 0];
    dv2 = w(3)*[2*a(3)+6*a(4) 0 0 2 0];
    dv = dv+dv1+dv2;
end
```

Kapitola 10

Příklady

10.1 Regresní model se střední hodnotou ve tvaru polynomu

Lineárním modelem máme na mysli model, jehož deterministickou část můžeme vyjádřit pomocí skalárního součinu $\varphi_t^T \theta$ (viz (4.7)), kde φ_t je regresní vektor a θ je vektor parametrů. Uvažujme regresní vektor s deterministickou částí ve tvaru polynomu

$$p(t) = p_k t^k + p_{k-1} t^{k-1} + \dots + p_1 t + p_0.$$

Určete:

- Regresní vektor a vektor parametrů takového modelu.
- Typ tohoto modelu.

ŘEŠENÍ

- Regresní model má tvar

$$y_t = p_k t^k + p_{k-1} t^{k-1} + \dots + p_1 t + p_0 + e_t = \varphi_t^T \theta + e_t$$

kde

$$\varphi_t = [t^k, t^{k-1}, \dots, t, 1]^T \quad \text{a} \quad \theta = [p_k, p_{k-1}, \dots, p_1, p_0]^T.$$

- Jedná se o lineární statický regresní model, kde konstanta modelu je časově proměnná a je vyjádřena pomocí polynomu. Viz str. 55.

Poznámka: Takový model je vhodný např. pro odhad závislosti intenzity dopravního proudu na jeho hustotě. Tato závislost má přibližně tvar konkávní paraboly. Připustíme-li odchylky od takového tvaru, pak dobrým základem pro modelování je polynom.

10.2 Regresní model popisující exponenciálu

Určete regresní model popisující odezvu systému popsaného diferenciální rovnicí

$$y' = -\alpha y, \quad y(0) = 1, \quad \alpha \geq 0$$

ŘEŠENÍ 1

Přibližné řešení dostaneme, nahradíme-li derivace diferencemi, tedy

$$y'(\tau) \rightarrow \frac{y(\tau+h) - y(\tau)}{h},$$

kde h je délka kroku diskretizace. Zadanou diferenciální rovnici pak lze přibližně zapsat takto

$$\frac{y(\tau+h) - y(\tau)}{h} = -\alpha y(\tau) \quad \Rightarrow \quad y(\tau+h) = (1 - h\alpha)y(\tau)$$

Zavedeme normovaný čas t tak, že označuje nejnovější časový okamžik, tady $t = \tau + h$ a $h = 1$, a budeme jej vyznačovat jako index příslušné proměnné. Bude tedy $y_t = y(\tau+h)$ a $y_{t-1} = y(\tau)$. Příslušná diferenční rovnice (hledaný regresní model) bude

$$y_t = ay_{t-1}, \quad y_0 = 1,$$

kde $a = 1 - h\alpha$.

Poznámka: Tato metoda selhává pro velmi malá h . Pro $h \rightarrow 0$ bude $a \rightarrow 1$, $y_t = y_{t-1}$ a odvozená diferenční rovnice degeneruje.

ŘEŠENÍ 2

Přesné řešení dostaneme tak, že určíme řešení zadané diferenciální rovnice, a pak nalezneme diferenční rovnici, která toto řešení generuje. Řešením diferenciální rovnice s počáteční podmínkou $y(0) = 1$ je funkce

$$y(\tau) = e^{-\alpha\tau}.$$

Pro posunutou hodnotu platí

$$y(\tau+h) = e^{-\alpha(\tau+h)} = e^{-\alpha\tau}e^{-\alpha h} = y(\tau)e^{-\alpha h}.$$

Odpovídající diferenční rovnice je

$$y_t = \tilde{a}y_{t-1}, \quad \tilde{a} = e^{-\alpha h}.$$

Poznámka: Porovnání obou výsledků zjistíme, že koeficient a z prvního přibližného postupu je prvním členem Taylorova rozvoje koeficientu \tilde{a} z druhého přesného postupu.

10.3 Regresní model popisující funkci sinus

Určete regresní model, který popisuje funkci $y(t) = \sin(t)$.

ŘEŠENÍ

Diferenciální rovnici generující funkci $y(t) = \sin(t)$ dostaneme tak, že tuto funkci dvakrát derivujeme: $y'(t) = \cos(t)$, $y''(t) = -\sin(t)$ a v druhé derivaci dosadíme: $y''(t) = -y(t)$. Hledaná diferenciální rovnice je

$$y''(t) + y(t) = 0, \quad y(0) = 0, \quad y'(0) = 1.$$

Další postup je obdobný jako v předchozím příkladě, proto jej jen naznačíme.

Přibližné řešení: Derivace nahradíme diferencemi a dostaneme

$$y_t - 2y_{t-1} + ay_{t-2} = 0, \quad a = 1 + h^2, \quad y(0) = 0, \quad y(-h) = -\sin(h).$$

Přesné řešení: Vyjádříme dvakrát a jednou posunutou funkci $y(\tau) = \sin(\tau)$:

$$\begin{aligned} y(\tau + 2h) &= \sin(\tau + 2h) = \cos(h) \sin(\tau + h) - (\sin(h))^2 \sin(\tau) + \sin(h) \cos(h) \cos(\tau) \\ y(\tau + h) &= \sin(\tau + h) = \cos(h) \sin(\tau) + \sin(h) \cos(\tau) \end{aligned}$$

Z druhé rovnosti vyjádříme $\cos(\tau)$ a dosadíme do první. Rovněž dosadíme $\sin(\tau) = y(\tau)$ a $\sin(\tau + h) = y(\tau + h)$. Tak dostaneme

$$y_t - 2\tilde{a}y_{t-1} + y_{t-2} = 0, \quad \tilde{a} = \cos(h), \quad y_0 = 0, \quad y_{-1} = -\sin(h)$$

Následuje program realizující uvedené výpočty.

```
% Diferenční rovnice generující funkci y=sin(x)
% -----
clc, clear all
N = 20;           % počet kroků generování
h = .05;         % délka kroku generování
tt= 0:h:N;       % hodnoty t pro generování

% přibližné řešení diskretizace
a = 1+h^2;       % koeficient diferenční rovnice
y1 = 0;         % jednou zpožděný výstup na počátku
y2 = -sin(h);   % dvakrát zpožděný výstup na počátku
yy1=[];
for t=tt
    y = 2*y1-a*y2; % hodnota výstupu modelu
    yy1= [yy1 y];
    y2 = y1;      % posun hodnot staré -> ještě starší
    y1 = y;       % nová -> stará
end

% přesné řešení diskretizace
a = cos(h);     % koeficient diferenční rovnice
y1 = 0;         % jednou zpožděný výstup na počátku
y2 = -sin(h);   % dvakrát zpožděný výstup na počátku
yy2=[];
for t=tt
    y=2*a*y1-y2; % hodnota výstupu modelu
    yy2=[yy2 y];
    y2=y1;      % posun hodnot staré -> ještě starší
    y1=y;       % nová -> stará
end

% zobrazení výsledků
figure(1),plot(tt,yy1,':',tt,yy2,'-')
title('Přibližná (..) a přesná (-) simulace funkce sin(x)')
```

Doporučené experimenty

1. S rostoucím počtem kroků simulace N se zvětšuje rozdíl mezi přibližným a přesným řešením.
2. S kratší periodou vzorkování se rozdíly mezi přibližným a přesným řešením zmenšují.
3. Zkuste odvodit vzorce a pozměnit program tak, aby generoval funkci $y = \cos(x)$.

10.4 Regresní model s rovnoměrným šumem

Určete model chyby zaokrouhlování uříznutím desetinné části čísla

$$\delta_t = y_t - [y_t],$$

kde symbol $[\cdot]$ značí příslušné zaokrouhlení. Určete střední hodnotu a rozptyl chyby δ_t .

ŘEŠENÍ

Chyba zaokrouhlení může být v intervalu $(0; 1)$ a předpokládáme-li, že zaokrouhlovaná čísla přicházejí zcela náhodně, bude rozdělena rovnoměrně. Model tedy bude – srv. odstavec 4.1.3

$$\delta_t = 0.5 + e_t, \quad f(e_t) = \begin{cases} 1 & \text{pro } e_t \in (-0.5; 0.5) \\ 0 & \text{jinde} \end{cases}$$

a pro šum modelu platí

$$E[e_t] = \int_{-0.5}^{0.5} e_t f(e_t) de_t = 0; \quad D[e_t] = \int_{-0.5}^{0.5} e_t^2 de_t - (E[e_t])^2 = 1/12.$$

Střední hodnotu a rozptyl chyby zaokrouhlení δ_t můžeme určit pomocí modelu

$$E[\delta_t] = E[0.5 + e_t] = 0.5 + E[e_t] = 0.5$$

a

$$D[\delta_t] = D[0.5 + e_t] = 0 + D[e_t] = \frac{1}{12}.$$

Poznámka: Výsledky, které jsme obdrželi s pomocí modelu můžeme ověřit přímým výpočtem: Hustotu pravděpodobnosti chyby zaokrouhlení d_t dostaneme transformací hp šumu podle rovnice modelu. Tato hustota má rovněž rovnoměrné rozdělení

$$f(\delta_t) = \begin{cases} 1 & \text{pro } \delta_t \in (0; 1) \\ 0 & \text{jinde} \end{cases}$$

Střední hodnota (v libovolném čase)

$$E[\delta] = \int_0^1 \delta \, 1 d\delta = 0.5$$

Rozptyl (v libovolném čase)

$$D[\delta] = E[\delta^2] - (E[\delta])^2 = \int_0^1 \delta^2 \, 1 d\delta - 0.5^2 = \frac{1}{12}.$$

Výsledky se shodují s těmi, které jsme obdrželi s pomocí modelu.

10.5 Diskrétní model

Sledujeme malou oblast městských komunikací a podle její úrovně provozu rozlišujeme stupně dopravy 1, 2 a 3. Napište model, který popisuje tuto situaci.

ŘEŠENÍ

Vzhledem k platnosti základních fyzikálních zákonů (zachování hmoty a setrvačnosti) předpokládáme, že bude existovat určitá vazba mezi dvěma následujícími stavy. Zvolíme proto dynamický model a pro jednoduchost jej ponecháme 1. řádu. Tento model lze vyjádřit podmíněnými pravděpodobnostmi – srv. odstavec 4.1.4

$$f(s_t = i | s_{t-1} = j, \Theta) = \Theta_{i,j}, \quad (10.1)$$

kde

s_t a s_{t-1} jsou stavy dopravy v časech t a $t-1$; $i, j \in \{1, 2, 3\}$ a

Θ je matice pravděpodobností přechodu mezi jednotlivými stavy.

Uvedený model lze psát také ve tvaru

$$f(s_t = i | s_{t-1} = j, \Theta) = \prod_i \prod_j \Theta_{i,j}^{\delta(s_t, i) \delta(s_{t-1}, j)},$$

kde $\delta(s, i)$ je Kroneckerova funkce $\delta(s, i) = 1$ pro $s = i$ a $\delta(s, i) = 0$ pro $s \neq i$. Toto druhé vyjádření modelu je vhodné pro další odhadování.

10.6 Spojitý model s diskretní veličinou

Na jednom místě dopravní komunikace modelujeme intenzitu dopravního proudu. Situace je komplikovaná tím, že v sousedství je určitá oblast, která bývá často uzavřena. Tato uzavírka má podstatný vliv na provoz v modelovaném místě, neboť tudy vede objížďka uzavřené oblasti. Napište model intenzity dopravního proudu ve sledovaném místě, jestliže

- zanedbáme vlastní dynamiku proudu a sledujeme jen vliv uzavírky v sousedství,
- chceme postihnout současně dynamiku proudu i vliv uzavírky,
- uvažujeme různé rozptyly modelů pro jednotlivé stavy uzavírky.

ŘEŠENÍ

Uvažujeme situaci bez uzavírky $s_t = 1$, s uzavírkou $s_t = 2$.

a) *Bez vlastní dynamiky*: Pro jeden konkrétní stav uzavírky $s_t = I$, $I \in \{1, 2\}$ bude intenzita y_t popsána statickým regresním modelem – srv. (4.27)

$$f(y_t | s_t = I, \Theta) = \frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} (y_t - \mu_I)^2 \right\},$$

kde μ_I jsou střední hodnoty, různé pro různé modely, a r je společný rozptyl šumu.

Tento model můžeme vyjádřit jako součin jednotlivých modelů, umocněných na příslušnou Kroneckerovu funkci

$$f(y_t | s_t = I, \Theta) = \prod_{i=1}^2 \left[\frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} (y_t - \mu_i)^2 \right\} \right]^{\delta(i, I)} \quad (10.2)$$

$$= \frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} \sum_{i=1}^2 (y_t - \mu_i)^2 \delta(i, I) \right\}. \quad (10.3)$$

Všimneme si sumy v exponentu modelu a doplníme ji na úplný čtverec

$$(y_t - \mu_1)^2 \delta(1, I) + (y_t - \mu_2)^2 \delta(2, I) = [y_t - \mu_1 \delta(1, I) - \mu_2 \delta(2, I)]^2.$$

Při úpravách jsme použili skutečnosti, že $\delta(1, I)\delta(2, I) = 0$, neboť vždy právě jeden z činitelů je roven nule. S touto úpravou exponentu má model tvar

$$f(y_t | s_t = I, \Theta) = \frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} [y_t - \mu_1 \delta(1, I) - \mu_2 \delta(2, I)]^2 \right\}.$$

Regresní vektor a vektor parametrů takového modelu je

$$\varphi_t = [\delta(1, I), \delta(2, I)]^T, \quad \Theta = [\mu_1, \mu_2]^T$$

Poznámka: *Problematikou kombinace spojitých a diskrétních veličin při modelování a odhadu se blíže zabývá článek [10].*

b) *S vlastní dynamikou:* Předchozí statický model rozšíříme na dynamický prvního řádu. Budeme tedy uvažovat závislost intenzity y_t na intenzitě y_{t-1} , zjištěné v předchozím okamžiku měření. Vzhledem k zavedení této dynamiky je rozumné uvažovat rovněž dynamiku v diskrétní proměnné – uzavírce oblasti v sousedství. Pro jednoduchost budeme stále uvažovat rozptýl šumu společný všem jednotlivým komponentám modelu. Nový model bude mít tvar

$$f(y_t | y_{t-1}, s_t = I, s_{t-1} = J, \Theta) = \frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} (y_t - a_{I,J} y_{t-1} + b_{I,J})^2 \right\},$$

kde $I, j \in \{1, 2\}$ jsou hodnoty stavu uzavírky v současném a minulém okamžiku měření. Model lze opět vyjádřit jako součin (10.2) a (10.3)

$$\begin{aligned} f(y_t | y_{t-1}, s_t = I, s_{t-1} = J, \Theta) &= \prod_{i,j} \left[\frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} (y_t - a_{i,j} y_{t-1} + b_{i,j})^2 \right\} \right]^{\delta(i,j; I, J)} = \\ &= \frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} \sum_{i,j} [(y_t - a_{i,j} y_{t-1} + b_{i,j})^2 \delta(i, j; I, J)] \right\} = \\ &= \frac{1}{\sqrt{2\pi r}} \exp \left\{ -\frac{1}{2r} (y_t - \sum_{i,j} a_{i,j} y_{t-1} \delta(i, j; I, J) + \sum_{i,j} b_{i,j} \delta(i, j; I, J))^2 \right\}, \end{aligned}$$

kde jsme opět využili skutečnosti, že $\sum_{i,j} \delta(i, j; I, J) = 1$ a $\delta(i_1, j_1; I, J)\delta(i_2, j_2; I, J) = 0$ pro $(i_1, j_1) \neq (i_2, j_2)$.

Odtud plyne, že regresní vektor a vektor parametrů budou

$$\varphi_t = [y_{t-1} \delta_{11}, \delta_{11}, y_{t-1} \delta_{12}, \delta_{12}, y_{t-1} \delta_{21}, \delta_{21}, y_{t-1} \delta_{22}, \delta_{22}]^T,$$

kde $\delta_{i,j} = \delta(i, j; I, J)$ a

$$\Theta = [a_{11}, b_{11}, a_{12}, b_{12}, a_{21}, b_{21}, a_{22}, b_{22}]^T.$$

c) *S různými rozptyly v jednotlivých stavech:* Předchozí dvě úlohy byly velmi zjednodušeny tím, že jsme uvažovali shodné rozptyly modelů v jednotlivých stavech. Díky tomu jsme obdrželi speciální řešení pro odhad s kompaktním regresním vektorem, společným pro všechny stavy. Tato úloha ukazuje standardní postup odhadování modelu se smíšenými (tj. spojitými i diskrétními) daty. Řešení lze stručně shrnout takto:

- kolik je různých hodnot diskrétní veličiny (tj. kolik je stavů systému), tolik je různých statistik pro odhad parametru,
- změřená hodnota diskrétní veličiny v okamžiku t vybere příslušnou statistiku (odpovídající příslušnému stavu) a ta se přepočte s daty v tomto okamžiku naměřenými,
- každá statistika poskytuje odhad parametrů modelu pro odpovídající stav.

Model systému popíšeme hp $f(y_t, s_t = I|\Theta) = f_I(y_t|\Theta_I)$, kde jsme vyjádřili skutečnost, že příslušný stav pouze vybere příslušný model se svými parametry (index u Θ) a případně svojí strukturou (index u f).

Poznámka: V případě, že bychom chtěli stavy modelovat, by rozklad sdruženého modelu byl následující: $f(y_t, s_t = I|\Theta) = f(y_t|s_t = I, \Theta_I)f(s_t|\Theta)$, kde druhá hp je tabulka, tady vektor, pravděpodobností jednotlivých stavů, případně parametrizovaný (s parametry zahrnutými do vektoru Θ).

Bayesův vzorec pro uvažovaný model a n_s různých stavů má tvar

$$f(\Theta|D(t)) \propto f(y_t, s_t|D(t-1), \Theta)f(\Theta|D(t-1)).$$

Vyjádříme-li hp parametrů v součinném tvaru a model systému zapíšeme pro příslušný změřený stav $s_t = I$, dostaneme

$$\prod_{i=1}^{n_s} f(\Theta_i|D(t)) \propto \prod_{i=1}^{n_s} f_i(y_t|s_t = i, \Theta_i)^{\delta(i,I)} f(\Theta_i|D(t-1)).$$

Obě strany předchozího výrazu můžeme integrovat přes všechna Θ_j , kde $j \neq I$. Dostaneme

$$f(\Theta_I|D(t)) \propto f_I(y_t|s_t = I, \Theta_I)f(\Theta_I|D(t-1)),$$

což je "obyčejný" Bayesův vztah pro přepočet statistik modelu, odpovídajícího stavu I .

Další postup odhadování je již standardní. Následuje program odhadující parametry statického modelu s třemi stavy a se separátním přepočtem jednotlivých statistik.

```
% Odhad modelu se smíšenými daty
% -----
% - simulovány 3 stavy s pravděp. al=0.2, 0.5, 0.3
% - model každého ze stavů je statický se
%   střední hodnotou mi a směrodatnou odchylkou sg
% - výsledné parametry jsou Md{i}.th
%
clc, clear all, randn('seed',123), rand('seed',456)
nd = 1000; % počet dat pro simulaci a odhad
mi = [1 3 5]; % střední hodnoty modelů
sg = [1 1 1]*.6; % směrodatné odchylky modelů
al = [.2 .7]; % pravděpodobnosti stavů: al(i)-al(i-1)
y = zeros(1,nd);
nk = length(al)+1;

% struktura pro statistiky a odhady parametrů
for i=1:nk,
    Md{i}.v = eye(2)*1e-5; % regularizace informační matice
    Md{i}.th = 0; % apriorní parametry jsou 0
end
```

```

% simulace a přepočítání informační matice
for t=1:nd
    k=genpt(al); % generátor aktivního stavu
    y(t) = mi(k)+sg(k)*randn; % simulovaná data
    fi = [y(t) 1]; % regresní vektor
    Md{k}.v = Md{k}.v+fi'*fi; % přepočítání informační matice
end

% výpočet odhadů středních hodnot a disperzí modelů
for i=1:nk
    Md{i}.th = Md{i}.v(1,2)/Md{i}.v(2,2);
    Md{i}.sg2 = (Md{i}.v(1,1)-Md{i}.v(1,2)^2/Md{i}.v(2,2))/Md{i}.v(2,2);
end

% tisk hodnot odhadu parametrů a obrázků
disp('Odhadnuté parametry modelů pro jednotlivé stavy:')
disp('      stř. hod. - sm. odch.')
```

```

for i=1:nk
    fprintf('Stav %u', i)
    disp([Md{i}.th, sqrt(Md{i}.sg2)])
end
figure(1), hist(y, 100)
title('Histogram simulovaných dat')
```

```

function k=genpt(al)
% k=genpt(al)
% náhodný generátor stavu
% al - kumulativní pravděpodobnosti stavů
% k - ukazovátka na aktivní stav
%
n=length(al);
rr=rand;
k=n+1;
for i=1:n
    if rr<al(i), k=i; break, end
end
```

Poznámka: Program odhaduje střední hodnoty (tj. polohy) jednotlivých stavů, reprezentovaných jako kopečky ve výsledném histogramu. Kvalitu odhadu lze rychle (tj. bez ohledu na odhadnuté rozptyly modelů, tj. šířky kopečků) posoudit porovnáním poloh vrcholů kopečků a odhadnutých středních hodnot. V Další fázi posuzování kvality odhadu je možno porovnávat i odhadnuté směrodatné odchylky a šířky kopečků.

Doporučené experimenty

1. Změna počtu dat $nd = 1000$ a kontrola kvality odhadu pro malý počet dat.
2. Změna simulovaných středních hodnot modelů $mi = [1 \ 3 \ 5]$ a kontrola jejich odhadů $Md\{i\}.th$.

3. Změna simulovaných směrodatných odchylek modelů $sg = [1 \ 1 \ 1] * .6a$ kontrola jejich odhadů $Md\{i\} .sg2$.
4. Změna pravděpodobnosti aktivit jednotlivých stavů soustavy, tj. změna prvků vektoru $a1 = [.2 \ .7]$ – ten definuje pravděpodobnosti aktivit jako 0.2, $0.7-0.2=0.5$ a $1-0.7=0.3$. Změny by se měly projevit např. horšími odhady pro málo pravděpodobné stavy.

10.7 Regresní model jako filtr

V základním kurzu statistiky se hovoří o regresní přímce, která je proložena body s lineárním trendem. Chceme-li předpovídat polohu dalších bodů, určíme je tak, aby ležely na této odhadnuté regresní přímce. Jiná úloha, podobná předpovědi, je filtrace. Při ní nepředpovídáme hodnoty dopředu, ale snažíme se odhadnout správnou okamžitou hodnotu dat, kterou jsme naměřili se šumem. To lze provést tak, že odhadneme model měřených dat a "správnou hodnotu" generujeme jako střední hodnotu z tohoto modelu (to odpovídá zmíněnému případu: odhad přímky a na ní předpovědi).

Pomocí metod bayesovského odhadování zkonstruujte datové filtry založené na těchto modelech

- a) $d_t = a + e_t$
- b) $d_t = bt + a + e_t$
- c) $d_t = dt^3 + ct^2 + bt + a + e_t$
- d) $d_t = bd_{t-1} + a + e_t$
- e) $d_t = \begin{cases} a_1 & \text{pro } j_t = 1 \\ a_2 & \text{pro } j_t = 2 \end{cases}$, kde j_t je měřená veličina.

ŘEŠENÍ

Datové vektory pro jednotlivé případy budou

- a) $\phi_t = [d_t, 1]$
- b) $\phi_t = [d_t, t, 1]$
- c) $\phi_t = [d_t, t^3, t^2, t, 1]$
- d) $\phi_t = [d_t, d_{t-1}, 1]$
- e) Zde budou dvě různé informační matice, pro každou hodnotu j jiná. Regresní vektor bude $\phi_t = [d_t, 1]$ a přepočte se s ním vždy příslušná informační matice.

Pro vyjádření regresního vektoru v případě e) lze také využít Kroneckerovu funkci $\delta(i, j)$. Model je možno zapsat takto

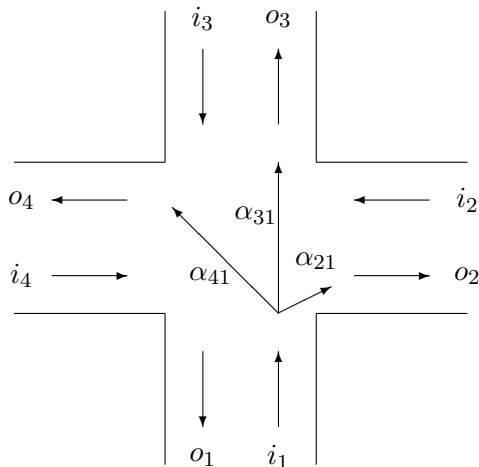
$$d_t = a_1\delta(1, j_t) + a_2\delta(2, j_t) + e_t.$$

Parametrický a regresní vektor takového modelu jsou

$$\Theta = [a_1, a_2]^T, \quad \varphi_t = [\delta(1, j_t), \delta(2, j_t)]^T.$$

10.8 Odhad směrových vztahů

Uvažujme křížovátku podle obrázku



kde symbol i značí vstupní intenzitu dopravního proudu do (in) křižovatky a o je výstupní intenzita ven (out) z křižovatky. Jednotlivým komunikacím, tvořícím křižovatku, jsou přiřazena čísla 1, 2, 3, 4.

Symbolem $\alpha_{k,m}$ značíme podíl intenzity, která odbočuje ze silnice m do silnice k . Předpokládáme, že v křižovatce se nesmí otáčet. Potom koeficienty $\alpha_{k,m}$ tvoří následující matici

$$\alpha = \begin{bmatrix} 0 & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & 0 & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & 0 & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & 0 \end{bmatrix}$$

Měříme intenzity $i_1, i_2, i_3, i_4, o_1, o_2, o_3, o_4$ v časových okamžicích $t = 1, 2, \dots, N$. Odhadněte směrové koeficienty α .

ŘEŠENÍ

Budeme uvažovat dvě různé situace.

a) Předpokládáme, že celý proud vozidel, který do křižovatky po jedné silnici vjede, z ní musí také okamžitě ostatními silnicemi vyjet. Tedy, v našem případě, určité procento odbočí doprava, další procento jede rovně a zbylé procento odbočí doleva. Protože v křižovatce nic nesmí zůstat, musí platit $\sum_{k \neq m} \alpha_{km} = 1$, pro $m = 1, 2, 3, 4$.

b) Jiný předpoklad je, že průjezd křižovatkou není okamžitý, a proto právě uvedený vztah pro směrové koeficienty platí jen přibližně.

Poznámka: Na těchto dvou předpokladech lze demonstrovat dva různé přístupy k odhadování s omezeními. V jednom se omezení respektují přesně a tím se redukuje počet odhadovaných parametrů, v druhém se odhadují všechny parametry, ale pomocí metody fiktivních dat je odhadování "tlačeno" k tomu, aby vazby alespoň přibližně respektovala. Oba dva přístupy lze využít také při odhadu spline-modelů.

Podle přístupu a) pro jednotlivé výstupní intenzity platí

$$o_k = \sum_{m \neq k} \alpha_{k,m} i_m + e_k,$$

kde e_k je šum modelu pro k -tou výstupní intenzitu. Podmínky kontinuity dopravního proudu dávají např. následující vazby mezi koeficienty

$$\alpha_{12} = 1 - \alpha_{13} - \alpha_{14}, \quad \alpha_{23} = 1 - \alpha_{24} - \alpha_{21}, \quad \alpha_{34} = 1 - \alpha_{31} - \alpha_{32}, \quad \alpha_{41} = 1 - \alpha_{42} - \alpha_{43}.$$

Dosadíme-li tyto podmínky do modelu, dostaneme, po roznásobení a úpravě, jeho jednotlivé části ve tvaru

$$\begin{aligned} o_1 &= \alpha_{13}(i_3 - i_2) + \alpha_{14}(i_4 - i_2) + i_2 + e_1 \\ o_2 &= \alpha_{21}(i_1 - i_3) + \alpha_{24}(i_4 - i_3) + i_3 + e_2 \\ o_3 &= \alpha_{31}(i_1 - i_4) + \alpha_{32}(i_2 - i_4) + i_4 + e_3 \\ o_4 &= \alpha_{42}(i_2 - i_1) + \alpha_{43}(i_3 - i_1) + i_1 + e_4. \end{aligned}$$

Zavedeme-li společný vektor parametrů

$$\Theta = [\alpha_{12}, \alpha_{14}, \alpha_{21}, \alpha_{24}, \alpha_{31}, \alpha_{32}, \alpha_{42}, \alpha_{43}]^T,$$

budeme k jejich odhadu v každém časovém okamžiku používat tyto modelované veličiny y_k a regresní vektory φ_k

$$\begin{aligned} y_1 = o_1 - i_2 & \quad \text{a} \quad \varphi_1 = [i_3 - i_2, i_4 - i_2, 0, 0, 0, 0, 0, 0]^T, \\ y_2 = o_2 - i_3 & \quad \text{a} \quad \varphi_2 = [0, 0, i_1 - i_3, i_4 - i_3, 0, 0, 0, 0]^T, \\ y_3 = o_3 - i_4 & \quad \text{a} \quad \varphi_3 = [0, 0, 0, 0, i_1 - i_4, i_2 - i_4, 0, 0]^T, \\ y_4 = o_4 - i_1 & \quad \text{a} \quad \varphi_4 = [0, 0, 0, 0, 0, 0, i_2 - i_1, i_3 - i_1]^T. \end{aligned}$$

Podle přístupu b) uvažujeme všechny koeficienty jako nezávislé a vazby mezi nimi "vnutíme" pomocí fiktivních dat. Společný vektor parametrů bude celá matice α (bez nul na diagonále) a narovnaná do vektoru

$$\Theta = [\alpha_{12}, \alpha_{13}, \alpha_{14}, \alpha_{21}, \alpha_{23}, \alpha_{24}, \alpha_{31}, \alpha_{32}, \alpha_{34}, \alpha_{41}, \alpha_{42}, \alpha_{43}].$$

Odpovídající modelované veličiny y_k a regresní vektory φ_k budou

$$\begin{aligned} y_1 = o_1 & \quad \text{a} \quad \varphi_1 = [i_2, i_3, i_4, 0, 0, 0, 0, 0, 0, 0, 0, 0]^T, \\ y_2 = o_2 & \quad \text{a} \quad \varphi_2 = [0, 0, 0, i_1, i_3, i_4, 0, 0, 0, 0, 0, 0]^T, \\ y_3 = o_3 & \quad \text{a} \quad \varphi_3 = [0, 0, 0, 0, 0, 0, i_1, i_2, i_4, 0, 0, 0]^T, \\ y_4 = o_4 & \quad \text{a} \quad \varphi_4 = [0, 0, 0, 0, 0, 0, 0, 0, 0, i_2, i_3, i_4]^T, \end{aligned}$$

Tyto rovnice však nerespektují vazby mezi koeficienty. Jestliže v těchto vazbách připustíme určitou volnost, můžeme je s pomocí šumu e zapsat takto

$$\begin{aligned} 1 &= \alpha_{21} + \alpha_{31} + \alpha_{41} + e_1, \\ 1 &= \alpha_{12} + \alpha_{42} + \alpha_{32} + e_2, \\ 1 &= \alpha_{13} + \alpha_{23} + \alpha_{43} + e_3, \\ 1 &= \alpha_{14} + \alpha_{24} + \alpha_{34} + e_4. \end{aligned}$$

Tyto vazby přidáme jako další (fiktivní) modely s modelovanými veličinami y_k a regresními vektory φ_k

$$\begin{aligned} y_1 = 1 & \quad \text{a} \quad \varphi_1 = [0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0]^T, \\ y_2 = 1 & \quad \text{a} \quad \varphi_2 = [1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0]^T, \\ y_3 = 1 & \quad \text{a} \quad \varphi_3 = [0, 1, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1]^T, \\ y_4 = 1 & \quad \text{a} \quad \varphi_4 = [0, 0, 1, 0, 0, 1, 0, 0, 1, 0, 0, 0]^T, \end{aligned}$$

Tyto rovnice přidáme do informační matice násobené vahou w , která vyjadřuje počet, kolikrát byla tato fiktivní podmínka "naměřena". V porovnání s počtem skutečných dat přidanych do informační matice se potom tyto podmínky více nebo méně prosadí.

Pro odhad použijeme algoritmus uvedený na str. 74. Oba způsoby odhadu směrových vztahů jsou doloženy příslušnými programy v jazyce MATLAB:

Odhad s přesnými podmínkami na dělení dopravního proudu

```
% Odhadování koeficientů směrových vztahů
% s přesnými podmínkami na dělení dopravního proudu
% -----
clc, clear all, randn('seed',123), rand('seed',456)
a=[0 .1 .2 .6;          % simulované směrové koeficienty
   .2 0 .4 .2;
   .5 .5 0 .2;
   .3 .4 .4 0];
nd=1000;                % počet dat pro simulaci a odhad
v=eye(9)*.001;         % počáteční informační matice

for t=1:nd
    % simulace vstupních a výstupních intenzit
    i=3*rand(4,1)+10;    % vstupní intenzity
    o=a*i+.01*randn(4,1); % výstupní intenzity

    % datové vektory pro odhad
    d=[o(1)-i(2) i(3) i(4) 0 0 0 -i(2) -i(2) 0;
        o(2)-i(3) -i(3) 0 i(1) i(4) 0 0 0 -i(3);
        o(3)-i(4) 0 -i(4) 0 -i(4) i(1) i(2) 0 0;
        o(4)-i(1) 0 0 -i(1) 0 -i(1) 0 i(2) i(3)];
    v=v+d'*d;           % přepočítání informační matice
end

vd=v(1,1);             %\
vdf=v(2:9,1);          % rozdělání informační matice
vf=v(2:9,2:9);         %/
alf=inv(vf)*vdf;       % výpočet odhadu parametrů

% konstrukce matice směrových koeficientů
disp('Odhad matice směrových vztahů')
al=[0 1-alf(6)-alf(7) alf(1:2)';
    alf(3) 0 1-alf(8)-alf(1) alf(4);
    alf(5:6)' 0 1-alf(2)-alf(4);
    1-alf(3)-alf(5) alf(7:8)' 0]
```

Odhad s volnými podmínkami na dělení dopravního proudu

```
% Odhadování koeficientů směrových vztahů
% s volnými podmínkami na dělení dopravního proudu
% -----
```

```

clc, clear all, randn('seed',123), rand('seed',456)
a=[0 .3 .1 .3;           % simulované směrové koeficienty
   .2 0 .5 .4;
   .2 .5 0 .3;
   .6 .2 .4 0];
nd= 1000;                % počet dat pro simulaci a odhad
w = 1000;                % váhy pro fiktivní modely
v = eye(13)*.001;       % počáteční informační matice

for t=1:nd
    % simulace vstupních a výstupních intenzit
    i=rand(4,1)+10;      % vstupní intenzity
    o=a*i+2*randn(4,1); % výstupní intenzity

    % datové vektory pro odhad + přepočítání informační matice
    d=[o(1) i(2) i(3) i(4) 0 0 0 0 0 0 0 0 0]; v=v+d'*d;
    d=[o(2) 0 0 0 i(1) i(3) i(4) 0 0 0 0 0 0]; v=v+d'*d;
    d=[o(3) 0 0 0 0 0 0 i(1) i(2) i(4) 0 0 0]; v=v+d'*d;
    d=[o(4) 0 0 0 0 0 0 0 0 0 i(1) i(2) i(3)]; v=v+d'*d;

    % fiktivní datové vektory + přepočítání informační matice
    d=[1 0 0 0 1 0 0 1 0 0 1 0 0]*w; v=v+d'*d;
    d=[1 1 0 0 0 0 0 0 1 0 0 1 0]*w; v=v+d'*d;
    d=[1 0 1 0 0 1 0 0 0 0 0 0 1]*w; v=v+d'*d;
    d=[1 0 0 1 0 0 1 0 0 1 0 0 0]*w; v=v+d'*d;
end

vd=v(1,1);              %\
vdf=v(2:13,1);          % rozdělení informační matice
vf=v(2:13,2:13);        %/
alf=inv(vf)*vdf;        % výpočet odhadu parametrů

% konstrukce matice směrových koeficientů a její tisk
disp('Odhad matice směrových vztahů')
al=[[0 alf(1:3)'];
    [alf(4) 0 alf(5:6)'];
    [alf(7:8)' 0 alf(9)];
    [alf(10:12)' 0]]
disp('Kontrola součtu sloupců matice směrových vztahů')
suma=sum(al)            % kontrola podmínek dělení dopravního proudu

```

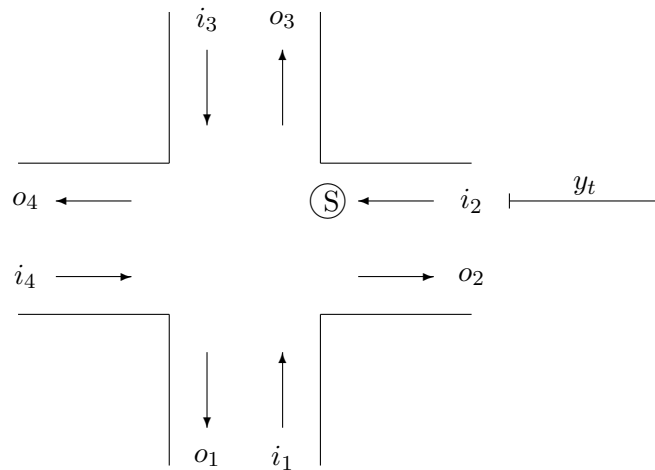
Doporučené experimenty

1. Změna počtu dat $nd=1000$ a kontrola přesnosti odhadu matice směrových vztahů vzhledem k simulované.
2. Definice vlastní matice směrových vztahů (matice a) a kontrola jejího odhadu. Zkuste nastavit speciální typ křižovatky, např. některé ulice jednosměrné!
3. U druhé verze programu zkuste měnit váhy $w=1000$ pro fiktivní modely, které zajišťují rovnici kontinuity pro křižovatku (tedy co do křižovatky vjede, musí také vyjet). Malé

hodnoty vah modelují velkou kapacitu křižovatky (auta v ní mohou zůstat a nebo samovolně vyjždět).

10.9 Spojité řízení

Uvažujme křižovatku podle následujícího obrázku



Zajímá nás silnice, která do křižovatky vstupuje zprava. Tady je světelná signalizace (označená S v kroužku) a díky ní se na této silnici tvoří kolona vozidel, jejíž délku v čase t jsme označili y_t . Naším cílem je řídit délku červené u_t (při pevné délce cyklu) tak, aby provoz v křižovatce byl optimální. Optimalitu definujeme takto

- Chceme, aby si kolona v křižovatce udržovala stálou délku w .
- Přejeme si udržet kompromis mezi délkou kolony a blokováním příčného směru, tj. na jedné straně chceme co nejkratší kolonu, na druhé straně nechceme příliš dlouhou zelenou, protože ta blokuje příčný směr.

ŘEŠENÍ

V obou případech budeme délku kolony modelovat řízeným regresním modelem prvního řádu – viz (4.8) s $\nu = 1$

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

kde okamžitou délku kolony y_t modelujeme v závislosti na délce kolony v minulém okamžiku y_{t-1} a nastavené délce zelené u_t .

V případě a) uvažujeme kritérium (7.2) s nulovou penalizací vstupu ω

$$J_a = \sum_{t=1}^N (y_t - w)^2 \rightarrow \min,$$

které udržuje délku kolony na předepsané velikosti w .

V případě b) je vhodné rovněž kritérium (7.2), ale s nulovou žádanou hodnotou w

$$J_a = \sum_{t=1}^N y_t^2 + \omega u_t^2 \rightarrow \min,$$

kteří hledá kompromis mezi délkou kolony a blokadou příčného směru. Penalizace řídicí veličiny pomocí $\omega \geq 0$ udává, v jakém poměru mají být protichůdné požadavky, minimální délky kolony a maximální průjezdnosti v příčném směru, v kritériu uplatňovány.

Řešení obou úloh provedeme postupným středováním a minimalizací podle algoritmu řízení ze str. 97

$$E_t = E [J_t + J_{t+1}^* | u_t, D(t-1)] = \int [J_t + J_{t+1}^*] f(y_t | u_t, D(t-1)) dy_t,$$

kde

J_t značí část kritéria, příslušející časovému okamžiku t a

J_{t+1}^* označuje minimální hodnotu kritéria z minulého kroku minimalizace.

Tuto funkci počítáme pro $t = N, N-1, \dots, 1$, s počáteční podmínkou $J_{N+1}^* = 0$. V každém kroku určíme hodnotu u_t , která kritérium minimalizuje. To je hodnota optimálního řízení pro daný časový okamžik.

V případě a) je

$$E_T = E [(y_T - w)^2 + J_{T+1}^*] = (ay_{T-1} + bu_T - w)^2$$

Minima je dosaženo pro $u_T = (w - ay_{T-1})/b$ a jeho hodnota je nula. Další kroky budou tedy stejné. Dostáváme jednokrokové řízení s řídicí veličinou

$$u_t = \frac{w - ay_{t-1}}{b}, \quad \text{pro } t = 1, 2, \dots, N.$$

V případě b) dostáváme pro obecný čas t

$$E_t = E [(y_t - w)^2 + \omega u_t^2 + J_{t+1}^* | u_t, D(t-1)] = (ay_{t-1} + bu_t - w)^2 + \omega u_t^2 + J_{t+1}^*.$$

Takovýmto typem kritéria jsme se již zabývali – viz program pro optimální řízení soustavy 1. řádu na str. 31.

10.10 Řízení a odhad s diskrétním modelem

Řízení

Budeme uvažovat křižovatku se světelnou signalizací. Sledujeme méně významnou vedlejší příjezdovou komunikaci, na které se tvoří kolona vozidel. Modelovaná veličina y_t je délka této kolony, přičemž rozlišujeme pouze tři stavy kolony: malá ($y_t = 1$), střední ($y_t = 2$) a velká ($y_t = 3$). Akční veličinou u_t , ovlivňující délku kolony, je: světelná signalizace zapnuta ($u_t = 1$) a světelná signalizace vypnuta ($u_t = 2$).

Pro konstrukci modelu pro délku kolony vezmeme v úvahu následující skutečnosti:

Kolona nemůže najednou celá zmizet nebo nebo vzniknout. Její délka může jen narůstat nebo se zmenšovat. Její okamžitý stav bude tedy závislý na stavu předchozím.

Zapnuté řízení křižovatky bude mít tendenci kolonu zkracovat, vypnuté (protože jsme na vedlejší silnici) bude naopak kolonu prodlužovat. Stav řízení v minulém okamžiku bude mít na okamžitou kolonu také určitý vliv, ten však pro jednoduchost zanedbáme a zahrneme jej do informace získané z minulé délky kolony.

Zapnuté řízení jednak zkracuje kolonu na vedlejší silnici, ale také dává vznik koloně na hlavní silnici. Proto se zapnutému řízení do jisté míry bráníme.

Model délky kolony bude následující – viz (10.1) a (4.20)

$$f(y_t|y_{t-1}, u_t) = P(y_t = i|y_{t-1} = j, u_t = k) = \Theta_{i|j,k}, \quad (10.4)$$

kde

Θ je pole známých parametrů modelu,

konkrétní parametr $\Theta_{i|j,k}$ je prvek třírozměrného pole s indexy i, j, k (první a druhé dva indexy jsme oddělili svislou čárou jen formálně) s významem pravděpodobnosti, že kolona bude mít délku i , když před tím měla délku j a hodnota řízení je k ,

všechny takové parametry tvoří třírozměrnou tabulku (pole, tenzor) s rozměry $(3, 3, 2)$.

Tento model budeme nyní konstruovat z "našich expertních znalostí". Jeho konstrukce spočívá v tom, že pro všechny možné kombinace hodnot $y_t = i$, $y_{t-1} = j$ a $u_t = k$ budeme definovat jejich pravděpodobnost $\Theta_{i|j,k}$. Konstrukce je uvedena v následující tabulce.

Ve stejné tabulce jsou také vyznačeny hodnoty penalizace $\Omega_{i|j,k}$ pro jednotlivé hodnoty y_t , y_{t-1} a u_t . Mohli bychom samozřejmě kritérium definovat v běžné kvadratické formě $y_t^2 + \omega u_t^2$. Uvedený způsob definice kritéria je ale obecnější a konečnost různých stavů (kombinací y_t , y_{t-1} a u_t) nám tento "komfort" dovolí.

č.	$y_t = i$	$y_{t-1} = j$	$u_t = k$	$\Theta_{i j,k}$	$\Omega_{i j,k}$
1.	1	1	1	0.7	1
2.	1	1	2	0.4	0
3.	1	2	1	0.5	0
4.	1	2	2	0.1	0
5.	1	3	1	0.5	0
6.	1	3	2	0.0	0
7.	2	1	1	0.2	3
8.	2	1	2	0.4	2
9.	2	2	1	0.4	2
10.	2	2	2	0.2	2
11.	2	3	1	0.5	1
12.	2	3	2	0.2	1
13.	3	1	1	0.1	5
14.	3	1	2	0.2	4
15.	3	2	1	0.1	3
16.	3	2	2	0.7	4
17.	3	3	1	0.0	5
18.	3	3	2	0.8	5

TABULKA: Definice modelu a kritéria řízení.

První řádek tabulky říká: jestliže minule byla kolona malá a řízení bylo zapnuto, pak pravděpodobnost, že kolona zůstane malá je 0.7. Tento stav penalizujeme 1, protože nebylo třeba řízení zapínat a tím blokovat hlavní směr v křižovatce.

Řádky 2. až 6. představují přechod na malou kolonu podmíněné ostatními kombinacemi předchozí kolony a stavu řízení. Mají své pravděpodobnosti a nejsou penalizovány – výsledek je malá kolona, což se požaduje.

Dále jsou uvedeny pravděpodobnosti a penalizace pro výslednou střední a velkou kolonu v závislosti na předchozí koloně a řízení. Penalizace jsou vyšší, protože nebylo dosaženo hlavního cíle – malé kolony. Bere se rovněž ohled na zapnuté řízení.

Poznámka: Protože z jednoho výchozího stavu (např. $y_{t-1} = 1$ a $u_t = 1$ – viz 1., 7. a 13. řádek tabulky) jsou jen tři možné výsledky ($y_t = 1$, $y_t = 2$ nebo $y_t = 3$), musí být součet jejich pravděpodobností roven jedné. V naší tabulce je to 0.7, 0.2 a 0.1. Podobně je tomu i pro ostatní výchozí stavy.

Poznámka: Koeficienty $\Theta_{i|j,k}$ lze odhadovat – viz pokračování příkladu na str. 162.

Úkolem této části příkladu je navrhnout optimální řízení pro horizont N tak, aby bylo minimalizováno kritérium

$$J = \sum_{t=1}^N \Omega_{y_t|y_{t-1}, u_t}.$$

ŘEŠENÍ

Řešení provedeme pomocí dynamického programování – viz str. 105, tj. postupným středováním a minimalizací kritéria od konce intervalu řízení. V každém kroku budeme počítat střední hodnotu

$$J_t = E[\Omega_t + J_{t+1}^* | u_t, D(t-1)] \quad (10.5)$$

a provádět minimalizaci

$$J_t^* = \min_{u_t} J_t, \quad (10.6)$$

s počáteční podmínkou $J_{t+1}^* = 0$,

kde

symbolem * jsme označili minimalizovanou část kritéria a

Ω_t označuje penalizaci v čase t .

Minimalizaci tedy začneme od $t = N$.

$$\begin{aligned} J_N &= E[\Omega_N + J_{N+1}^* | u_N, D(N-1)] = \sum_{i=1}^3 [\Omega_{i|j,k} + 0] \Theta_{i|j,k} = \\ &= \begin{bmatrix} \sum_{i=1}^3 \Omega_{i|1,1} \Theta_{i|1,1}, & \sum_{i=1}^3 \Omega_{i|1,2} \Theta_{i|1,2} \\ \sum_{i=1}^3 \Omega_{i|2,1} \Theta_{i|2,1}, & \sum_{i=1}^3 \Omega_{i|2,2} \Theta_{i|2,2} \\ \sum_{i=1}^3 \Omega_{i|3,1} \Theta_{i|3,1}, & \sum_{i=1}^3 \Omega_{i|3,2} \Theta_{i|3,2} \end{bmatrix} = \begin{bmatrix} J_{N;1,1} & J_{N;1,2} \\ J_{N;2,1} & J_{N;2,2} \\ J_{N;3,1} & J_{N;3,2} \end{bmatrix}. \end{aligned}$$

Posledním vyjádřením jsme chtěli ukázat, že střední hodnota kritéria J_N v čase N je matice čísel, kde řádky odpovídají hodnotám u_N a sloupce hodnotám y_{N-1} .

Dalším krokem ve stejném časovém okamžiku je minimalizace. V každém řádku vybereme minimum a zapamatujeme hodnotu u_N , pro kterou bylo minima dosaženo. Dostaneme tak tříprvkový sloupec – první prvek odpovídá hodnotě $y_{N-1} = 1$ a označuje hodnotu optimálního řízení. Druhý, resp. třetí prvek jsou hodnoty optimálního řízení pro $y_{N-1} = 2$, resp. $y_{N-1} = 3$. Dostáváme tak optimální řízení jako funkci y_{N-1} . Tento vektor je řídicím zákonem pro čas $t = N$. Až bude naměřena hodnoty y_{N-1} , tj. v okamžiku, kdy budeme řídit na hodnotu y_N , vybereme příslušnou hodnotu optimálního řízení a použijeme ji. Zatím jsme dostali pouze předpis

$$u_N^* = \begin{bmatrix} u_{N;1}^* \\ u_{N;2}^* \\ u_{N;3}^* \end{bmatrix}, \quad \text{kterému odpovídá} \quad J_N^* = \begin{bmatrix} J_{N;1}^* \\ J_{N;2}^* \\ J_{N;3}^* \end{bmatrix},$$

kde u^* označuje optimální řízení.

Pro konkrétní hodnoty z tabulky dostaneme střední hodnotu, řídicí zákon a hodnoty kritéria po minimalizaci

$$J_N = \begin{bmatrix} 1.8 & 1.6 \\ 1.1 & 3.2 \\ 0.5 & 4.2 \end{bmatrix}, \quad u_N^* = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad J_N^* = \begin{bmatrix} 1.6 \\ 1.1 \\ 0.5 \end{bmatrix}.$$

V dalším kroku minimalizace, pro $t = N - 1$, bude středování

$$\begin{aligned} J_{N-1} &= E[\Omega_{N-1} + J_N^* | u_{N-1}, D(N-2)] = \sum_{i=1}^3 [\Omega_{i|j,k} + J_{N;i}^*] \Theta_{i|j,k} = \\ &= \begin{bmatrix} \sum_{i=1}^3 (\Omega_{i|1,1} + J_{N;i}^*) \Theta_{i|1,1}, & \sum_{i=1}^3 (\Omega_{i|1,2} + J_{N;i}^*) \Theta_{i|1,2} \\ \sum_{i=1}^3 (\Omega_{i|2,1} + J_{N;i}^*) \Theta_{i|2,1}, & \sum_{i=1}^3 (\Omega_{i|2,2} + J_{N;i}^*) \Theta_{i|2,2} \\ \sum_{i=1}^3 (\Omega_{i|3,1} + J_{N;i}^*) \Theta_{i|3,1}, & \sum_{i=1}^3 (\Omega_{i|3,2} + J_{N;i}^*) \Theta_{i|3,2} \end{bmatrix} = \begin{bmatrix} J_{N-1;1,1} & J_{N-1;1,2} \\ J_{N-1;2,1} & J_{N-1;2,2} \\ J_{N-1;3,1} & J_{N-1;3,2} \end{bmatrix}. \end{aligned}$$

Minimalizace opět vybere v každém řádku minimální prvek a zapamatuje hodnotu řídicí veličiny, pro kterou bylo tohoto minima dosaženo. Tak dostaneme

$$u_{N-1}^* = \begin{bmatrix} u_{N-1;1}^* \\ u_{N-1;2}^* \\ u_{N-1;3}^* \end{bmatrix} \quad \text{kterému odpovídá} \quad J_{N-1}^* = \begin{bmatrix} J_{N-1;1}^* \\ J_{N-1;2}^* \\ J_{N-1;3}^* \end{bmatrix},$$

a pro konkrétní hodnoty z tabulky

$$J_N = \begin{bmatrix} 3.19 & 2.78 \\ 2.39 & 3.93 \\ 1.85 & 4.82 \end{bmatrix}, \quad u_N^* = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \quad J_N^* = \begin{bmatrix} 2.78 \\ 2.39 \\ 1.85 \end{bmatrix}.$$

Tyto dva kroky (středování a minimalizace) se ve stejné formě dále opakují pro $t = N - 2, N - 3, \dots, 1$. V okamžiku, kdy určíme řídicí zákon pro čas 1, který bude funkcí y_0 – které známe, jsme schopni skutečně realizovat první akční zásah u_1 a změřit první modelovanou (řízenou) veličinu y_1 . S ní lze opět určit další akční zásah u_2 , atd.

Celý návrh i realizace optimálního řízení je uveden v následujícím programu.

```
% Řízení s diskretním modelem f(y_t|y_(t-1),u_t)
% -----
% y=1 (kolona malá), 2 (kolona střední), 3 (kolona velká);
% u=1 (řízení zapnuto), 2 (řízení vypnuto)
%
% fy      : parametry modelu Theta{i}(j,k)
% om      : kritérium omega{i}(j,k)
% N       : počet kroků řízení (horizont)
% E       : vystředovaná část kritéria
% Em      : minimalizovaná část kritéria
% yy,uu   : okamžité hodnoty výstupu a vstupu
% y,u     : průběhy výstupu a vstupu
%
clc, clear all, rand('seed',123);
```



```

% Zadání hodnot Theta a omega
fy{1}=[.7 .4;.5 .1;.5 0]; % parametry pro malou kolonu
[ny nu]=size(fy{1}); % počet úrovní veličiny y a řízení u
fy{2}=[.2 .4;.4 .2;.5 .2]; % parametry pro střední kolonu
fy{3}=ones(ny,nu)-fy{1}-fy{2}; % parametry pro velkou kolonu
if sum(sum(fy{3}<0))>0,
    disp(' Chybné hodnoty parametrů modelu'),
end
om={[1 0;0 0;0 0] [3 2;2 2;1 1] [5 4;3 4;5 5]};
% hodnoty penalizace omega
ny=length(fy); % počet hodnot y
nu=size(fy{1},2); % počet hodnot u
N=20; % horizont řízení
ur=[];
Em=zeros(1,ny);

% Výpočet řízení na horizontě N
for i=1:N
    E=zeros(ny,nu); % start s nulovou počáteční podmínkou

    % středování
    for k=1:ny, E=E+(om{k}+ones(ny,nu)*Em(k)).*fy{k}; end

    % minimalizace
    for k=1:ny, [Em(k) ut(k)]=min(E(k,:)); end,
    ur=[ut; ur]; % hodnoty řízení (od poslední k první)
end

% Simulace a aplikace řízení
yy=2; % počáteční hodnota y(0)
y=yy; u=[]; kk=0;
for t=1:N
    % výpočet optimálního řízení
    ut=ur(t,:); % výběr řízení pro čas t
    uu=ut(yy); % optimální řízení u(t)=fce(y(t-1))
    %uu=1; % manuální řízení (srovnání s optimálním)
    u=[u uu];
    y1=yy;

    % simulace diskrétního modelu
    r=rand;
    if r<fy{1}(yy,uu), yy=1;
    elseif r<(fy{1}(yy,uu)+fy{2}(yy,uu)), yy=2;
    else yy=3;
    end
    kk=kk+om{yy}(y1,uu); % skutečná hodnota kritéria
    y=[y yy];
end

```

```
% Závěry
disp('Dosažená hodnota kriteria po minimalizaci')
kk
figure(1),plot(0:N,y(1:N+1),'o',1:N,u,'x')
title('Časový graf hodnot kolony (o) a řízení (x)')
```

Doporučené experimenty

1. V příkladu je především možno nastavit vlastní model $f(y_t|y_{t-1}, u_t)$ křížovatky změnou hodnot jeho parametrů $\text{fy}\{1\}$, $\text{fy}\{2\}$ a $\text{fy}\{3\}$. Vymyslete vlastní závislosti délky kolony na její minulé délce a řízení a zadejte je pomocí těchto parametrů.
2. Další experiment se týká změny penalizační matice om , která vyjadřuje cíle, jež chceme řízením dosáhnout. Vyslovte vlastní požadavky na řízení a zadejte je pomocí nové penalizační matice. (Současné kritérium vyjadřuje požadavek na co nejkratší kolonu, ale s ohledem na přílišnou blokádu kolmé hlavní silnice).
3. V místě generování optimálního řízení $\text{uu}=\text{ut}(\text{yy})$ je připraveno manuální řízení $\% \text{uu}=1$. Když odstraníme znak $\%$ bude aktivní a výsledek lze srovnat s kvalitou optimálního řízení podle hodnoty minimalizovaného kriteria kk . Porovnejte pro obě možné hodnoty řízení.

Odhadování

Proveďte odhad parametrů Θ modelu (10.4).

ŘEŠENÍ

Budeme postupovat podle příkladu 5.4.4. Pomocí součinu Kroneckerových funkcí δ (5.65) vyjádříme model (10.4) v součinném tvaru

$$f(y_t|y_{t-1}, u_t) = \prod_{i,j,k} \Theta_{i|j,k}^{\delta(i,y_t)\delta(j,y_{t-1})\delta(k,u_t)}, \quad (10.7)$$

kde součin se provádí přes všechny kombinace indexů i, j, k , tj. $i, j = 1, 2, 3$; $k = 1, 2$.

Apriorní hp zvolíme v konjugovaném tvaru (5.68)

$$f(\Theta|D(0)) = \prod_{i,j,k} \Theta_{i|j,k}^{\kappa_{i|j,k}(0)}, \quad (10.8)$$

kde $\kappa_{i|j,k}(0)$ je apriorní statistika příslušející stavu $y_0 = i$, $y(-1) = j$ a $u_0 = k$.

Bayesův vzorec (5.18) $f(\Theta|D(t)) \propto f(y_t|y_{t-1}, u_t, \Theta)f(\Theta|D(t-1))$ má tvar

$$\prod_{i,j,k} \Theta_{i|j,k}^{\kappa_{i|j,k}(t)} \propto \prod_{i,j,k} \Theta_{i|j,k}^{\delta(i,y_t)\delta(j,y_{t-1})\delta(k,u_t)} \Theta_{i|j,k}^{\kappa_{i|j,k}(t-1)}, \quad (10.9)$$

odkud plyne rekurze pro přepočet statistik, srv (5.70)

$$\kappa_{y_t|y_{t-1}, u_t}(t) = \kappa_{y_t|y_{t-1}, u_t}(t-1) + 1, \quad t = 1, 2, \dots, N. \quad (10.10)$$

Ostatní statistiky zůstávají v každém časovém okamžiku t bezezměny. Rekurze začíná s apriorními statistikami $\kappa_{i|j,k}(0)$ pro $i, j = 1, 2, 3$, $k = 1, 2$.

Z vypočtené aposteriorní hp pro $t = 1, 2, \dots, N$ určíme bodové odhady parametrů jako podmíněné střední hodnoty – viz (5.11)

$$\hat{\Theta}_{i|j,k} = E[\Theta_{i|j,k}|D(N)] = \int_0^1 \Theta_{i|j,k} f(\Theta|D(N)) d\Theta_{i|j,k}, \quad \forall i, j, k.$$

Podobně jako v (5.71) lze pro bodové odhady parametrů odvodit vztah, který lze očekávat

$$\hat{\Theta}_{q|j,k} = \frac{\Theta_{q|j,k}}{\sum_{i=1}^3 \Theta_{i|j,k}}, \quad (10.11)$$

pro $q = 1, 2, 3$ a všechny kombinace indexů $j = 1, 2, 3$; $k = 1, 2$.

Uvedené odhadování je realizováno v následujícím programu

```
% Odhad s diskretnim modelem f(y_t|y_(t-1),u_t)
% -----
% y=1 (kolona malá), 2 (kolona střední), 3 (kolona velká);
% u=1 (řízení zapnuto), 2 (řízení vypnuto)
%
% fy      parametry modelu Theta{i}(j,k)
% N      počet kroků řízení (horizont)
% s      statistika pro odhad parametrů
% Th     odhady parametrů modelu
% yy,uu  okamžité hodnoty výstupu a vstupu
% y,u    průběhy výstupu a vstupu
%
clc, clear all, rand('seed',123);

% Parametry pro simulaci
N=1000;                                % horizont simulace a odhadu
fy{1}=[.7 .4;.5 .1;.5 0];              % simulované parametry modelu
[ny nu]=size(fy{1});
fy{2}=[.2 .4;.4 .2;.5 .2];              % pro malou, střední
fy{3}=ones(ny,nu)-fy{1}-fy{2};          % a velkou kolonu
if sum(sum(fy{3}<0))>0,
    disp(' Chybné hodnoty parametrů modelu'),
end
uu=1; yy=1; y=[]; u=[];

% Simulační smyčka
for t=1:N
    if rand>.8, uu=3-uu; end              % generování vstupu
    r=rand;                               % generování výstupu
    if r<fy{1}(yy,uu),                    yy=1;
    elseif r<(fy{1}(yy,uu)+fy{2}(yy,uu)), yy=2;
    else                                   yy=3;
    end
    y=[y yy]; u=[u uu];
end
figure(1),plot(0:N,y,':o',1:N,u,':x') % případné zobrazení
```

```

% Odhad
for i=1:ny, s{i}=zeros(ny,nu); end
for t=1:N
    ys=y(t); yy=y(t+1); uu=u(t);

    % přepočítání statistiky
    s{yy}(ys,uu)=s{yy}(ys,uu)+1;
end
ss=zeros(size(s{1}));
for i=1:ny, ss=ss+s{i}; end

% Výpočet bodového odhadu parametrů
for i=1:ny, Th{i}=s{i}./ss; end
disp(' ')
disp('Porovnání simulovaných a odhadnutých parametrů')
fprintf(' pro %i kroků simulace\n',N)
disp(' ')
disp(' - simulované parametry')
disp([fy{1} fy{2} fy{3}])
disp(' ')
disp(' - odhadnuté parametry')
disp([Th{1} Th{2} Th{3}])

```

Doporučené experimenty

1. Podobně jako v první části tohoto příkladu o řízení lze experimentovat s vytvořením vlastního modelu křižovatky pomocí parametrů f_y a sledovat jejich odhady.
2. Je také možno měnit počet dat $N=1000$ pro odhad při sledování kvality zobrazených bodových odhadů ve srovnání s parametry simulovanými.

Kapitola 11

Dodatky

11.1 Doplnění na čtverec

Při minimalizaci kritéria (2.5) budeme často používat doplnění kvadratického dvojčlenu na úplný čtverec, a to jak pro skalární, tak i pro vektorové funkce. Připomeňme nyní základní myšlenku.

Pro skalární koeficienty je zřejmá

$$ax^2 - 2bx + c = a \left[x^2 - 2\frac{b}{a}x + \left(\frac{b}{a}\right)^2 - \left(\frac{b}{a}\right)^2 \right] + c = a \left(x - \frac{b}{a} \right)^2 + c - \frac{b^2}{a}.$$

Pro maticové koeficienty je úprava poněkud méně přehledná.

Nechť A je čtvercová, symetrická, pozitivně definitní matice n -tého stupně, nechť w, x jsou n -rozměrné sloupcové vektory a nechť c je skalár. Ukážeme, že platí rovnost

$$x^T Ax - x^T w - w^T x + c = (x - A^{-1}w)^T A(x - A^{-1}w) + c - w^T A^{-1}w. \quad (11.1)$$

Podle předpokladu je matice A symetrická, a tedy $A^T = A$, $(A^{-1})^T = A^{-1}$. Roznásobíme součin na pravé straně

$$\begin{aligned} (x - A^{-1}w)^T A(x - A^{-1}w) &= x^T Ax - x^T AA^{-1}w - w^T A^{-1}Ax + w^T A^{-1}AA^{-1}w = \\ &= x^T Ax - x^T w - w^T x + w^T A^{-1}w. \end{aligned}$$

Nyní přičteme k oběma stranám $c - w^T A^{-1}w$ a dostaneme dokazovanou rovnost (11.1).

Provedená úvaha ihned ukazuje, jak lze při úpravě na úplný čtverec postupovat.

Kvadratický člen $x^T Ax$ ponecháme a do lineárních členů vložíme jednotkovou matici $A^{-1}A = AA^{-1}$. Dostaneme

$$x^T Ax - x^T AA^{-1}w - w^T A^{-1}Ax + c.$$

Nyní stačí přičíst a odečíst $(A^{-1}w)^T A(A^{-1}w) = w^T A^{-1}AA^{-1}w = w^T A^{-1}w$, abychom dostali úplný čtverec, jak je na pravé straně rovnosti (11.1).

11.2 Operátor průměrování a jeho vlastnosti

Operátor průměrování je pro posloupnost $x = [x_1, x_2, \dots, x_N]$ definován předpisem

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i.$$

Vlastnosti operátoru průměrování: Necht' x, y jsou konečné posloupnosti, a konstanta. Pak platí:

Průměr násobku

$$\overline{ax} = \frac{1}{n} \sum_{i=1}^N ax_i = a \frac{1}{N} \sum_{i=1}^N x_i = a\bar{x}.$$

Průměr součtu

$$\overline{x+y} = \frac{1}{N} \sum_{i=1}^N (x_i + y_i) = \frac{1}{N} \sum_{i=1}^N x_i + \frac{1}{N} \sum_{i=1}^N y_i = \bar{x} + \bar{y}.$$

Průměr z konstanty

$$\bar{a} = \frac{1}{N} \sum_{i=1}^N a = \frac{1}{N} Na = a.$$

Průměr z průměru

$$\overline{\bar{x}} = \bar{x}.$$

Pro průměr součinu a kvadrátu obecně platí nerovnosti

$$\overline{xy} \neq \bar{x} \bar{y}, \quad \overline{x^2} \neq \bar{x}^2.$$

Průměr kvadrátů odchylek od průměru

$$\overline{(x - \bar{x})^2} = \overline{x^2 - 2x\bar{x} + \bar{x}^2} = \overline{x^2} - \overline{2x\bar{x}} + \overline{\bar{x}^2} = \overline{x^2} - \bar{x}^2.$$

Průměr součinů odchylek od průměrů

$$\overline{(x - \bar{x})(y - \bar{y})} = \overline{xy} - \bar{x} \bar{y}.$$

Příklad [Střední hodnota modelu pomocí operátoru průměrování]

Pro $N = 3$ a model $y_\tau = bu_\tau + c + e_\tau$ platí

$$\bar{y} = \frac{1}{3} \sum_{\tau=1}^3 y_i = \frac{1}{3} \sum_{\tau=1}^3 (bu_i + c + e_i) = b \frac{1}{3} \sum_{\tau=1}^3 u_i + \frac{1}{3} \sum_{\tau=1}^3 c + \frac{1}{3} \sum_{\tau=1}^3 e_i = b\bar{u} + c + \bar{e},$$

nebo totéž pomocí operátoru průměrování

$$\bar{y} = \overline{bu + c + e} = \overline{bu} + \bar{c} + \bar{e} = b\bar{u} + c + \bar{e}.$$

11.3 Funkce gama a beta

Pro výpočty s diskrétními modely potřebujeme funkci beta, kterou lze vyjádřit pomocí funkce gama. Zde budeme obě funkce definovat a uvedeme některé vztahy, které pro ně platí.

Funkce Gamma

Pro z reálné (komplexní) definujeme funkci gama jako integrál

$$\Gamma(z) = \int_0^{\infty} e^{-z} t^{z-1} dt. \quad (11.2)$$

Pro $z = n \in N$ platí $\Gamma(n+1) = n!$, proto se o funkci gama mluví často i jako o zobecněném faktoriálu.

Funkce Beta

Pro $t, x, y \in R$, $0 < t < 1$, $x > 1$, $y > 1$ definujeme funkci beta vztahem

$$B(x, y) = \int_0^1 t^{x-1} (1-t)^{y-1} dt \quad (11.3)$$

Pomocí funkce gama můžeme vyjádřit funkci beta ve tvaru

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \quad (11.4)$$

z něhož plyne důležitá vlastnost funkce beta

$$B(x+1, y) = \frac{x}{x+y} B(x, y) \quad (11.5)$$

Odvození tohoto vzorce ukážeme pro $x = m \in N$ a $y = n \in N$

$$B(m, n) = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} = \frac{(m-1)!(n-1)!}{(m+n-1)!},$$

$$B(m+1, n) = \frac{(m)!(n-1)!}{(m+n)!} = \frac{m(m-1)!(n-1)!}{(m+n)(m+n-1)!} = \frac{m}{m+n} B(m, n).$$

11.4 Jednorázový výpočet hp předpovědi

Ukážeme odvození vzorce pro jednorázový výpočet hp předpovědi, který jsme již bez odvození použili v (5.16) a (7.12). Vyjdeme ze vztahu (7.11) a aposteriorní hp podle (5.17) vyjádříme jako normovaný součin hp modelů v naměřených datech a apriorní hp ve tvaru

$$f(\Theta|D(t-1)) = \frac{[\prod_{\tau=1}^{t-1} f(y_{\tau}|\varphi_{\tau}, \Theta)] f(\Theta|D(0))}{\int [\prod_{\tau=1}^{t-1} f(y_{\tau}|\varphi_{\tau}, \vartheta)] f(\vartheta|D(0)) d\vartheta}, \quad (11.6)$$

kde ϑ je jen jiné označení pro parametry Θ .

Dosadíme-li do (7.11) můžeme vzorec pro hp předpovědi upravit do tvaru

$$\begin{aligned} f(y_t|u_t, D(t-1)) &= \int f(y_t|\varphi_t, \Theta) \frac{[\prod_{\tau=1}^{t-1} f(y_{\tau}|\varphi_{\tau}, \Theta)] f(\Theta|D(0))}{\int [\prod_{\tau=1}^{t-1} f(y_{\tau}|\varphi_{\tau}, \vartheta)] f(\vartheta|D(0)) d\vartheta} d\Theta = \\ &= \frac{\int [\prod_{\tau=1}^t f(y_{\tau}|\varphi_{\tau}, \Theta)] f(\Theta|D(0)) d\Theta}{\int [\prod_{\tau=1}^{t-1} f(y_{\tau}|\varphi_{\tau}, \vartheta)] f(\vartheta|D(0)) d\vartheta}. \end{aligned}$$

Integrál z věrohodnostní funkce násobené apriorní hp označíme I_t . Hp předpovědi lze potom psát jako podíl těchto integrálů v časech t a $t - 1$

$$f(y_t|u_t, D(t-1)) = \frac{I_t}{I_{t-1}}, \quad (11.7)$$

kde

$$I_t(\Theta) = \int \left[\prod_{\tau=1}^t f(y_\tau|\varphi_\tau, \Theta) \right] f(\Theta|D(0)) d\Theta = \int L_t(\Theta) f(\Theta|D(0)) d\Theta, \quad (11.8)$$

kde $L_t(\Theta)$ je věrohodnostní funkce (5.5) a $f(\Theta|D(0))$ je apriorní hp (5.6).

11.5 Bodová předpověď výstupu soustavy

Počítáme optimální bodovou předpověď \hat{y}_t výstupu soustavy, která minimalizuje kvadratické kritérium optimality

$$J_{\text{pre}} = E[(\hat{y}_t - y_t)^2]$$

a pro kterou platí, že závisí nejvýše na u_t a $D(t-1)$, tj. předpověď ve tvaru $\hat{y}_t = \hat{y}_t(u_t, D(t-1))$. Za těchto předpokladů bude

$$\begin{aligned} \min_{\hat{y}_t} J &= \min_{\hat{y}_t} \int (y_t - \hat{y}_t)^2 f(y_t, u_t, D(t-1)) d(y_t, u_t, D(t-1)) = \\ &= \min_{\hat{y}_t} \int \int (y_t - \hat{y}_t)^2 f(y_t|u_t, D(t-1)) f(u_t, D(t-1)) dy_t d(u_t, D(t-1)) = \\ &= \int \min_{\hat{y}_t} \left[\int (y_t - \hat{y}_t)^2 f(y_t|u_t, D(t-1)) dy_t \right] f(u_t, D(t-1)) d(u_t, D(t-1)) = \\ &= E \left[\min_{\hat{y}_t} E[(y_t - \hat{y}_t)^2 | u_t, D(t-1)] \right]. \end{aligned}$$

Úloha se tak převádí z minimalizace nepodmíněné střední hodnoty na minimalizaci podmíněné střední hodnoty. To je velice důležitý krok, neboť podmíněnou střední hodnotu jsme schopni počítat pomocí modelu. Nepodmíněnou střední hodnotu přímo počítat neumíme!

Minimalizaci výrazu $E[(y_t - \hat{y}_t)^2 | u_t, D(t-1)]$ provedeme doplněním na čtverec a anulováním kvadratického členu.

$$\begin{aligned} E[\hat{y}_t^2 - 2\hat{y}_t y_t + y_t^2 | u_t, D(t-1)] &= \\ &= \hat{y}_t^2 - 2\hat{y}_t E[y_t | u_t, D(t-1)] + E[y_t^2 | u_t, D(t-1)] = \\ &= \hat{y}_t^2 - 2\hat{y}_t E[y_t | u_t, D(t-1)] + (E[y_t | u_t, D(t-1)])^2 + \\ &\quad + E[y_t^2 | u_t, D(t-1)] - (E[y_t | u_t, D(t-1)])^2 = \\ &= (\hat{y}_t - E[y_t | u_t, D(t-1)])^2 + D[y_t | u_t, D(t-1)], \end{aligned}$$

kde $D[.]$ značí podmíněný rozptyl.

Minima kritéria je dosaženo pro $\hat{y}_t = E[y_t | u_t, D(t-1)]$, kdy je kvadratický člen anulován, a tedy minimální. Zbytek kritéria po minimalizaci, který již na argumentu minimalizace nezávisí, je $E[y_t^2 | u_t, D(t-1)] - E[y_t | u_t, D(t-1)]^2 = D[y_t | u_t, D(t-1)] = D[y_t]$.

11.6 Přímý přepoččet bodových odhadů

Odvození provedeme pro statický model $y_t = \theta u_t + e_t$. Odvození pro dynamický model je podobné, ale vyžaduje speciální maticové operace.

Přepoččet kovariance parametrů C_t : Vydeme ze vztahu $S_u^2(t) = S_u^2(t-1) + u_t^2$ (7.3), který plyne z přepočtu informační matice $V(t-1) \rightarrow V(t)$. Protože $C_t = \frac{1}{S_u^2(t)}$ (7.5), budeme se snažit uvedený přepoččet převést pro inverze funkcí S_u^2

$$C_t = \frac{1}{S_u^2(t)} = \frac{1}{S_u^2(t-1) + u_t^2}.$$

Přičteme a odečteme člen $\frac{1}{S_u^2(t-1)} = C_{t-1}$ a dostaneme

$$\begin{aligned} C_t &= C_{t-1} + \frac{1}{S_u^2(t-1) + u_t^2} - \frac{1}{S_u^2(t-1)} = C_{t-1} - \frac{u_t^2}{(S_u^2(t-1) + u_t^2)S_u^2(t-1)} = \\ &= C_{t-1} - \frac{1}{1 + \frac{u_t^2}{S_u^2(t-1)}} \left(\frac{u_t}{S_u^2(t-1)} \right)^2. \end{aligned}$$

Označme $\zeta_t = \frac{u_t^2}{S_u^2(t-1)}$. Pak $\frac{1}{S_u^2(t-1)} = C_{t-1} = \frac{\zeta_t}{u_t^2}$, a tedy

$$C_t = C_{t-1} - \frac{1}{1 + \zeta_t} (C_{t-1} u_t)^2, \quad (11.9)$$

což je požadovaný vztah pro C_t .

Poznámka: Ve skalárním vyjádření, kterého jsme se drželi, není tak dobře patrný význam uvedených vzorců vzhledem k prováděným inverzím. Zatímco při přímém přepočtu informační matice (algoritmus I) je třeba v každém kroku provádět výpočet kovarianční matice C pomocí inverze, jediné dělení, které provádíme zde, je dělení výrazem $1 + \zeta_t$, a to je vždy číslo. Pro regresní vektor φ_t ve tvaru sloupcového vektoru je $\zeta_t = \varphi_t^T C_{t-1} \varphi_t$, což je skalár.

Přepoččet bodových odhadů parametrů $\hat{\theta}_t$: Pro odhad parametrů platí

$$\hat{\theta}_t = \frac{S_{yu}(t)}{S_u^2(t)} = C_t S_{yu}(t).$$

Do posledního výrazu dosadíme za C_t z předchozí rekurze a $S_{yu}(t) = S_{yu}(t-1) + y_t u_t$. Dostaneme

$$\begin{aligned} \hat{\theta}_t &= \left[C_{t-1} - \frac{1}{1 + \zeta_t} (C_{t-1} u_t)^2 \right] [S_{yu}(t-1) + y_t u_t] = \\ &= C_{t-1} S_{yu}(t-1) + C_{t-1} y_t u_t - \frac{1}{1 + \zeta_t} [(C_{t-1} u_t)^2 S_{yu}(t-1) + (C_{t-1} u_t)^2 y_t u_t] = \\ &= \hat{\theta}_{t-1} + \frac{1}{1 + \zeta_t} [(1 + \zeta_t) C_{t-1} y_t u_t - (C_{t-1} u_t)^2 S_{yu}(t-1) - (C_{t-1} u_t)^2 y_t u_t], \end{aligned}$$

kde jako první člen jsme obdrželi minulé odhady parametrů, druhý člen jsme vsunuli do hranaté závorky. Jelikož je

$$\zeta_t C_{t-1} y_t u_t = C_{t-1}^2 u_t^2 y_t u_t,$$

ruší se dva sčítance v hranaté závorce. Dále vytkneme $C_{t-1}u_t$ a využijeme rovnost $C_{t-1}S_{yu}(t-1) = \hat{\theta}_{t-1}$. Dostaneme

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{1 + \zeta_t} [C_{t-1}y_t u_t - C_{t-1}^2 u_t^2 S_{yu}(t-1)] = \hat{\theta}_{t-1} + \frac{1}{1 + \zeta_t} C_{t-1} u_t (y_t - \hat{\theta}_{t-1} u_t), \quad (11.10)$$

což je hledaný vztah.

Přepočít zbytku kvadratické formy Λ_t : Vyjdeme ze vztahu pro Λ_t

$$\Lambda_t = S_y^2(t) - C_t [S_{yu}(t)]^2.$$

Za všechny výrazy na pravé straně dosadíme z rekurentních vzorců

$$\Lambda_t = S_y^2(t-1) + y_t^2 - \left(C_{t-1} - \frac{1}{1 + \zeta_t} (C_{t-1} u_t)^2 \right) (S_{yu}(t-1) + y_t u_t)^2.$$

Jelikož je $\zeta_t = C_{t-1} u_t^2$, prostřední závorka je

$$C_{t-1} - \frac{1}{1 + \zeta_t} (C_{t-1} u_t)^2 = C_{t-1} \left(1 - \frac{\zeta_t}{1 + \zeta_t} \right) = \frac{C_{t-1}}{1 + \zeta_t}.$$

Abychom dostali Λ_{t-1} odečteme a přičteme člen $\hat{\theta}_{t-1} = C_{t-1} S_{yu}(t-1)$. Přidaný člen spolu s y_t^2 přesuneme za zlomek a závorku za zlomkem umocníme. Některé členy za zlomkem se po úpravě vyruší a dostaneme

$$\Lambda_t = S_y^2(t-1) - C_{t-1} [S_{yu}(t-1)]^2 + \frac{1}{1 + \zeta_t} [y_t^2 - 2y_t \hat{\theta}_{t-1} u_t + (\hat{\theta}_{t-1} u_t)^2].$$

Odtud plyne dokazovaný vztah

$$\Lambda_t = \Lambda_{t-1} + \frac{1}{1 + \zeta_t} (y_t - \hat{\theta}_{t-1} u_t)^2. \quad (11.11)$$

Index

- řízení
 - důvěřivá strategie, 100
 - horizont, 20
 - kauzální, 21
 - kritérium, 20
 - minimalizace, 93
 - opatrná strategie, 102
 - středování, 93
- řetězové pravidlo, 47
- algoritmus
 - řízení, 31, 97
 - odhadování, 29, 32, 33, 35, 74
- Bayesův vzorec, 48, 111
 - aproximace, 111
- Bayesův vztah, 71
- distribuční funkce, 38
 - sdružená, 39
- doplnění na úplný čtverec, 20, 26, 29
- exponenciální třída modelů, 108
- funkce
 - Kroneckerova, 109
 - spline, 131
 - 1. stupně, 136
 - 3. stupně, 137
 - aproximační, 134
 - fundamentální, 132
 - interpolační, 134
 - zobecněná, 134
- hustota pravděpodobnosti, 39
 - aposteriorní, 66, 70
 - apriorní, 66
 - marginální, 39
 - předpovědi, 67, 70
 - podmíněná, 40
 - sdružená, 39
- indikátor, 43, 58
- kritérium
 - řízení, 94, 100, 102
- minimalizace
 - jednorázová, 20
 - postupná, 21
- minimum variance, 23
- odhady, 67
- předpověď, 67
- zbytek, 27
- matice
 - informační
 - apriorní, 28
 - rozšířená, 28
- metoda klouzavého horizontu, 99
- MixTools, 116
- model
 - řád, 18
 - deterministický, 11, 12
 - diskrétní, 61
 - dynamický, 56
 - komponenta multimodelu, 108
 - maticový, 27
 - mnohorozměrný, 60
 - parametry, 18, 53
 - pravděpodobnostní, 13
 - regresní, 18, 53, 54
 - regresní koeficienty, 18, 24, 53
 - regresní vektor, 18
 - s diskrétním šumem, 59
 - s normálním šumem, 56, 57
 - s rovnoměrným šumem, 58
 - součinný tvar, 108, 110
 - spline, 137
 - statický, 57–59
 - typy modelů, 55
- multimodel, 107
- náhodná veličina, 37
- normovaný argument, 134
- obdélníkové schéma, 50
- odhad
 - apriorní, 10

- bodový, 24, 67
- maximálně věrohodný, 13
- středně kvadratický, 14
- odhadování, 23
 - kritérium, 24, 28
 - model a diskretními parametry, 87
 - obecný regresní model, 72
 - statický model
 - diskrétní šum, 84
 - normální šum, 75
 - rovnoměrný šum, 81
- operátor průměrování, 24
- předpověď
 - bodová, 67
 - jednorázová, 103, 168
- přepínač, 109
- přirozené podmínky řízení, 66
- podmínky hladkosti, 138
- pravděpodobnost
 - úplná, 48
 - marginální, 47
 - podmíněná, 47
 - sdužená, 47
- pravděpodobnostní strom, 51
- rezidua, 24
- rozdělení
 - diskrétní, 40
 - náhodné veličiny, 38
 - spojité, 42
- směs hustot pravděpodobnosti, 110
- statistika, 10
 - pro multimodel, 113
- systém
 - data, 17
 - současná, 17
 - stará, 17
 - porucha, 17
 - výstup, 14, 17
 - vstup, 14, 17
- ukazatel, 109
- věrohodnostní funkce, 66
- virtuální regresní vektor, 138

Literatura

- [1] S. Kubík, Z. Kotek, M. Razím, J. Hrušák, and J. Branžovský, *Teorie automatického řízení II - Optimální, adaptivní a učící se systémy*, SNTL, Alfa, Praha, 1982.
- [2] J. Hátle, J. Likeš, *Základy počtu pravděpodobnosti a matematické statistiky*, SNTL, Praha, 1974.
- [3] V. Peterka, “Bayesian approach to system identification”, in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.
- [4] G.J. Bierman, *Factorization Methods for Discrete Sequential Estimation*, Academic Press, New York, 1977.
- [5] J. Anděl, *Matematická statistika*, SNTL, Praha, 1978.
- [6] R. C. Rao, *Lineární metody statistické indukce a jejich aplikace*, ACADEMIA, Praha, 1978.
- [7] M. Kárný, A. Halousková, J. Böhm, R. Kulhavý, and P. Nedoma, “Design of linear quadratic adaptive control: Theory and algorithms for practice”, *Kybernetika*, vol. 21, 1985, Supplement to No. 3, 4, 5, 6.
- [8] D.M. Titterton, A.F.M. Smith, and U.E. Makov, *Statistical Analysis of Finite Mixtures*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1985, ISBN 0 471 90763 4.
- [9] M. Kárný, J. Kadlec, and E. L. Sutanto, “Quasi-Bayes estimation applied to normal mixture”, in *Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing*, J. Rojíček, M. Valečková, M. Kárný, and K. Warwick, Eds., Praha, September 1998, pp. 77–82, ÚTIA AV ČR.
- [10] M. Kárný, I. Nagy, and Jana Novovičová, “Quasi-Bayes Approach to Multi-Model Fault Detection and Isolation”, *Adaptive Control and Signal Processing*, , John Willey and Sons, vol. 16, no. 1, pp. 61–83, 2002.
- [11] Nedoma P., Kárný M., Nagy I., Guy T.V, and Andrysek J., “Mixtools: learning and prediction with normal mixtures”, Tech. Rep. 2059, ÚTIA AV ČR, Prague, 2002.
- [12] M Papageorgiou, *Applications of Automatic Control Concepts to Traffic Flow Modelling and Control.*, vol. 50, 1983.
- [13] D. Helbing, “Traffic data and their implications for consistent traffic flow modelling.”, in *Transportation Systems, Chania*, pp. 809–814. Greece, 1997.
- [14] B.S. Kerner and P. Konhäuser, “Structures and parameters of clusters in traffic flow.”, *Phys. Rev.E* 50, pp. 54–83, 1994.

- [15] G. Confessore Bianco, L. and P. Reverberi, "Optimal location of traffic counting points for transport network control.", in *IFAC Transportation Systems, Chania*, pp. 890–895. Grece, 1997.
- [16] Saeed Ishtiaq Hounsell, N. and F. McLeod, "Short term forecasting of urban traffic congestion.", in *6th World Conference on Transport Research, Lyon*. France, 1992.
- [17] G.T. Bowen and R.D. Bretherton, "Latest developments in SCOOT - version 3.1.", in *International Conference on Road Traffic Monitoring and Control. Inst. of Electrical Engineers*. London, 1996.
- [18] J. Viegas and L. Baichuan, "Traffic control system with intermittent bus lines.", in *IFAC Transportation Systems, Chania*, pp. 905–910. Greece, 1997.
- [19] P. Příbyl, "Metody klasifikace dopravy a řízení dopravních sítí v městských aglomeracích", *Docentská habilitační práce, FD ČVUT*, 1998.
- [20] Ja. S. Zavjalov, B. I. Kvasov, and V. L. Mirošničenko, *Metody splajn-funkcij*, Nauka, Moskva, 1980.
- [21] N. P. Kornějčuk, *Splajny v teorii približenija*, Nauka, Moskva, 1984.
- [22] D. V. Lindley, *Bayesian Statistics, a Review*, SIAM, 1971.
- [23] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley Publ. Comp., 1973.
- [24] T.S. Fergusson, "A Bayesian analysis of some nonparametric problems", *The Annals of Statistics*, vol. 1, pp. 209–230, 1973.
- [25] O. Barndorff-Nielsen, *Information and exponential families in statistical theory*, Wiley, New York, 1978.
- [26] S.R. Dalal and G.J. Hall Jr., "On approximating parametric Bayes models by nonparametric Bayes models", *The Annals of Statistics*, vol. 8, pp. 664–672, 1980.
- [27] E.F. Daum, "New exact nonlinear filters", in *Bayesian Analysis of Time Series and Dynamic Models*, J.C. Spall, Ed. Marcel Dekker, New York, 1988.
- [28] A.F.M. Smith and A.E. Gelfand, "Bayesian statistics without tears: a smapling-resampling perspective", *The American Statistician*, vol. 46, pp. 84–88, 1992.
- [29] R.M. Neal, *Bayesian Learning for Neural Networks*, vol. 118, Springer, New York, Berlin Heidelberg, 1996, Lecture Notes in Statistics, ISBN 0-387-94724-8.
- [30] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore, 1997, 2nd edition.
- [31] S. Richardson and P.J. Green, "On Bayesian analysis of mixtures with an unknown number of components, with discussion", *Journal of the Royal Statistical Society, Series B*, vol. 59, no. 4, pp. 731–792, 1997.
- [32] J. Stutz and P. Cheeseman, "AutoClass - a Bayesian approach to classification", in *Maximum Entropy and Bayesian Methods*, J. Skilling and S. Sibisi, Eds. Kluwer Academic Publishers, Dordrecht, 1995.

- [33] V. Peterka and M. Kárný, “Bayesian system classification”, in *Preprints of the 5th IFAC Symposium on Identification and System Parameter Estimation*, vol. 1, pp. 349–356. Darmstadt, 1979.
- [34] M. Kárný, “Bayesian estimation of model order”, *Problems of Control and Information Theory*, vol. 9, pp. 33–46, 1980.
- [35] M. Kárný and K.M. Hangos, “Approximation of Bayes rule”, in *Preprints of the 7th IFAC/IFIP Symposium on Identification and System Parameter Estimation*. York, 1985.
- [36] M. Kárný, A. Halousková, and I. Nagy, “Modelling, identification and adaptive control of cross-direction basis weight of paper sheets”, in *Proceedings IEE Int. Conf. CONTROL 88*, pp. 159–164. The Chameleon Press Limited, Wandsworth, London, Oxford, 1988.
- [37] M. Kárný and R. Kulhavý, “Structure determination of regression-type models for adaptive prediction and control”, in *Bayesian Analysis of Time Series and Dynamic Models*, J.C. Spall, Ed. Marcel Dekker, New York, 1988, chapter 12.
- [38] M. Kárný and I. Nagy, “Recursive least squares approximation of Bayesian non-gaussian/non-linear estimation”, in *IFAC Workshop on Mutual Impact of Computing Power and Control Theory. MICC '92*, Kárný M. and Warwick K., Eds., Prague, September 1992, pp. 71–76, UTIA CSAV.
- [39] R. Kulhavý and M. B. Zarrop, “On general concept of forgetting”, *International Journal of Control*, vol. 58, no. 4, pp. 905–924, 1993.
- [40] M. Kárný, L. Berec, and P. Nedoma, “Bayesian structure estimation: State and progress”, in *Preprints of the Summer School Course*, J. Böhm and J. Rojíček, Eds., pp. 13–27. ÚTIA AVČR, Prague, 1995.
- [41] L. Berec and M. Kárný, “Identification of reality in Bayesian context”, in *Preprints of the 2nd European IEEE Workshop on Computer Intensive Methods in Control and Signal Processing, CMP'96*, L. Berec, J. Rojíček, M. Kárný, and K. Warwick, Eds., pp. 211–216. ÚTIA AV ČR, Prague, 1996.
- [42] L. Berec and M. Kárný, “Identification of reality in Bayesian context”, in *Computer-Intensive Methods in Control and Signal Processing: Curse of Dimensionality*, K. Warwick and M. Kárný, Eds. Birkhauser, 1997.
- [43] M. Valečková and M. Kárný, “Markov chains in non-linear identification: Bayesian clustering for reduced parametrisation”, in *Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing*, J. Rojíček, M. Valečková, M. Kárný, and K. Warwick, Eds., Praha, September 1998, pp. 273–278, ÚTIA AV ČR.
- [44] M. Valečková, M. Kárný, and E. L. Sutanto, “Bayesian M-T clustering for reduced parametrisation of Markov chains used for non-linear adaptive elements”, in *Preprints of the IFAC Workshop on Adaptive Systems in Control and Signal Processing*, Glasgow, August 1998, pp. 381–386, IFAC.
- [45] M. Kárný, “Probabilistic support of operators”, *ERCIM News*, , no. 40, pp. 25–26, 2000.
- [46] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, A. Kanouras, I. Nagy, P. Nedoma, A. Quinn, L. Tesař, D. Parry, and M. Tichý, “Productool background – theory, algorithms and software”, Tech. Rep., 2001, Draft of the report, 313 pp.

- [47] I. Nagy, M. Kárný, P. Nedoma, and Š. Voráčová, “Bayesian Estimation of Traffic Lane State”, *Adaptive Control and Signal Processing*, vol. 17, no. 1, pp. 51–65, 2003.