Ivan Nagy

ROZHODOVÁNÍ A ŘÍZENÍ

v dopravě

Část II. Identifikace a řízení

Obsah

Předmluva

Předkládaná zpráva je pokračováním seriálu zpráv, pojednávajících o využití výsledků bayesovské statistiky pro automatické řízení v dopravě. V této druhé části pojednáme o základech bayesovské identifikace a optimálním řízení pro stochastické modely, popsané pomocí podmíněných hustot pravděpodobností. Protože se tato práce zaměřuje především na ty, kteří se s danou problematikou teprve seznamují, je výklad zaměřen zejména na konkrétní příklady. Ty jsou ale voleny tak, aby postupy v nich použité byly snadno použitelné i pro obecnější případy.

Probírané úlohy předpokládají znalost předchozího dílu tohoto pojednání a matematiku na úrovni základního kurzu na některé z technických vysokých škol.

Pokračování uvedených dvou zpráv se bude zabývat speciálními modely a jejich identifikací. Jsou to modely, využívající spline-aproximaci hladkých funkcí nebo směsové modely.

Dále potom budou navazovat pojednání o konkrétních dopravních problémech a jejich řešení, založené na bayesovském přístupu k identifikaci a řízení.

Zpráva je určena především studentům, přicházejícím do styku s problematikou automatického řízení, aby jim pomohla překonat základní bariéru, kterou vytváří jednak poněkud rozšířený matematický aparát užívaný v této oblasti, ale zejména jeho použití, které se běžnému studentu může zdát neobvyklé.

Komentář k textu:

Poznámky jsou tištěny odlišným písmem a mají za úkol průběžně komentovat vykládaný text. Jsou odlišeny proto, aby bylo dobře možno je při dalším čtení přeskakovat a nerušily souvislost výkladu.

Odstavce označené znakem "piky" vyjadřují důležité poznatky. V zájmu snadnějšího zapamatování jsou často vyjádřeny "lidovou mluvou".

Cvičení mají za úkol upozornit na zajímavé problémy, nebo dávají čtenáři příležitost ověřit si porozumění vykládané problematice.

Kapitola 1

Model jako hustota pravděpodobnosti

Bayesovský přístup k identifikaci a řízení je založen na modelu soustavy, který je chápán jako základní prvek, předpovídající modelovanou veličinu. Umožňuje určit hp výstupu soustavy y_t na základě známých veličin (starších dat) a konstant (parametrů modelu). Predikce je potom základem pro další odhadování i řízení.

Nejčastěji používaný lineární model je regresní model. 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.

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. Podle bayesovského přístupu jsou neznámé parametry rovněž pokládány za náhodné.

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 identifikaci a dále pro řízení.

Značení

V předchozí kapitole jsme používali symbolické značení pro pravděpodobnosti. Jestliže bylo známo o jaký jev X se jedná, pak jsme, z důvodů přehlednosti, místo označení $P(X = x)$ zavedli symbol $P(x)$. Podobně i zde zavedeme odpovídající symboliku pro hp popisující signály nebo parametry. Tato symbolika je v literatuře často používána, zlepšuje přehlednost, ale při nedostatku přesného porozumění může být matoucí. Zjednodušení spočívá v tom, že ve značení přestaneme rozlišovat mezi náhodnou veličinou X a její hodnotou x. Budeme důsledně 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.

$$f(y_t|u_t) \text{ znamená } f_{Y_t|U_t}(y_t|u_t).$$

Poznámka: *Signály jsou náhodné posloupnosti, tedy indexované náhodné veličiny. Odtud index, který se u náhodných veličin, popisujících signály, vyskytuje.*

1.1 Regresní model

Rovnice modelu

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

$$\boxed{y_t = m(\varphi_t, \theta) + e_t,} \tag{1.1}$$

kde

y_t je modelovaná veličina – výstup soustavy,

φ_t je regresní vektor, obsahující aktuální řídící veličinu u_t, data, zpožděná až do ν kroků zpět (ν je řád modelu) a jedničku pro respektování konstanty modelu, tj.

$$\varphi_t = [u_t, y_{t-1}, u_{t-1}, ..., y_{t-\nu}, u_{t-\nu}, 1]^T,$$

θ je vektor regresních koeficientů (známé nebo odhadnuté konstanty), jejichž pořadí koresponduje s pořadím dat v regresním vektoru

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

$m(\varphi_t, \theta)$ 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; nejčastější je lineární vyjádření

$$m(\varphi_t, \theta) = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + ... + +a_\nu y_{t-\nu} + b_\nu u_{t-\nu} + c$$

e_t je náhodná veličina (šum) o které předpokládáme:

- nulovou střední hodnotou $E[e_t] = 0, \ \forall t$,
- konstantní rozptyl $D[e_t] = \sigma^2, \ \forall t$,
- nekorelovanost e_t s daty obsaženými v regresním vektoru $\varphi_t, \ \forall t$.

Symbolem Θ označíme množinu všech parametrů modelu, tj

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

Hp modelu

Uvedená rovnice modelu definuje rozdělení náhodné veličiny y_t ve formě podmíněné hp

$$\boxed{f(y_t|\varphi_t, \Theta).} \tag{1.2}$$

Jde skutečně o podmíněnou hp, protože jedině při znalosti veličin φ_t a Θ z podmínky hp jsme schopni rovnici pro predikci výstupu y_t využít. Pro regresní vektor φ_t a parametry θ vypočteme hodnotu funkce $m(\varphi_t, \theta)$. Rozdělení výstupu y_t je pak dáno transformací (posunutím) daného rozdělení náhodné veličiny e_t do bodu určeného funkcí m.

`P o z n á m k a:` *Tato transformace odpovídá základní transformaci náhodné veličiny normovaného normálního rozdělení $Z \to X$ pomocí transformační rovnice*

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

která spočívá v tom, že v hp náhodné veličiny nahradíme proměnné z výrazem $\frac{x-\mu}{\sigma}$, který plyne z transformační rovnice, a násobení celé hp Jakobiánem transformace $1/\sigma$.

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

$$E[y_t|\varphi_t, \theta] = E[m(\varphi_t, \theta)|\varphi_t, \Theta] + E[e_t|\varphi_t, \Theta] = m(\varphi_t, \theta) \tag{1.3}$$

▷ `Rozptyl` y_t

$$D[y_t|\varphi_t, \Theta] = D[y_t|\Theta] = \sigma^2 \tag{1.4}$$

▷ `Variance` y_t

$$E[y_t^2|\varphi_t, \Theta] = (E[y_t|\varphi_t, \Theta])^2 + D[y_t|\Theta] = m(\varphi_t, \theta)^2 + \sigma^2 \tag{1.5}$$

`P o z n á m k a:` *Uvedené charakteristiky jsou podmíněné. Samozřejmě existují i nepodmíněné, ty ale nejsme schopni s pomocí modelu určit a musíme je konstruovat jinak. Na první pohled je např. zřejmé, že při jejich konstrukci bude zapotřebí odhadovat parametry, tj. provádět identifikaci. Model vyžaduje znalost parametrů, ty ale v praxi většinou známé nejsou.*

Typy regresních modelů

Uvedené vztahy pro rovnici modelu a jeho charakteristiky jsou obecné. Jsou založeny pouze na skutečnosti, že model je konstruován jako součet deterministické funkce a náhodné veličiny. Základní předpoklad je, že náhodná veličina e_t je nekorelovaná s daty obsaženými v regresním vektoru φ_t. Většinou požadujeme, aby šum měl nulovou střední hodnotu. Pokud nemá, je střední hodnota modelu dána hodnotou funkce m plus touto střední hodnotou šumu.

Takto definovaný model se nazývá *regresní model.* Podle konkrétní specifikace funkce m, případně rozdělení náhodné veličiny e_t, dostaneme konkrétní typ regresního modelu.

- Je-li funkce m lineární funkcí vzhledem k parametrům, např. je-li

$$m = \varphi_t^T \theta = \sum_i \varphi_{i,t} \theta_i,$$

  potom se tento model nazývá *lineární regresní model.* V případě, že m není lineární, nazývá se *nelineární regresní model.*

- Jsou-li parametry θ konstanty, nazýváme tento model *regresní model s konstantními koeficienty.* Jsou-li koeficienty funkcemi času, jde o *regresní model s proměnnými koeficienty.*

- Jestliže regresní vektor obsahuje také zpožděné modelované veličiny, tj. $y_{t-1}, y_{t-2}, \ldots$ 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 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 signály, se kterými model pracuje, vícerozměrové, tj. v každém časovém okamžiku tvoří vektor, nazývá se takový model *mnoharozměrový regresní model.* Jsou-li signály skalární veličiny, nazývá se *skalární regresní model.*

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

1.2 Příklady regresních modelů

PŘÍKLAD 1.1 Dynamický regresní model s normálním šumem

Plným názvem je to skalární lineární dynamický regresní model řádu ν s konstantními koeficienty a normálním rozdělením šumu. V oblasti lineárních modelů, je takový model nejčastěji používán. Lineární vyjádření střední hodnoty a normální rozdělení šumu pokrývá velkou většinu praktických příkladů a jeho použití pro identifikaci a řízení je výpočetně výhodné. Je to stochastická obdoba známé diferenční rovnice.

Rovnice modelu

Model je určen rovnicí

$$\boxed{y_t = \varphi_t^T \theta + e_t,} \tag{1.6}$$

kde

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

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

e_t je skalární náhodná veličina s rozdělením $N(0, \sigma^2)$, tedy její rozdělení je

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

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

Model lze rovněž rozepsat do tvaru diferenční rovnice

$$y_t = b_0 u_t + a_1 y_{t-1} + b_1 u_{t-1} + ... + a_\nu y_{t-\nu} + b_\nu u_{t-\nu} + c + e_t. \tag{1.7}$$

Momenty modelu

Model určuje rozdělení, nebo je-li třeba charakteristiky modelovaného výstupu y_t

▷ `Střední hodnota a rozptyl`

$$E[y_t|\varphi_t, \Theta] = \varphi_t^T \theta, \qquad D[y_t|\varphi_t, \Theta] = \sigma^2. \tag{1.8}$$

▷ `Variance`

$$E[y_t^2|\varphi_t, \Theta] = D[y_t|.] + E[y_t|.]^2 = (\varphi_t^T \theta)^2 + \sigma^2. \tag{1.9}$$

Hp modelu

Hp modelu je dána transformací hp šumu $f(e_t)$ podle rovnice modelu. Bude mít opět normální rozdělení, ale se střední hodnotou a rozptylem podle (1.10)

$$\boxed{f(y_t|\varphi_t, \Theta) = \tfrac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\tfrac{1}{2\sigma^2}(y_t - \varphi_t^T \theta)^2\right\}} \tag{1.10}$$

`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.*

PŘÍKLAD 1.2 Statický regresní model s normálním šumem

A. Řízený systém

Statický model je speciálním případem dynamického regresního modelu, kdy regresní vektor neobsahuje žádné zpožděné výstupy. Regresní vektor může tedy obsahovat řídící veličinu a jedničku pro respektování konstanty.

Nejprve budeme uvažovat model bez konstanty, kde modelovaná veličina je úměrná řídící veličině s neznámým koeficientem úměrnosti – neznámým parametrem modelu θ. Pro $u_t = 1,\ \forall t$ bude možno tento parametr chápat jako konstantu modelu.

Navíc budeme předpokládat známý rozptyl šumu $\sigma^2 = 1$. Tento model bude mít tedy jedinou řídící veličinu u_t a jediný neznámý parametr $\Theta = \theta$.

Rovnice modelu

$$y_t = \theta u_t + e_t, \tag{1.11}$$

kde

θ je skalární parametr (konstanta),

y_t, u_t jsou výstup a vstup soustavy,

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

Momenty modelu

Střední hodnota: θu_t, Rozptyl: 1, Variance: $\theta^2 u_t^2 + 1$

Hp modelu

$$f(y_t|u_t,\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y_t^2 - 2\theta y_t u_t + \theta^2 u_t^2)\right\} \tag{1.12}$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\,[-1 \quad \theta] \left(\begin{bmatrix} y_t \\ u_t \end{bmatrix} [y_t \quad u_t]\right) \begin{bmatrix} -1 \\ \theta \end{bmatrix}\right\}. \tag{1.13}$$

Exponent prvního vyjádření hp přímo ukazuje, jaké budou statistiky pro identifikaci. [1] Druhé vyjádření reprezentuje model jako funkci neznámých parametrů a je vhodné pro konstrukci věrohodnostní funkce.

[1] Budou to kvadráty $y_t^2,\ u_t^2$ a součiny $y_t u_t$ pro t na časovém intervalu identifikace.

B. Neřízený systém

Ještě jednodušší variantou tohoto modelu je model neřízeného systému, který obdržíme, jestliže v předchozím modelu položíme $u_t = 1,\ \forall t$.

Rovnice modelu

$$y_t = \theta + e_t, \qquad e_t \sim N(0, \sigma^2). \tag{1.14}$$

Hp modelu

Pro známý rozptyl $\sigma^2 = 1$ je hp modelu

$$f(y_t|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y_t^2 - 2\theta y_t + \theta^2)\right\}. \tag{1.15}$$

P o z n á m k a: *V případě, kdy by rozptyl šumu σ^2 nebyl znám, měl by šum rozdělení $N(0, \sigma^2)$ a hp modelu by byla následující*

$$f(y_t|\Theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_t^2 - 2\theta y_t + \theta^2)\right\}. \tag{1.16}$$

Neznámé parametry v tomto případě jsou $\Theta = [\theta, \sigma^2]$.

PŘÍKLAD 1.3 Statický regresní model s rovnoměrným šumem

Pro vyjádření hp rovnoměrného rozdělení budeme používat tzv. *indikátorovou funkci* χ intervalu $I = [a, b] \subset R$

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

Rovnice modelu

Model je popsán rovnicí

$$y_t = \theta + e_t, \tag{1.17}$$

kde

θ je neznámý spojitý parametr,

e_t je náhodná veličina s rovnoměrným rozdělením s hp

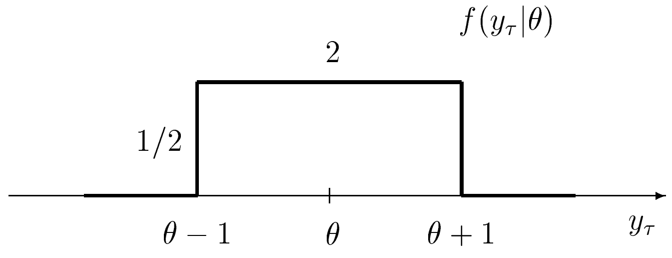
$$f(e_t) = \frac{1}{2}\chi(e_t \in [-1; 1]).$$

P o z n á m k a: *Opět pro jednoduchost uvažujeme hp šumu se známými parametry. Její rozsah je od -1 do 1. Pokud meze pro hp nebudou jednotkové, ale budou známé, je situace stejná. Složitější model dostaneme v případě, kdy tyto meze nejsou známé a představují další parametry, které je třeba identifikovat.*

Hp modelu

Tato hp má také rovnoměrné rozdělení, které opět dostaneme transformací hp šumu

$$f(y_t|\theta) = \frac{1}{2}\chi(y_t \in [\theta - 1;\ \theta + 1]),\ y_t \in R. \tag{1.18}$$



PŘÍKLAD 1.4 Statický regresní model s diskrétním šumem

Jednoduché, ale poněkud odlišné jsou operace prováděné při identifikaci modelů s diskrétními veličinami. Tyto veličiny mohou být zásadně dvojího druhu: signály a parametry. Zde se budeme věnovat modelům s diskrétními signály a se spojitými parametry. Příští příklad bude věnován modelům s diskrétními parametry.

A. Známý parametr modelu

Jedná se snad o nejjednodušší statistický model, popisující např. hod mincí. Abychom tento příklad udělali trochu zajímavější, uvažujeme poškozenou minci, kde pravděpodobnost rubu se nerovná pravděpodobnosti lícu. Předpokládáme ale, že pravděpodobnost padnutí rubu (a tedy i lícu) je známa (např. z dřívějších experimentů) a rovná se θ_0 pro rub a $1 - \theta_0$ pro líc.

Uvažujeme posloupnost hodů v časech $\tau = 1, 2, ..., t$, 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

Rovnice modelu je triviální

$$y_t = e_t, \tag{1.19}$$

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

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

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

Hp modelu

Tato hp je shodná s hp šumu

$$f(y_t) = \theta_0^{1-y_t}(1-\theta_0)^{y_t}, \tag{1.20}$$

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 rubu (a tedy i lícu) není známa. Takový model má velmi dobrou dopravní interpretaci, kterou jsme se zabývali v úvodním příkladu minulého dílu. Při ní uvažujeme křižovatku ve tvaru T. Do ní přijíždějí automobily a odbočují buď vpravo nebo vlevo. My 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 identifikace směrových vztahů na křižovatce.

Rovnice modelu

Rovnice modelu je shodná s předchozím příkladem. Jak jsme již uvedli, v takto jednoduchém příkladě není zajímavá.

Hp modelu

Hp modelu je podmíněná hp s alternativním rozdělením. Závisí na jednom neznámém parametru $\theta \in [0,\ 1]$ s významem pravděpodobnosti padnutí rubu (R). Má tvar

$$f(y_t|\theta) = \theta^{1-y_t}(1-\theta)^{y_t}, \qquad y_t \in \{0, 1\}. \tag{1.21}$$

Je tedy $\quad f(0|\theta) = P(y_t = 0|\theta) = \theta,$
$\quad f(1|\theta) = P(y_t = 1|\theta) = 1 - \theta.$

Definuje tedy: pravděpodobnost padnutí rubu $(y_t = 0)$ je θ a pravděpodobnost padnutí lícu $(y_t = 1)$ je $1 - \theta$.

PŘÍKLAD 1.5 Modely s diskrétními parametry

Modely s diskrétním parametrem mohou mít dvojí význam. Jak jsme již uvedli, neumíme identifikovat časově proměnné parametry. Proto předpokládáme, že parametr je konstanta, ale nám neznámá. Naše neznalost je příčinou neurčitosti a proto parametry považujeme za náhodné veličiny, popsané svými hp. Pokud máme důvod předpokládat, že tato náhodná veličina může nabývat jen konečný počet různých hodnot, jedná se o diskrétní náhodnou veličinu s diskrétní hp. Předpokládáme, že (při daném experimentu) správná je jen jedna hodnota a tu odhadujeme. Přitom pravděpodobnosti jednotlivých možných hodnot parametru mohou být známé, jako budeme předpokládat zde, nebo neznámé, a potom tyto pravděpodobnosti musíme odhadovat. To je ale již úloha odhadování spojitého parametru z diskrétních dat, kterou jsme se zabývali při odhadování pravděpodobností rubu a líce falešné mince.

Jiná situace nastane, když připustíme, že skutečné hodnoty parametru se během experimentu doopravdy mění a parametr nabývá vždy jednu z konečného počtu možných hodnot. Přitom předpokládáme, že pravděpodobnosti jednotlivých hodnot parametru jsou pevné (konstantní). Tuto situaci lze také interpretovat tak, že soustava má konečný počet vnitřních stavů, mezi nimiž přechází, a v každém stavu je popsatelná modelem s konstantními koeficienty. Nyní hovoříme o situaci, kdy modely jsou stejného typu a liší se jen v rozdílných hodnotách parametrů. Modely mohou být ale také zcela odlišné. Soustava je tak popsána celým systémem modelů, ve kterém nejen identifikujeme parametry jednotlivých modelů (komponent) ale také odhadujeme, který z modelů je v dané chvíli aktivní (predikce aktivní komponenty). Takový model nazýváme směsovým modelem. Jeho praktická identifikace je možná jen při aproximaci Bayesova vzorce. Zmíněným modelům je věnován následující díl tohoto pojednání.

`Poznámka:` *Směsové modely mají velmi dobré uplatnění v dopravě. Právě v této oblasti se velmi často setkáváme se soustavami, vyskytujícími se v několika různých, velmi odlišných stavech. Připomeňme např. silniční křižovatku v centru města v různých denních dobách: ráno, dopoledne, odpoledne, večer a v noci.*

Modely pro obě zmíněné situace jsou stejného typu. Situace se liší jen předpokladem o změnách parametrů soustavy.

Uvažujme tedy statický regresní model se známými charakteristikami ve formě hp $f(e)$ a $f(\theta)$. Protože předpokládáme, že pravděpodobnosti hodnot parametru jsou známé, nebudeme v tomto příkladě provádět akumulaci dat pro identifikaci. Provedeme pouze jeden krok identifikace v čase. V něm provedeme predikci výstupu a po změření výstupu odhadneme hodnotu parametru, který jej generoval. Abychom zdůraznili tuto situaci, nebudeme u signálů uvádět časový index.

A. Diskrétní parametr i signál

Rovnice modelu

$$y = \theta + e, \tag{1.22}$$

kde

$f(e)$ a $f(\theta)$ jsou dány tabulkami

e	-1	1
$f(e)$	0.4	0.6

,

θ	-2	0	2
$f(\theta)$	0.2	0.5	0.3

,

obě náhodné veličiny jsou nezávislé.

Hp modelu

Je to hp $f(y|\theta)$ zadaná opět pomocí tabulky. Ve sloupcích tabulky jsou modely vždy pro určitou hodnotu parametru v podmínce.

$f(y\|\theta)$ $y\backslash\theta$	-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

(1.23)

Tuto tabulku sestavíme z tabulek hp e a θ takto: Nejdříve určíme všechny možné hodnoty výstupu y tak, že pro každou hodnotu θ vezmeme všechny hodnoty e a dosazujeme do rovnice modelu. Potom určíme pravděpodobnosti tak, že pro danou hodnotu θ určíme hodnotu e, která se na ní podílela a vezmeme její pravděpodobnost. Pravděpodobnosti hodnot θ se neuplatní. Proč? [2]

Ve sloupcích tabulky, představující 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

Rovnice má opět stejný tvar

$$y = \theta + e, \tag{1.24}$$

kde

θ je diskrétní a e spojitá náhodná veličina s rozděleními

θ	-2	0	2
$f(\theta)$	0.2	0.5	0.3

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

Hp modelu

dostaneme jako transformaci nv e na nv y

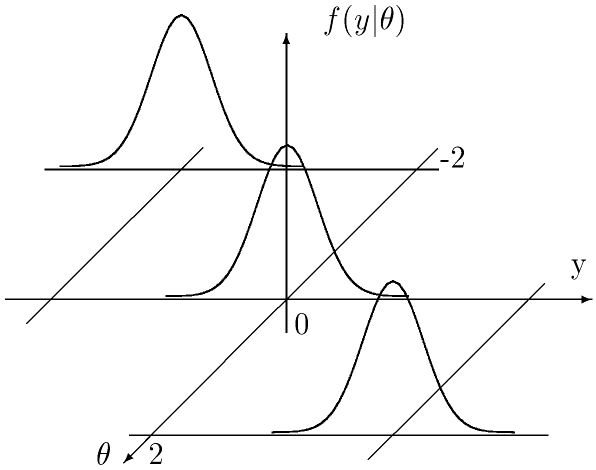
$$f(y|\theta) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y-\theta)^2\right\}, \qquad \theta = -2, 0, 2. \tag{1.25}$$

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

$$f(y|\theta=-2) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y+2)^2\right\} \qquad \text{pro} \quad \theta = -2$$

[2]Protože model je hp podmíněná parametry. Ty zde tedy vystupují ne jako náhodné veličiny, ale jako hodnoty – tj jako by byly známé. Proto model nezávisí na pravděpodobnostech těchto hodnot.

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



PŘÍKLAD 1.6 Mnoharozměrový regresní model

Mnoharozměrovou identifikací a řízením se zde podrobně nezabýváme. Protože však ve většině praktických úloh je třeba modelovat několik veličin, a protože mnoharozměrový model je možno převést na soustavu tzv. jednorozměrných faktorů, uvedeme pro úplnost také zápis mnoharozměrového regresního modelu.

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

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

Rovnice modelu

Rovnice tohoto 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

$$
y_t = \theta^T \varphi_t + e_t, \tag{1.26}
$$

kde

y_t je sloupcový vektor dimenze n_y,

$\theta^T = [B_0, A_1, B_1, ..A_\nu, B_\nu, C]$, kde A_i, C jsou matice typu $(n_y \times n_y)$ a B_i jsou matice typu $(n_y \times n_u)$,

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

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

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

Hp modelu

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

$$f(y_t|\varphi_t, \Theta) = (2\pi)^{-\frac{n_y}{2}} |R|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(y_t - \theta^T\varphi_t)^T R^{-1} (y_t - \theta^T\varphi_t) \right\} \tag{1.27}$$

Příklad pro $n_y = 2,\ n_u = 1,\ \nu = 1$

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

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

Střední hodnota a kovarianční matice šumu

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

Kapitola 2

Predikce a identifikace

V této kapitole si všimneme identifikace pravděpodobnostních modelů. Ta je těsně spjata, podobně jako i řízení, s predikcí budoucího výstupu soustavy. Pojednáme tedy o predikci a identifikaci společně.

Důležité úvodní poznámky

Značení integrálů
Bayesovská statistika pracuje se sdruženými, marginálními a podmíněnými hp. Při výpočtech s nimi se často používá integrace. 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 pravidlo: 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 R^n.

Přirozené podmínky řízení
Při odhadů parametrů modelu soustavy, řízené ve zpětné vazbě prostřednictvím řídící veličiny, uvažujeme platnost tzv. přirozených podmínek řízení. *Ty jsou založeny na předpokladu, že skutečné parametry modelu Θ jsou neznámé, a nelze je tedy použít při konstrukci aktuální řídící veličiny u_t. Tato řídící veličina, je podle našeho rozhodnutí zpětnovazební, a tedy smí záviset pouze na starých datech $D(t-1) = [y_{t-1}, u_{t-1}, y_{t-2}, ...]^T$. Platí tedy*

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

Použitím Bayesova vzorce lze ukázat, že předchozí podmínka je ekvivalentní s podmínkou

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

která říká, že pro konstrukci odhadu parametrů (která je popsána právě touto hp) nelze využít aktuální řídící veličinu. Ta je totiž podle našeho rozhodnutí konstruována právě na základě dat $D(t-1)$ a tedy nenese žádnou novou informaci než tu, která je již v těchto datech obsažena.

Přehled použitých hp

Než začneme s vlastním výkladem predikce a identifikace, uvedeme přehled jednotlivých prvků (většinou hp), které se dále budou objevovat. Jsou to: *model soustavy*, reprezentovaný rovnicí nebo svou hp, *prediktivní hp*, která předpovídá výstup soustavy bez znalosti parametrů a *hp parametrů*, která se před aktualizací naměřenými daty nazývá apriorní a po aktualizaci aposteriorní hp.

Model soustavy

Budeme uvažovat regresní model obecného řádu ν s jedním vstupem a výstupem podle vztahu (1.26)

$$y_t = \theta^T \varphi_t + e_t, \tag{2.1}$$

kde
φ_t je regresní vektor, obsahující aktuální řízení a starší data až do řádu modelu;
e_t je šum s rozdělením $N(0, \sigma^2)$.

Hp modelu

Model lze reprezentovat ve formě podmíněné hp

$$f(y_t|\varphi_t, \Theta), \tag{2.2}$$

kde $\Theta = \{\theta, \sigma^2\}$ je množina všech neznámých parametrů modelu.

`Poznámka:` *Hp (2.2) je rovnicí (2.1) definována. Samotná hp je ale obecnější. Lze pomocí ní vyjádřit např. i modely nelineární. Takové modely však v tomto okamžiku nejsou předmětem našeho zájmu.*

Prediktivní hp

Prediktivní hp je hp výstupu, podmíněná příslušnou akční veličinou a staršími daty

$$f(y_t|u_t, D(t-1)), \tag{2.3}$$

kde $D(t-1)$ značí všechna data naměřená na soustavě až do okamžiku $t-1$, včetně počátečních dat $D(t-1) = [d_{t-1}^T, d_{t-2}^T, ..d_1^T, d_0^T, ..]^T = [y_{t-1}, u_{t-1}, y_{t-2}, ..y_0, u_0, ..]^T$.

`Poznámka:` *Tato hp je podobná hp modelu, neobsahuje ale v podmínce parametry θ a místo nich obsahuje všechna data od počátku identifikace. Hp modelu se vyskytuje v úlohách se známými parametry, prediktivní hp v úlohách, kde parametry modelu známy nejsou. Problém spočívá v tom, že prediktivní hp v převážné většině případů není známá a je třeba ji konstruovat. Její konstrukce, volně řečeno, spočívá v tom, že data z podmínky prediktivní hp se implicitně použijí k odhadu parametrů, které pro výpočet výstupu z hp modelu potřebujeme.*

Konstrukcí prediktivní hp se budeme zabývat dále v této kapitole.

Hp parametrů

Je to podmíněná hp udávající rozdělení neznámých parametrů, které podle bayesovského principu považujeme za náhodné veličiny. Je podmíněná všemi dosud známými daty, tj. naměřenými i apriorními. V dalším výkladu se bude objevovat v následujících podobách:

– podmíněná všemi (i současnými) daty

$$f(\Theta|D(t)) \qquad \text{(aposteriorní hp)}, \tag{2.4}$$

– podmíněná všemi minulými daty (ne současnými)

$$f(\Theta|D(t-1)) \qquad \text{(aposteriorní hp z minulého kroku)}, \tag{2.5}$$

– podmíněná jen apriorními daty (nebo jinou apriorní informací)

$$f(\Theta|D(0)) \qquad \text{(apriorní hp)}. \tag{2.6}$$

Poznámka: *Aposteriorní hp z minulého kroku se také někdy nazývá apriorní hp pro současný krok.*

Poznámka: *Aposteriorní hp přímo neznáme a musíme ji rekurentně počítat z apriorní hp a měřených dat, s pomocí modelu soustavy.*

Její konstrukcí se budeme zabývat dále v této kapitole.

2.1 Bodová predikce a bodové odhady

Pro motivaci rekurzivního odhadování parametrů vyjdeme z praktické úlohy – bodové predikce 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 predikce, nebo bodový odhad neznámého parametru.

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

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

Tento typ kriteria vede na bodové odhady ve formě střední hodnoty, podmíněné minulými daty. Predikce výstupu a bodový odhad parametrů jsou

$$\boxed{\hat{y}_t = E[y_t|u_t, D(t-1)] \qquad \text{a} \qquad \hat{\Theta}_t = E[\Theta|D(t-1)],} \tag{2.7}$$

kde ve vyjádření bodového odhadu parametrů jsou využity přirozené podmínky řízení, zmíněné na začátku této kapitoly.

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

Odvození tohoto tvrzení pro predikci je uvedeno v Doplňku 5.2. Odvození pro odhad parametru je obdobné.

Bodová predikce podle kvadratického kriteria je dána podmíněnou střední hodnotou, jak jsme právě odvodili. Výpočet střední hodnoty nyní záleží na tom, zda známe nebo neznáme parametry modelu. I když v této kapitole hovoříme o identifikaci, a tedy máme na mysli model s neznámými parametry, pojednáme zde o obou možnostech pohromadě.

Bodová predikce pro model se známými parametry

Model

Uvažujme model (2.1) s normálním rozdělením šumu $e_t \sim \mathcal{N}(0, \sigma^2)$ a se známými parametry $\Theta = [\theta, \sigma^2]$. Tento model definuje hp

$$f(y_t|\varphi_t) = \mathcal{N}(\theta^T \varphi_t, \sigma^2).$$

kde $\Theta = [\theta, \sigma^2]$ známe.

`Poznámka:` *Protože parametry jsou známá čísla, nevyskytují se v podmínce, tj. když nám někdo prozradí jejich hodnoty, nic se na našich znalostech o y_t nezmění (protože jsme je stejně znali). Tato informace neovlivní naše znalosti, lze ji tedy z podmínky hp vyloučit.*

`Poznámka:` *Známe-li regresor, známe vše, co je potřebné pro konstrukci hp výstupu y_t. Starší data než ta, která jsou obsažena v regresním vektoru, jsou zbytečná. Platí proto*

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

Bodová predikce

Predikci určíme jako střední hodnotu

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

Dostáváme tedy

$$\boxed{\hat{y}_t = \varphi_t^T \theta.} \tag{2.8}$$

`Poznámka:` *Výraz, který jsme obdrželi je možno ihned použít pro výpočet. Za regresní vektor dosadíme změřená data a vektor parametrů známe. Můžeme tedy vypočítat hodnotu predikce.*

`Poznámka:` *Jako predikci jsme dostali hodnotu spočtenou z modelu, jestliže za šum dosadíme jeho střední hodnotu, tj. nulu.*

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

Bodová predikce pro model s neznámými parametry

Situace je složitější. Předchozí výsledek není přijatelný, protože do něho nyní neumíme dosadit – neznáme θ. Jak uvidíme, součástí predikce bude odhad parametrů z naměřených dat, tj. identifikace.

Model

Model (2.1) je nyní zadán pomocí hp

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

Lze jej použít pro generování výstupu y_t jen tehdy, když "zjistíme" hodnoty jinak neznámých parametrů θ i σ^2. Tím máme většinou na mysli zjištění všech možných hodnot a jejich pravděpodobností (tj. rozdělení neznámého parametru). Model pak použijeme ve vzorci pro úplnou pravděpodobnost místo prostého dosazení bodových odhadů.

Prediktivní hp

Predikce je definována stejně jako pro známé parametry

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

avšak prediktivní hp $f(y_t|u_t, D(t-1))$ nyní neznáme a musíme ji určit pomocí hp modelu. Použijeme k tomu definici marginální a podmíněné pravděpodobnosti

$$f(y_t|u_t, D(t-1)) = \int f(y_t, \Theta|u_t, D(t-1))\mathrm{d}\Theta = \qquad \text{(marginální pravděp.)}$$

$$= \int f(y_t|u_t, D(t-1), \Theta) f(\Theta|u_t, D(t-1))\mathrm{d}\Theta \qquad \text{(podmíněná pravděp.)}$$

V posledním řádku můžeme ještě podle poznámek v úvodu provést dvě úpravy

1. $f(y_t|u_t, D(t-1), \Theta) = f(y_t|\varphi_t, \Theta)$ při známých parametrech závisí y_t jen na φ_t;

2. $f(\Theta|u_t, D(t-1)) = f(\Theta|D(t-1))$ *přirozené podmínky řízení* - viz poznámka v úvodu kapitoly.

Je tedy

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

P o z n á m k a: *Interpretace tohoto výsledku je názorná a odpovídá metodě úplné pravděpodobnosti. Mohu použít jen model s dosazenými parametry. Jejich správné hodnoty však 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 "pravděpodobnosti hodnot parametrů"* $f(\Theta|D(t-1))$.

Bodová predikce

Nyní můžeme přistoupit k výpočtu střední hodnoty $E[y_t|u_t, D(t-1)]$, podle (2.7). S použitím předchozího vyjádření prediktivní hp je

$$E[y_t|u_t, D(t-1)] =$$

$$= \int_{y_t} y_t f(y_t|u_t, D(t-1))\mathrm{d}y_t$$

$$
\begin{aligned}
&= \int_{y_t} y_t \int_\Theta f(y_t|\varphi_t,\Theta) f(\Theta|D(t-1)) \mathrm{d}\Theta \mathrm{d}y_t \quad \text{dosazeno z (2.10)} \\
&= \int_\Theta \left[\int_{y_t} y_t f(y_t|\varphi_t,\Theta)\mathrm{d}y_t\right] f(\Theta|D(t-1))\mathrm{d}\Theta \quad \text{záměna integrace} \\
&= \int_\Theta \varphi_t^T \theta f(\Theta|D(t-1))\mathrm{d}\Theta \quad \text{dosazeno z (2.8)} \\
&= \varphi_t^T \int_\Theta \theta f(\Theta|d(t-1))\mathrm{d}\Theta = \varphi_t^T \hat{\theta}_{t-1} \quad \text{označen bod. odhad } \theta
\end{aligned}
$$

Protože bodová predikce se rovná této střední hodnotě, dostáváme

$$
\boxed{\hat{y}_t = \varphi_t^T \hat{\theta}_{t-1}.} \tag{2.9}
$$

♣ Bodová predikce výstupu při neznámých parametrech regresního modelu je dána skalárním součinem regresního vektoru s vektorem bodových odhadů parametrů.

2.2 Predikce a identifikace

Podobně jako jsme pro model soustavy uvažovali dvojí možné vyjádření: rovnicí nebo pomocí hp, můžeme takové dvě (částečně pomyslné) linie sledovat i při vytváření algoritmů predikce a identifikace. První linie počítá se statistikami hp (tj. měřená data akumuluje do statistik, případně jako výsledky počítá bodové odhady), druhá linie počítá se samotnými hustotami pravděpodobnosti. První cesta dává praktické výsledky (hodnoty odhadovaných veličin), druhá je obecnější, ale její výsledky nelze přímo využít jako algoritmy realizovatelné na počítači.

Vlastní identifikace často není cílem ale jen prostředkem k dosažení jiného cíle. V předchozím odstavci jsme jako cíl postavili predikci výstupu pro model s neznámými parametry. Jako podúloha, která vznikla při řešení tohoto problému, se objevila právě identifikace. Protože jsme ale požadovali bodovou predikci, byly jako výsledek identifikace potřeba také bodové odhady parametrů.

Vezmeme-li ale za cílovou úlohu např. optimální řízení s pravděpodobnostním modelem, minimalizující střední hodnotu kriteria, penalizujícího kvadráty vstupu a výstupu (tak, jak jej budeme uvažovat na konci této kapitoly), dostaneme také jako podúlohu identifikaci. V tomto případě se ale úloha identifikace neomezí na pouhý výpočet bodových odhadů a jejich dosazení za neznámé parametry. Pracuje se zde s celými hp parametrů, na základě pravidla o úplné pravděpodobnosti. Tj., 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 předání celé informace o odadovaném parametru včetně jeho neurčitosti je velkou předností bayesovského přístupu.

Konstrukce prediktivní hp

Jak jsme ukázali v minulé kapitole, lze prediktivní hp konstruovat podle vzorce pro úplnou pravděpodobnost

$$\boxed{f(y_t|u_t, D(t-1)) = \int f(y_t|\varphi_t, \Theta) f(\Theta|D(t-1)) \mathrm{d}\Theta.} \tag{2.10}$$

První hp na pravé straně je hp definovaná modelem (podobně jako pro známé parametry). Druhá je hp parametrů, podmíněná znalostí dat až do okamžiku $t-1$. Tu neznáme a musíme ji postupně napočítávat s měřenými daty pomocí identifikace. Konstrukcí této hp se budeme zabývat v následujícím odstavci, věnovaném identifikaci.

♣ Konstrukce prediktivní hp probíhá podle vzorce pro úplnou pravděpodobnost. Vstupuje do ní model soustavy a hp parametrů, kterou neznáme a musíme ji postupně napočítávat z dat a apriorní hp v identifikaci.

P o z n á m k a: *Prediktivní hp je možno počítat průběžně, tak jak jsme to zatím ukázali, ale také jednorázově, pro všechna dosud naměřená data najednou. Odvození této jednorázové predikce provedeme v kap. 7, která pojednává o řízení při neznalosti parametrů modelu. Zde pro úplnost uvedeme jen výsledek.*

$$f(y_t|u_t, D(t-1)) = \frac{I_t}{I_{t-1}}, \tag{2.11}$$

kde

$$I_\vartheta = \int \prod_{\tau=1}^{\vartheta} [f(y_\tau|\varphi_\tau, \Theta)] f(\Theta|D(0)) d\Theta, \quad \text{pro } \vartheta = t,\ t-1.$$

a $f(\Theta|D(0))$ je apriorní hp z počátku identifikace.

Konstrukce hp parametrů

Při použití vztahu (2.10) pro prediktivní hp je třeba použít hp parametrů $f(\Theta|D(t-1))$. Tuto hp přímo neznáme. Její konstrukce probíhá rekurzivně, tak jak měříme data. Vychází z tzv. apriorní hp, která vyjadřuje naše předběžné znalosti parametrů, nebo se opírá o předem změřená data. K této hp je postupně s využitím Bayesova vzorce přidávána informace získaná z měřených dat, a tak je apriorní hp přepočítávána na tzv. aposteriorní hp. Ta je buď žádaným výsledkem, jako například pro výpočet predikce, nebo jsou z ní počítány bodové odhady, 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 minulé aposteriorní hp $f(\Theta|D(t-1))$

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

$$= \frac{f(y_t|u_t, D(t-1), \Theta) f(\Theta|u_t, D(t-1))}{f(y_t|u_t, D(t-1))}. \quad \text{(Bayesův vzorec)}$$

Použitím stejných argumentů jako pro predikci (2.10) lze v podmínce první hp v čitateli zlomku odstranit data starší, než jsou v regresoru, a v podmínce druhé hp v čitateli odstranit veličinu u_t - přirozené podmínky řízení. Hp ve jmenovateli zlomku nezávisí na Θ, představuje tedy jen konstantu, která zajišťuje jednotkový integrál aposteriorní hp, tj. $\int f(\Theta|D(t))\mathrm{d}\Theta = 1$. 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

$$\boxed{f(\Theta|D(t)) \propto f(y_t|\varphi_t, \Theta) f(\Theta|D(t-1)).} \tag{2.12}$$

Jedná se o rekurentní výpočet hp parametrů, který začíná výpočet s apriorní hustotou

$$f(\Theta|D(0)) = f(\Theta)$$

a v jednotlivých časových okamžicích $\tau = 1, 2, ..t$ poskytuje požadovanou (aposteriorní) hp.

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

$$\boxed{f(\Theta|D(t)) \propto L_t(\Theta) f(\Theta),} \tag{2.13}$$

kde $L_t(\Theta)$ je věrohodnostní funkce (likelihood)

$$L_t(\Theta) = \prod_{\tau=1}^{t} f(y_\tau|\varphi_\tau, \Theta), \tag{2.14}$$

kterou chápeme jako funkci parametrů Θ, za data jsou dosazeny změřené hodnoty.

Tyto vztahy dostaneme postupným dosazováním rekurentního vzorce (2.12) pro časové okamžiky $\tau = 1, 2, ..t$.

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

Algoritmus

Predikce výstupu a odhad parametrů

Start *apriorní hp*

$$f(\Theta|D(0)) = f(\Theta)$$

For $\tau = 1$ **to** t **do**

predikce

$$f(y_\tau|u_\tau, D(\tau-1)) = \int f(y_\tau|\varphi_\tau, \Theta) f(\Theta|D(\tau-1)) \mathrm{d}\tau$$

identifikace

$$f(\Theta|D(\tau)) = \frac{f(y_\tau|\varphi_\tau, \Theta) f(\Theta|D(\tau-1))}{f(y_\tau|u_\tau, D(\tau-1))}$$

End for

2.3 Příklady

V následujících příkladech ukážeme, jak probíhá predikce a identifikace s konkrétními tvary modelů. Budeme sledovat ty typy modelů, o kterých jsme hovořili v příkladové části předchozí kapitoly. Uvidíme, že konečný algoritmus identifikace je velmi podobný algoritmům pro deterministický model, uvedeným v kap. 2 minulé části.

PŘÍKLAD 2.1 Regresní model řádu ν

V tomto příkladě navážeme na příklad ze str. 8 a budeme demonstrovat postup identifikace pro nejběžnější případ skalárního regresního modelu s obecným řádem ν s rovnicí (1.6)

$$y_t = \varphi_t^T \theta + e_t$$

a hp modelu (1.10)

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

Součin modelů

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

$$\prod_{\tau=1}^{t} f(y_t|\varphi_t, \Theta) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^t \exp\left\{-\frac{1}{2\sigma^2}\sum_{\tau=1}^{t}(y_\tau - \varphi_\tau^T\theta)^2\right\} \tag{2.15}$$

Věrohodnostní funkce

Součin modelů vyjádřený jako funkce parametrů je věrohodnostní funkce. Zavedeme-li sloupcové vektory $[-1, \theta^T]^T$ a $[y_t, \varphi_t^T]^T$, můžeme předchozí součin modelů uspořádat jako funkci prvního z vektorů, tedy parametrů. Jde o stejný postup a značení jako jsme použili v Příkladu 2.3 kapitoly 2. o identifikaci v předchozím díle této práce

$$L_t(\Theta) = (2\pi\sigma^2)^{-\frac{t}{2}} \exp\left\{-\frac{1}{2\sigma^2} \begin{bmatrix} -1 & \theta^T \end{bmatrix} \sum_{\tau=1}^{t} \left(\begin{bmatrix} y_\tau \\ \varphi_\tau \end{bmatrix} \begin{bmatrix} y_\tau & \varphi_\tau^T \end{bmatrix} \right) \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}. \tag{2.16}$$

Apriorní hp

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

$$f(\Theta|D(0)) \propto \sigma^{-\frac{\kappa(0)}{2}} \exp\left\{-\frac{1}{2\sigma^2} \begin{bmatrix} -1 & \theta^T \end{bmatrix} V(0) \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}, \tag{2.17}$$

kde $V(0)$ je pozitivně definitní symetrická čtvercová matice odpovídajícího stupně a κ_0 je skalár.

Aposteriorní hp

Tato hp je normovaným součinem věrohodnostní funkce a apriorní hp

$$f(\Theta|D(t)) = \text{konst.}\, L_t(\Theta)\, f(\Theta|D(0)).$$

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

$$f(\Theta|D(t)) \propto \sigma^{-\frac{\kappa(t)}{2}} \exp\left\{-\frac{1}{2\sigma^2} \begin{bmatrix} -1 & \theta^T \end{bmatrix} V(t) \begin{bmatrix} -1 \\ \theta \end{bmatrix} \right\}, \tag{2.18}$$

kde

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

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

Označíme

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

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

$$\Lambda_t = V_y(t) - V_{y\varphi}^T(t)V_\varphi^{-1}(t)V_{y\varphi}(t), \quad (2.19)$$
$$\hat{\theta}_t = V_\varphi^{-1}(t)V_{y\varphi}(t). \quad (2.20)$$

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

$$\hat{\theta}_t = V_\varphi^{-1}(t)V_{y\varphi}(t) \qquad \text{odhad regresních koeficientů,} \quad (2.21)$$
$$\hat{\sigma}_t^2 = \Lambda_t/\kappa(t) \qquad \text{odhad rozptylu šumu.} \quad (2.22)$$

Poznámka: *Odhad kovariance parametrů je*

$$C_t = V_\varphi^{-1}(t).$$

Prediktivní hp

Tato hp je dána vztahem (2.10)

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

Výsledek integrace a tedy prediktivní hp je

$$f(y_t|u_t, D(t-1)) = \frac{\Gamma\left(\frac{\kappa_t - \rho + 2}{2}\right)}{2^{\frac{\kappa_t + 2}{2}} \pi^\rho} |V_{\varphi,t}|^{\frac{1}{2}} \Lambda_{t-1}^{\frac{\kappa_t - \rho + 3}{2}}, \quad (2.23)$$

kde ρ je dimenze regresního vektoru φ_t^T.

Poznámka: *Odvození uvedeného vztahu je uvedeno v Doplňku.*

Bodová predikce výstupu

Bodovou predikci výstupu dostaneme jako střední hodnotu y_t, podmíněnou staršími daty (nikoliv parametry). Podle (2.9) je

$$\hat{y}_t = \hat{\theta}_{t-1}^T \varphi_t, \quad (2.24)$$

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

Poznámka: *V tomto příkladě uvádíme nejjednodušší způsob výpočtů charakteristik hp, ze kterých počítáme bodové odhady. Postupně v čase (nebo najednou) počítáme rozšířenou informační matici V, tu rozdělíme na submatice $V_y, V_{y\varphi}$a V_φ. Z nich počítáme odhady parametrů a predikci výstupu. Existují ale i další algoritmy, které přepočítávají přímo odhady a jejich charakteristiky, bez počítání informační matice. O jednom z nich se zmíníme v kapitole o řízení při neznámých parametrech modelu.*

Algoritmus

Bayesovská identifikace

(regresní model, normální šum)

Off-line: před sběrem dat

`Připrav nebo změř apriorní data`

$$y_0, u_0, y_{-1}, u_{-1}, ...y_{-n}, u_{-n}$$

`Sestav apriorní statistiky` (např. pomocí apriorních dat)

$$\kappa, \qquad V$$

On-line: sběr dat

For $\tau = 1$ **to** t **do**

`Generuj vstup a změř výstup`

$$u_\tau = ..., \qquad y_\tau = ...$$

`Sestav rozšířený regresor`

$$\check{\varphi}^T = [y_\tau, u_\tau, ...y_{\tau-n}, u_{\tau-n}]$$

`Proveď aktualizaci statistik`

$$V = V + \check{\varphi}\check{\varphi}^T, \qquad \kappa = \kappa + 1$$

end for

Off-line: po sběru dat

`Rozděl informační matici`

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

`Vypočti`

– zbytek kriteria

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

– odhad parametrů

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

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

PŘÍKLAD 2.2 Statický regresní model s normálním šumem

A. Model se známým rozptylem

Statický model je velmi jednoduchý. Není pro něho třeba zavádět vektorovou symboliku, tak jako v případě regresního modelu obecného řádu, a odvození identifikačního algoritmu je průhlednější. Jak uvidíme, budeme pracovat v podstatě s prvním a druhým momentem datového výběru, tj. s průměrem a rozptylem výběru.

Identifikace tohoto modelu samozřejmě probíhá rovněž podle již uvedeného algoritmu. Jeho jednoduchost však umožňuje celé odvození demonstrovat jednodušším postupem. Budeme postupovat tak, že nejprve pomocí hp modelu a aposteriorní hp z minulého kroku určíme sdruženou hp pro výstup y_t a parametry Θ. Z ní pak lze doplněním exponentu normálního rozdělení na čtverec určit vše ostatní – aposteriorní i prediktivní hp. Budeme tedy sledovat následující schéma pro hp v jednom časovém okamžiku t

$$\underbrace{f(y_t|\theta)}_{\text{model}}\underbrace{f(\theta|D(t-1))}_{\text{apost. } t-1} = \underbrace{f(y_t,\theta|D(t-1))}_{\text{sdružená}} = \underbrace{f(\theta|D(t))}_{\text{apost. } t}\underbrace{f(y_t|D(t-1))}_{\text{prediktivní}} \tag{2.25}$$

Jednotlivé vzorce budou představovat speciální případ pro vzorce z předchozího příkladu pro obecný regresní model.

Model

Budeme uvažovat model s rovnicí (1.17)

$$y_t = \theta + e_t, \tag{2.26}$$

s hp podle (1.18). Je to model s jediným neznámým parametrem θ, protože rozptyl uvažujeme známý $\sigma^2 = 1$. Je tedy $\Theta = \theta$.

Součin modelů

Uvedeme výraz pro součin modelů od času 1 obecně do času ϑ.

$$\begin{aligned} L_t(\theta) = \prod_{\tau=1}^{\vartheta} f(y_\tau|\theta) &= (2\pi)^{-\frac{\vartheta}{2}} \exp\left\{-\frac{1}{2}\sum_{\tau=1}^{\vartheta}(y_\tau^2 - 2\theta y_\tau + \theta^2)\right\} = \\ &= (2\pi)^{-\frac{\vartheta}{2}} \exp\left\{-\frac{1}{2}(\vartheta\theta^2 - 2\theta\sum_{\tau=1}^{\vartheta} y_\tau + \sum_{\tau=1}^{\vartheta} y_\tau^2)\right\} \end{aligned} \tag{2.27}$$

Z předešlého výrazu je zřejmé, že exponent součinu modelů lze upravit jako kvadratickou formu v proměnné θ a celý součin je určen následujícími statistikami

$$\vartheta \qquad \text{počet dat} \tag{2.28}$$

$$\sum_{\tau=1}^{\vartheta} y_\tau \qquad \text{součet dat} \tag{2.29}$$

$$\sum_{\tau=1}^{\vartheta} y_\tau^2 \qquad \text{součet kvadrátů dat} \tag{2.30}$$

Apriorní hp

Podle předchozího vyjádření součinu modelů a v souladu s regresním modelem obecného řádu volíme apriorní hp v tomto tvaru

$$\boxed{f(\theta|D(0)) \propto \exp\left\{-\tfrac{1}{2}\left(\kappa(0)\theta^2 - 2\theta S_{y1}(0) + S_y^2(0)\right)\right\}} \tag{2.31}$$

kde $\kappa(0)$, $S_{y1}(0)$, $S_y^2(0)$ jsou statistiky této hp (jednička v indexu druhé statistiky formálně označuje u, které je zde jedna).

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

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

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.

Aposteriorní hp v čase $t-1$

Aposteriorní hp má stejný tvar jako apriorní, se statistikami v příslušném časovém okamžiku, zde $t-1$

$$f(\theta|D(t-1)) \propto \exp\left\{-\frac{1}{2}\left(\kappa(t-1)\theta^2 - 2\theta S_{y1}(t-1) + S_y^2(t-1)\right)\right\} = \tag{2.32}$$

$$= \exp\left\{-\frac{1}{2}\left[\kappa(t-1)\left(\theta - \hat{\theta}_{t-1}\right)^2 + \Lambda_{t-1}\right]\right\} \tag{2.33}$$

kde

$\hat{\theta}_{t-1} = \frac{S_{y1}(t-1)}{\kappa(t-1)}$

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

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

pro $\kappa(0) = 0$, $S_{y1}(0) = 0$ a $S_y^2(0) = 0$

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

P o z n á m k a: *Při použití nenulových apriorních statistik bychom je museli k uvedeným statistikám ještě přičíst jako v (2.18).*

Sdružená hp

Podle schématu hp (2.25) konstruujeme nyní sdruženou hp pro výstup y_t a parametr θ, vynásobením hp modelu a aposteriorní hp pro čas $t-1$.

$$\begin{aligned} f(y_t,\theta|D(t)) &\propto (2\pi)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(y_t-\theta)^2\right\}(2\pi)^{-\frac{\kappa(t-1)}{2}}\cdot \\ &\quad\cdot\exp\left\{-\frac{1}{2}\left[\kappa(t-1)\left(\theta-\hat{\theta}_{t-1}\right)^2+\Lambda_{t-1}\right]\right\}\propto \qquad (2.34) \\ &\propto \exp\left\{-\frac{1}{2}\left[(y_t-\theta)^2+\kappa(t-1)\left(\theta-\hat{\theta}_{t-1}\right)^2+\Lambda_{t-1}\right]\right\}, \qquad (2.35) \end{aligned}$$

kde v posledním výrazu jsme vynechali konstantu.

Upravíme exponent tak, že jej doplníme na čtverec v proměnné θ.

$$y_t^2-2\theta y_t+\theta^2+\kappa(t-1)\left(\theta^2-2\theta\hat{\theta}_{t-1}+\hat{\theta}_{t-1}^2\right)+\Lambda_{t-1}=$$

$$=(\kappa(t-1)+1)\,\theta^2-2\theta\left(y_t+\kappa(t-1)\hat{\theta}_{t-1}\right)+y_t^2+\kappa(t-1)\hat{\theta}_{t-1}^2+\Lambda_{t-1}=$$

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

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

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 (2.20) při $\kappa(0)=0$ platí $\hat{\theta}_t=\frac{\sum_{\tau=1}^t y_t}{\kappa(t)}$

a označíme $\Lambda_t=\Lambda_{t-1}+\frac{\kappa(t-1)}{\kappa(t)}(y_t-\hat{\theta}_{t-1})^2$.

Po této úpravě exponentu bude mít sdružená hp následující tvar

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

$$=\exp\left\{-\frac{1}{2}\kappa(t)\left(\theta-\hat{\theta}_t\right)^2\right\}\exp\left\{-\frac{1}{2}\frac{\kappa(t-1)}{\kappa(t)}\left(y_t-\hat{\theta}_{t-1}\right)^2\right\}\exp\left\{-\frac{1}{2}\Lambda_{t-1}\right\} \qquad (2.36)$$

Tímto jsme se dostali k hlavnímu bodu celého příkladu, ve kterém jsme provedli přepočet (2.25). Tento přepočet je dán rovností (2.34) = (2.35) = (2.36). 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 kriteria nejmenších čtverců.

Výraz v (2.34) 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 (2.35). Poslední výraz (2.36) 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 prediktivní hp, předpovídající výstup jen na základě starých dat, nikoliv parametrů.

Aposteriorní hp v čase t

Aposteriorní hp v čase t má podle předchozího odvození normální rozdělení a po normování má tvar

$$\boxed{f(\theta|D(t)) = \frac{1}{\sqrt{2\pi/\kappa(t)}} \exp\left\{-\frac{1}{2}\kappa(t)\left(\theta - \hat{\theta}_t\right)^2\right\}} \tag{2.37}$$

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

Bodové odhady parametrů

Bodové odhady parametrů dostaneme jako podmíněnou střední hodnotu parametrů a spočteme ji 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

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{\kappa(t)}\left(y_t - \hat{\theta}_{t-1}\right) = \frac{S_{y1}(t)}{\kappa(t)} = \frac{\sum_{\tau=-\kappa(0)}^{t} y_\tau}{\kappa(0)+t} \tag{2.38}$$

kde

$\hat{\theta}_0$ je apriorní odhad parametru θ,

$\kappa(0)$ 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).

Prediktivní hp

Tato hp je určena marginálou ze sdružené hp (2.35) po doplnění exponentu na čtverec v proměnné θ. Oddělíme podmíněnou hp a po normování dostaneme

$$\boxed{f(y_t|D(t-1)) = \frac{1}{\sqrt{2\pi\frac{\kappa(t)}{\kappa(t-1)}}}\exp\left\{-\frac{1}{2}\frac{\kappa(t-1)}{\kappa(t)}\left(y_t-\hat{\theta}_{t-1}\right)^2\right\}} \tag{2.39}$$

Ta část sdružené hp (2.36), obsahující funkci Λ_{t-1} je zahrnuta v konstantě.

Bodová predikce

Bodová predikce je bodovým odhadem budoucího výstupu, vzhledem k změřeným datům. Je to opět střední hodnota výstupu, určená pomocí prediktivní hp

$$\hat{y}_t = \hat{\theta}_{t-1} = \frac{1}{\kappa(t-1)}\sum_{\tau=-\kappa(0)}^{t-1} y_\tau \tag{2.40}$$

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

Rozptyl bodové predikce je

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

♣ Rozptyl bodové predikce postupně klesá (čím více máme dat, tím přesněji jsme schopni predikovat) a ustaluje se na jedné (s hodně velkým počtem dat máme jistotu 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 kapitole o základech identifikace) je funkce Λ zbytek kriteria 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. Podle označení pod vztahem (2.33) a významu statistik v (2.31) je

$$\Lambda_t = S_y^2(t) - \frac{(S_{y1}(t))^2}{\kappa(t)} = \sum_{\tau=1}^{t} y_\tau^2 - \frac{\left(\sum_{\tau=1}^{t} y_\tau\right)^2}{t} = t\overline{y_\tau^2} - t\bar{y}^2 =$$

$$= t\overline{(y-\bar{y})^2} = \sum_{\tau=1}^{t}(y_\tau - \bar{y})^2 \qquad (2.41)$$

kde

$\bar{y} = \frac{1}{t}\sum_{\tau=1}^{t} y_\tau$ je operátorová symbolika pro výběrový průměr,

$S_y^2(t) = t\cdot\overline{y^2}$ a $S_{y1}(t) = t\cdot\bar{y}$ 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 reziduum v bodě τ pro parametry odhadnuté z dat až do času t. Tvrzení ale platí obecně.

♣ 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 součinu* 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* $\sum_{\tau=1}^{t} y_\tau = t\cdot\bar{y}$.

První člen posledního výrazu odpovídá exponentu aposteriorní hp, druhý člen exponentu prediktivní hp. Exponent této hp tvoří funkce Lambda v čase t.

B. Model s neznámým rozptylem

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

Model

Model v tomto příkladě bude mít stejnou rovnici jako v předchozím příkladu

$$y_t = \theta + e_t, \tag{2.42}$$

ale hp modelu bude jako v (1.16)

$$f(y_t|\Theta) = \frac{1}{\sqrt{2\pi r}} \exp\left\{-\frac{1}{2r}(y_t^2 - 2\theta y_t + \theta^2)\right\}. \tag{2.43}$$

s neznámými parametry $\Theta = [\theta, r]$.

Věrohodnostní funkce

$$L_t(\Theta) \propto r^{-\frac{t}{2}} \exp\left\{-\frac{1}{2r}\left(t\theta^2 - 2\theta\sum_{\tau=1}^{t} y_\tau + \sum_{\tau=1}^{t} y_\tau^2\right)\right\} \tag{2.44}$$

Aposteriorní hp pro t

$$\boxed{f(\Theta|D(t)) \propto r^{-\frac{\kappa(t)}{2}} \exp\left\{-\tfrac{1}{2r}\left[\kappa(t)\left(\theta - \hat{\theta}_t\right)^2 + \Lambda_t\right]\right\}} \tag{2.45}$$

Dalším potřebným krokem je integrace přes neznámé parametry. Tak dostaneme prediktivní hp. Nejdříve budeme integrovat přes θ (to je stejný krok jako v případě známého rozptylu), potom přes rozptyl r.

I_1 – integrál přes θ

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

I_2 – integrál přes $\sigma^2 = r$

$$\begin{aligned} I_2 &\propto \kappa(t)^{-\frac{1}{2}} \int_0^{\infty} r^{-\frac{\kappa(t)-1}{2}} \exp\left\{-\frac{\Lambda_t}{2r}\right\} \mathrm{d}r = \\ & \left| \begin{array}{cc} \frac{\Lambda_t}{2r} = z, & r = \frac{\Lambda_t}{2z} \\ -\frac{\Lambda_t}{2r^2}\mathrm{d}r = \mathrm{d}z, & \text{meze: } \infty, .., 0 \end{array} \right| \\ &= \kappa(t)^{-\frac{1}{2}} \int_0^{\infty} \left(\frac{\Lambda_t}{2z}\right)^{-\frac{\kappa(t)-1}{2}} \exp\{-z\} \frac{2}{\Lambda_t}\left(\frac{\Lambda_t}{2z}\right)^2 \mathrm{d}z = \\ &= \kappa(t)^{-\frac{1}{2}} \int_0^{\infty} \left(\frac{\Lambda_t}{2}\right)^{-\frac{\kappa(t)-1}{2}+2-1} z^{\frac{\kappa(t)-1}{2}-2} \exp\{-z\} \mathrm{d}z = \end{aligned}$$

$$
\begin{aligned}
&= \kappa(t)^{-\frac{1}{2}} \left(\frac{\Lambda_t}{2}\right)^{-\frac{\kappa(t)-3}{2}} \int_0^\infty z^{\frac{\kappa(t)-3}{2}-1} \exp\{-z\}\, \mathrm{d}z = \\
&= \kappa(t)^{-\frac{1}{2}} \left(\frac{\Lambda_t}{2}\right)^{-\frac{\kappa(t)-3}{2}} \Gamma\left(\frac{\kappa(t)-3}{2}\right) = \\
&= \kappa(t)^{-\frac{1}{2}} 2^{\frac{\kappa(t)-3}{2}} \Gamma\left(\frac{\kappa(t)-3}{2}\right) \Lambda_t^{-\frac{\kappa(t)-3}{2}}
\end{aligned}
$$

Prediktivní hp

Vyjádříme-li Λ_t rekurzivně, pomocí Λ_{t-1}, můžeme prediktivní hp vyjádřit ve formě Studentova rozdělení jako funkci proměnné y_t.

$$
\boxed{f(y_t|D(t-1)) \propto \kappa(t)^{-\frac{1}{2}} 2^{\frac{\kappa(t)-3}{2}} \Gamma\left(\frac{\kappa(t)-3}{2}\right) \left(\frac{\kappa(t-1)}{\kappa(t)}\left(y_t - \hat{\theta}_{t-1}\right)^2 + \Lambda_{t-1}\right)^{-\frac{\kappa(t)-3}{2}}} \qquad (2.46)
$$

PŘÍKLAD 2.3 Statický regresní model s rovnoměrným šumem

V tomto příkladě ukážeme strukturu identifikačních vzorců pro ne tak často používaný, ale velmi 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 identifikace pro hp je stejné jako pro normální šum, konstrukce statistik pro akumulaci dat je značně odlišná.

Model

Identifikovanou soustavu popíšeme modelem (1.17)

$$y_t = \theta + e_t,$$

s hp (1.18)

$$f(y_t|\theta) = \frac{1}{2}\chi(y_t \in [\theta-1;\ \theta+1]).$$

Součin modelů

Určíme součin hp modelů v časech $\tau = 1, 2, ..., t$, abychom odhadli strukturu apriorní a aposteriorní hp, které s tímto součinem musí být strukturálně shodné.

$$\prod_{\tau=1}^{t} f(y_t|\theta) = \prod_{\tau=1}^{t} \frac{1}{2}\chi(y_\tau \in [\theta-1;\ \theta+1]) = \frac{1}{2^t}\chi(\text{"podmínka"}),$$

kde "podmínka" je dána soustavou nerovnic

$$
\begin{array}{ccc}
y_1 \geq \theta - 1 & \wedge & y_1 \leq \theta + 1 \\
y_2 \geq \theta - 1 & \wedge & y_2 \leq \theta + 1 \\
\dots & & \dots \\
y_t \geq \theta - 1 & \wedge & y_t \leq \theta + 1.
\end{array}
$$

Soustavu lze shrnout takto

$$\min\{y_i\} \geq \theta - 1 \qquad \wedge \qquad \max\{y_i\} \leq \theta + 1$$

odkud

$$\max\{y_i\} - 1 \leq \theta \leq \min\{y_i\} + 1 \quad \Rightarrow \quad \theta \in [\max\{y_i\} - 1;\ \min\{y_i\} + 1].$$

Věrohodnostní funkce

Součin hp modelů vyjádříme jako funkci parametrů a dostaneme věrohodnostní funkci

$$L_t(\theta) = \frac{1}{2^t}\chi(\theta \in [\max\{y_i\} - 1;\ \min\{y_i\} + 1]). \tag{2.47}$$

P o z n á m k a: *Je-li* $\max\{y_i\} - \min\{y_i\} > 2$, *je* $L_t(\theta) = 0$*! Proč?* [1]

Apriorní hp

Apriorní hp volíme s tak, aby měla strukturu věrohodnostní funkce, tj. rovnoměrnou, s apriorními statistikami B_0^L a B_0^U.

$$\boxed{f(\theta) = \tfrac{1}{B_0^U - B_0^L}\chi(\theta \in [B_0^L;\ B_0^U])} \tag{2.48}$$

Aposteriorní hp

Součin apriorní hp a věrohodnostní funkce, podle (**??**) dá aposteriorní hp

$$\boxed{f(\theta|y(t)) = \tfrac{1}{B_t^U - B_t^L}\chi(\theta \in [B_t^L;\ B_t^U]),} \tag{2.49}$$

kde

$$B_t^L = \max\{B_0^L;\ \max\{y_\tau\}_{\tau=1}^t - 1\} = \max\{B_{t-1}^L;\ y_t - 1\},$$

$$B_t^U = \min\{B_0^L;\ \min\{y_\tau\}_{\tau=1}^t + 1\} = \min\{B_{t-1}^U;\ y_t + 1\},$$

s počáteční podmínkou B_0^L a B_0^U.

[1] Protože interval v argumentu indikátorové funkce v (2.47) je prázdný.

Bodový odhad parametrů

Zajímáme-li se o bodový odhad parametru, můžeme jej určit jako střední hodnotu s hp (2.49). Odvození je uvedeno v Doplňku.

$$\hat{\theta} = E[\theta|y(t)] = \int \theta f(\theta|y(t))\mathrm{d}\theta = \frac{B_t^L + B_t^U}{2}. \tag{2.50}$$

Výsledek je takový, jaký bychom očekávali. Při rovnoměrném rozdělení výskytu parametru na intervalu $[B_t^L, B_t^U]$ budeme odhadovat jeho správnou hodnotu uprostřed.

Prediktivní hp

Hledáme hp, která bude předpovídat budoucí výstup na základě minulých dat. Dostaneme ji podle obecného schématu (**??**).

$$f(y_t|y(t-1)) = \int f(y_t|\theta)f(\theta|y(t-1))\mathrm{d}\theta =$$

$$= \begin{cases} \frac{1}{2}\frac{y_t+1-B_{t-1}^L}{B_{t-1}^U-B_{t-1}^L} & \text{pro} \quad y_t \in [B_{t-1}^L - 1;\ B_{t-1}^U - 1] \\ \frac{1}{2} & \text{pro} \quad y_t \in [B_{t-1}^U - 1;\ B_{t-1}^L + 1] \\ \frac{1}{2}\frac{B_{t-1}^U+1-y_t}{B_{t-1}^U-B_{t-1}^L} & \text{pro} \quad y_t \in [B_{t-1}^L + 1;\ B_{t-1}^U + 1] \\ 0 & \text{jinak.} \end{cases} \tag{2.51}$$

Odvození a průběh prediktivní hp jsou uvedeny v Doplňku.

Bodová predikce výstupu

Je to střední hodnota výstupu, podmíněná minulými výstupy. Vypočteme ji použitím prediktivní hp (odvození je uvedeno v Doplňku)

$$\hat{y}_t = E[y_t|y(t-1)] = \int y_t f(y_t|y(t-1))\mathrm{d}y_t = \hat{\theta}_{t-1}. \tag{2.52}$$

Výsledek je opět to, co lze čekat. Výstup je modelován jako konstanta (parametr θ) na kterou je superponován šum s nulovou střední hodnotou. Nejlepším odhadem výstupu je tedy nejlepší odhad $\hat{\theta}_{t-1}$ parametru θ.

PŘÍKLAD 2.4 Statický regresní model s diskrétním šumem

Příklad se zabývá 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 $\tau = 1, 2, ..t$ bude

$$y_\tau \in \{0, 1\}, \qquad f(0) = P(V_\tau = R) = \theta \ \text{ a } \ f(1) = P(V_\tau = L) = 1 - \theta,$$

kde

– parametr modelu $\theta \in (0,\ 1)$ značí pravděpodobnost padnutí rubu (R) a

– f je pravděpodobnostní funkce diskrétní náhodné veličiny y.

Měřená data jsou $D(t) = y(t) = \{y_t, y_{t-1}, y_{t-2}, ...\}$.

Model

Popsaný proces je modelován hp (1.21). Pro její přehlednější vyjádření a jednodušší rozšíření na náhodné veličiny, s více než dvěma hodnotami, vyjádříme tuto hp pomocí diskrétní *indikátorové funkce* δ (Kroneckerovo delta). Tato funkce je analogií indikátorové funkce χ, použité v předchozím příkladě. Je definována pro diskrétní argument takto

$$\delta(i,j) = \begin{cases} 1 & \text{pro } i = j, \\ 0 & \text{pro } i \neq j \end{cases} . \tag{2.53}$$

S takto zavedeným značením můžeme hp modelu zapsat takto

$$f(y_\tau|\theta) = \theta^{\delta(y_\tau,0)}(1-\theta)^{\delta(y_\tau,1)}. \tag{2.54}$$

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 κ_t^R - počet padlých rubů a κ_t^L - počet padlých líců. Pomocí indikátorové funkce je můžeme vyjádřit takto

$$\kappa_t^R = \sum_{\tau=1}^{t} \delta(y_\tau, 0) \quad \text{a} \quad \kappa_t^L = \sum_{\tau=1}^{t} \delta(y_\tau, 1)$$

Věrohodnostní funkce

S použitím zavedených statistik můžeme součin modelů vyjádřit jako funkci parametrů – věrohodnostní funkci

$$L_t(\theta) = \theta^{\kappa_t^R}(1-\theta)^{\kappa_t^L} \tag{2.55}$$

Apriorní hp

Tuto hp volíme v podobném tvaru jako likelihood

$$\boxed{f(\theta) = \frac{1}{K_0}\theta^{\kappa_0^R-1}(1-\theta)^{\kappa_0^L-1}} \tag{2.56}$$

kde κ_0^R a κ_0^L jsou apriorní statistiky a $1/K_0$ je normalizační konstanta.

`Poznámka:` *Minus jednička v exponentech je zavedena formálně pro pozdější vhodné označení pomocí Beta funkce.*

`Poznámka:` *Jiná možná volba statistik je např. počet padlých rubů κ^R a počet pokusů t.*

Aposteriorní hp

Vynásobením apriorní hp a věrohodnostní funkce dostáváme

$$\boxed{f(\theta|y(t)) = \tfrac{1}{K}\theta^{\kappa_t^R-1}(1-\theta)^{\kappa_t^L-1}} \tag{2.57}$$

kde

$\kappa_\bullet^R$ a $\kappa_\bullet^L$ jsou statistiky, pro které platí rekurze

$$\kappa_t^R = \kappa_{t-1}^R + \delta(y_t, 0)$$

$$\kappa_t^L = \kappa_{t-1}^R + \delta(y_t, 1)$$

s počátečními hodnotami κ_0^R a κ_0^L; tj. po hodu mincí přičteme jedničku k té statistice, jejíž strana padla,

$K = \int \theta^{\kappa_t^R-1}(1-\theta)^{\kappa_t^L-1}\mathrm{d}\theta = B(\kappa_t^R, \kappa_t^L)$,

$B(.,.)$ je Beta funkce, která je popsána v Doplňku části I.

Bodový odhad parametrů

$$\hat{\theta}_t = E[\theta|y(t)] = \int \theta f(\theta|y(t))\mathrm{d}\theta = \frac{\kappa_t^R}{\kappa_t^R + \kappa_t^L} \tag{2.58}$$

Odvození tohoto vztahu je uvedeno v Doplňku.

Prediktivní hp

Integrací sdružené hp pro výstup a parametr, dané pomocí hp modelu a apriorní hp z kroku $t-1$, dostaneme prediktivní hp

$$f(y_t|y(t-1)) = \int f(y_t|\theta)f(\theta|y(t-1))\mathrm{d}\theta =$$

$$= \begin{cases} \frac{\kappa_{t-1}^R}{\kappa_{t-1}^R+\kappa_{t-1}^L} = \hat{\theta}_{t-1} & \text{pro } y_t = 0 \\ \\ \frac{\kappa_{t-1}^L}{\kappa_{t-1}^R+\kappa_{t-1}^L} = 1-\hat{\theta}_{t-1} & \text{pro } y_t = 1 \end{cases} \tag{2.59}$$

Bodová predikce výstupu

Je to střední hodnota výstupu vypočtená pomocí prediktivní hp. Tato hp je diskrétní, proto místo integrace musíme použít sumaci

$$\hat{y}_t = E[y_t|y(t-1)] = \sum_{y_t=0}^{1} y_t f(y_t|y(t-1)) = 0{\cdot}\hat{\theta}_{t-1} + 1{\cdot}(1-\hat{\theta}_{t-1}) = 1-\hat{\theta}_{t-1}. \tag{2.60}$$

Poznámka: *Pokud bychom chtěli, aby predikce byla skutečně existující hodnotou výstupu, mohli bychom za predikci vzít maximálně pravděpodobnou hodnotu. Kriterium optimality je v tomto případě argument maxima (tj. modus) prediktivní hp*

$$\hat{y}_t = \begin{cases} 0 & \text{pro } \hat{\theta}_{t-1} \geq 0.5 \\ 1 & \text{pro } \hat{\theta}_{t-1} < 0.5 \end{cases}$$

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 predikce bude trvale ta strana, která má vyšší pravděpodobnost. Tento výsledek konec konců odpovídá zdravému rozumu.

PŘÍKLAD 2.5 Modely s diskrétními parametry

V tomto příkladě budeme sledovat trochu umělý příklad modelu s diskrétním parametrem. O parametru modelu, který neznáme, předpokládáme, že může nabývat jen konečný počet různých hodnot. Pravděpodobnosti těchto hodnot předpokládáme známé. (Identifikací takových pravděpodobností jsme se zabývali v předchozím příkladě o identifikaci spojitého parametru z diskrétních dat.) Známe tedy rozdělení pravděpodobnosti parametrů (stará aposteriorní hp) a naším úkolem je před změřením nového výstupu určit jeho predikci a na základě změřeného výstupu provést přepočet 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, uvedené v minulé kapitole. Jsou to modely diskrétním parametrem a jak s diskrétním, tak i se spojitým signálem.

A. Diskrétní parametr i signál

Model

Pro popis soustavy vezmeme model (1.22)

$$y = \theta + e,$$

s diskrétním parametrem θ jehož rozdělení je zadáno tabulkou

θ	-2	0	2
$f(\theta)$	0.2	0.5	0.3

Tento model má podmíněnou diskrétní hp (1.23)

$f(y\|\theta)$ $y\backslash\theta$	-1	0	1
-3	0.4	0	0
-1	0.6	0.4	0
1	0	0.6	0.4
3	0	0	0.6

Prediktivní hp

Je to hp výstupu, která není podmíněná znalostí parametrů

$$f(y) = \sum_{\theta} f(y|\theta)f(\theta) = \tag{2.61}$$

$$= f(y|\theta=-2)f_\theta(-2) + f(y|\theta=0)f_\theta(0) + f(y|\theta=2)f_\theta(2) \tag{2.62}$$

`Poznámka:` *Z rozepsaného vzorce je dobře vidět princip Bayesovské statistiky. Kdybychom znali správnou hodnotu θ, byl by výstup predikován jedinou hp. Protože správnou hodnotu θ neznáme, je pro nás náhodnou veličinou, pro kterou známe pouze množinu všech možných hodnot a jejich pravděpodobnosti. Proto vezmeme v úvahu všechny podmíněné hp, vážíme je jejich pravděpodobnostmi $f(\theta)$ a ze všech dohromady sestavíme predikci.*

Zavedené hp lze výhodně reprezentovat pomocí vektorů a matic.

$$f(y|\theta) = \begin{bmatrix} 0.4 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 0.6 \end{bmatrix} \quad \text{a} \quad f(\theta) = \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix}$$

Potom předchozí vzorce lze vyjádřit jako součin matice a vektoru

$$f(y) = \sum_{j=1}^{3} f(y|\theta_j)f(\theta_j) = \begin{bmatrix} 0.4 & 0 & 0 \\ 0.6 & 0.4 & 0 \\ 0 & 0.6 & 0.4 \\ 0 & 0 & 0.6 \end{bmatrix} \begin{bmatrix} 0.2 \\ 0.5 \\ 0.3 \end{bmatrix} =$$

$$= 0.2f(y|\theta=-1) = 0.5f(y|\theta=0) = 0.3f(y|\theta=1) = \begin{bmatrix} 0.08 \\ 0.32 \\ 0.42 \\ 0.18 \end{bmatrix}$$

♣ Prediktivní pravděpodobnost je vážená kombinace jednotlivých modelů (hp), váhy jsou pravděpodobnosti parametrů (tj. pravděpodobnosti jednotlivých modelů).

Bodová predikce výstupu

$$\hat{y} = E[y] = \sum_{y} y f(y) = [-3, -1,\ 1,\ 3] \begin{bmatrix} 0.08 \\ 0.32 \\ 0.42 \\ 0.18 \end{bmatrix} = 0.4 \tag{2.63}$$

♣ Bodová predikce je vážená kombinace všech hodnot y, váhy jsou prediktivní pravděpodobnosti.

`Poznámka:` *Kdyby se požadovalo, aby predikovaná hodnota byla skutečně přípustnou hodnotou, mohli bychom jako predikci vzít hodnotu s maximální pravděpodobností*

$$\hat{y} = \arg\max f(y) = 1$$

Odhad parametru

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 $f(y,\theta) = f(y|\theta)f(\theta)$; marginální $f(y) = \sum_\theta f(y|\theta)$. Sdruženou hp počítáme tak, že vždy celý sloupec hp modelu násobíme příslušným prvkem vektoru $f(\theta)$; marginální je součet prvků v řádce sdružené hp.

$f(y,\theta)$				$f(y)$
$y\backslash\theta$	-1	0	1	
-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

Hledáme: $f(\theta|y=1)$. Tu určíme z podmíněné hp $f(\theta|y) = \frac{f(y|\theta)}{f(y)}$, kterou dostaneme tak, že normujeme řádky sdružené hp, tj. každý řádek sdružené hp dělíme příslušným prvkem marginální hp $f(y)$. Dostaneme

$f(\theta|y)$					
$y\backslash\theta$	-1	0	1		
-3	1	0	0	$\longleftarrow$	$f(\theta|y=-3)$
-1	0.375	0.625	0	$\longleftarrow$	$f(\theta|y=-1)$
1	0	0.714	0.286	$\longleftarrow$	$f(\theta|y=1)$
3	0	0	1	$\longleftarrow$	$f(\theta|y=3)$

Z druhého řádku tabulky je

θ	-1	0	1
$f(\theta|y=-1)$	0.375	0.625	0

Odkud je odpověď na naši otázku:

♣ Hodnota výstupu $y = -1$ vzešla nejpravděpodobněji z hodnoty $\theta = 0$ (s pravděpodobností 0.625), nebo méně pravděpodobně z hodnoty $\theta = -1$ (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 spojitý signál

Dále si všimneme práce s modelem, který má diskrétní parametr, ale pracuje se spojitými funkcemi. Hp takového modelu je kombinovaná, v některých proměnných diskrétní a v jiných spojitá. Představuje soubor spojitých funkcí, pro každou hodnotu diskrétní proměnné jednu funkci.

Model

Rovnice modelu je (1.24)

$$y = \theta + e,$$

kde diskrétní parametr θ má rozdělení

θ	-2	0	2
$f(\theta)$	0.2	0.5	0.3

Model má podmíněnou kombinovanou hp (1.25)

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

Sdružená hp

Sdružené pravděpodobnosti pro jednotlivé hodnoty θ určíme podle vzorce

$$f(y,\theta) = f(y|\theta)f(\theta)$$

tak, že každou spojitou hp modelu násobíme příslušným prvkem vektoru $f(\theta)$

$$f(y,\theta) = \begin{cases} 0.2f(y|\theta=-2) = \frac{0.2}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y+2)^2\right\} & \text{pro } \theta = -2 \\ 0.5f(y|\theta=0) = \frac{0.5}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y^2\right\} & \text{pro } \theta = 0 \\ 0.3f(y|\theta=2) = \frac{0.3}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y-2)^2\right\} & \text{pro } \theta = 2 \end{cases} \tag{2.64}$$

Prediktivní hp

Tato hp je marginální hp ze sdružené a můžeme ji určit podle vzorce pro úplnou pravděpodobnost

$$f(y) = \sum_j f(y|\theta_j)f(\theta_j).$$

Dostaneme ji jako součet prvků z předchozího vyjádření sdružené hp

$$f(y) = \frac{1}{\sqrt{2\pi}}\left[0.2\exp\left\{-\frac{1}{2}(y+2)^2\right\} + 0.5\exp\left\{-\frac{1}{2}y^2\right\} + 0.3\exp\left\{-\frac{1}{2}(y-2)^2\right\}\right] \tag{2.65}$$

Bodová predikce výstupu

Je to opět střední hodnota $\hat{y} = E[y] = \int_{-\infty}^{\infty} y\, f(y)\mathrm{d}y$

$$\hat{y} = 0.2E[y|\theta=-2] + 0.5E[y|\theta=0] + 0.3E[y|\theta=2] =$$

$$= 0.2 \cdot (-2) + 0.5 \cdot 0 + 0.3 \cdot 2 = 0.2\ . \tag{2.66}$$

Aposteriorní hp

Dostaneme ji podle vzorce

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

tak, že každý prvek sdružené hp dělíme hodnotou $f(y)$ v příslušném bodě y z podmínky. Formálně ji můžeme zapsat takto (y_0 je pevné y v podmínce)

$$f(\theta|y = y_0) \quad \begin{cases} \frac{1}{f(y_0)} \frac{0.2}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y_0 + 2)^2\right\} & \text{pro } \theta = -2 \\ \\ \frac{1}{f(y_0)} \frac{0.5}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}y_0^2\right\} & \text{pro } \theta = 0 \\ \\ \frac{1}{f(y_0)} \frac{0.3}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(y_0 - 2)^2\right\} & \text{pro } \theta = 2 \end{cases} \tag{2.67}$$

Kapitola 3

Řízení při známých parametrech modelu

V této kapitole navážeme na Kapitolu ??, kde jsme se zabývali základními výpočetními principy optimálního řízení založeného na kvadratickém kriteriu optimality. Z důvodu názornosti jsme uvažovali deterministický model. Zde naznačíme odvození syntézy řízení pro stochastický model obecného řádu se známými parametry. Řízením s modelem, jehož parametry neznáme, se budeme zabývat v následující kapitole.

Při syntéze řízení, podobně jako tomu bylo i v identifikaci, je možno sledovat dvě úrovně. Na jedné úrovni obdržíme rekurzivní algoritmus pro hustoty pravděpodobností, aniž bychom specifikovali 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 zadat konkrétní tvar rozdělení jednotlivých hustot pravděpodobností – zde budeme předpokládat normální rozdělení, a 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 v Kapitole ??. Rovněž výsledky odvození jsou pro odpovídající si modely a kriteria prakticky shodné. Jediným rozdílem je, že zbytek kriteria po minimalizaci je pro stochastický model větší, neboť je do něho kromě chyb řízení zahrnut ještě rozptyl šumu, který nelze řídit.

Vlastní algoritmus se sestává ze dvojice kroků, které se vždy v každém časovém okamžiku opakují. Jsou to kroky:

Středování, při kterém se s pomocí modelu řízené soustavy počítá střední hodnota příslušné části kriteria (té části, která obsahuje nejnovější data a zbytku kriteria z předchozí minimalizace). Tento krok převádí statistickou úlohu na deterministickou.

Minimalizace, která hledá optimální hodnotu nejnovějšího členu akční veličiny, pomocí které soustavu řídíme. Tento krok koresponduje s minimalizací kriteria pro deterministický model.

3.1 Odvození v integrální formě

Model

Syntézu řízení odvodíme pro stejný tvar modelu, který jsme uvažovali v druhé kapitole předešlého dílu, tj pro model (2.1)

$$y_t = \theta^T \varphi_t + e_t,$$

kde

y_t je výstup modelu,

θ jsou parametry modelu,

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

u_t je řídící veličina,

e_t je porucha modelu (šum).

Dále budeme používat následující symboliku:

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

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

$D(t)$ značí množinu všech datových dvojic, až do okamžiku s časovým indexem t.

Symbol $D(0)$ tedy označuje apriorní data (všechna data, naměřená před začátkem řízení). Data $D(t)$ získaná až do okamžiku t lze rozdělit takto: $D(t) = \{D(1..\, t), D(0)\}$, tj na data apriorní a data získaná v průběhu řízení až do času t.

Poznámka: *Připomeňme ještě, že všechny zapsané integrály jsou určité. Nejsou-li explicitně zapsány meze integrace, máme na mysli integrál přes celý obor hodnot proměnné přes kterou integrujeme. To je většinou reálná osa nebo příslušně dimenzionální Euclidův prostor.*

Kriterium

Kriterium, které nyní zahrnuje signály ve formě posloupností náhodných veličin, má tvar

$$J = E\left[\sum_{\tau=1}^{N} y_\tau^2 + \omega u_\tau^2 | D(0)\right]$$

kde $E[.|D(0)]$ značí operátor podmíněné střední hodnoty definovaný pomocí hp, podmíněné apriorními daty $f(.|D(0))$ – tedy např. pro středovanou funkci $h(x)$ je

$$E[h(x)|D(0)] = \int h(x) f(x|D(0)) \mathrm{d}x.$$

Syntéza řízení pro hp

Naším úkolem je nalézt v obecném případě posloupnost hp $f(u_t|D(t-1))$ pro $t = 1, .., N$, které minimalizují zvolené kriterium. Postup minimalizace naznačíme, podobně jako jsme to udělali v kapitole 2.2 předchozího dílu, na horizontu řízení $N = 3$.

Tedy, hledáme posloupnost hp

$$f_1, f_2, .., f_N, \qquad f_\tau = f(u_\tau|D(\tau))$$

tak, aby bylo dosaženo minima

$$J = \min_{f_1,f_2,f_3} E\left[\sum_{\tau=1}^{3}\left(y_\tau^2 + \omega u_\tau^2\right)|D(0)\right] =$$

$$= \min_{f_1,f_2,f_3} \int \sum_{\tau=1}^{3}\left(y_\tau^2 + \omega u_\tau^2\right) f(y_3, u_3, y_2, u_2, y_1, u_1|D(0))\mathrm{d}(y_3, u_3, y_2, u_2, y_1, u_1),$$

kde symbol $\int .\ \mathrm{d}(y_3, u_3, ..)$ představuje integraci přes celé obory hodnot veličin $y_3, u_3, ..$

Minimalizaci, stejně jako tomu bylo dříve, začneme od konce, tedy pro časový okamžik $t = N = 3$. V kriteriu oddělíme veličiny s indexem 3 a sdruženou hp rozdělíme podle řetězového pravidla.

$$J = \min_{f_1,f_2,f_3} \int\int \left[\sum_{\tau=1}^{2}\left(y_\tau^2 + \omega u_\tau^2\right) + \left(y_3^2 + \omega u_3^2\right)\right]\cdot$$

$$\cdot f(y_3, u_3|D(2))\ f(y_2, u_2, y_1, u_1|D(0))\ \mathrm{d}(y_3, u_3)\mathrm{d}(y_2, u_2, y_1, u_1)$$

První část integrací 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

$$J = \min_{f_1,f_2} \int \left\{\sum_{\tau=1}^{2}\left(y_\tau^2 + \omega u_\tau^2\right) + \min_{f_3} \int \left(y_3^2 + \omega u_3^2\right) f(y_3, u_3|D(2))\ \mathrm{d}(y_3, u_3)\right\}\cdot$$

$$\cdot f(y_2, u_2, y_1, u_1|D(0))\ \mathrm{d}(y_2, u_2, y_1, u_1)$$

Dále upravíme argument minima druhého členu ve složené závorce. Označíme jej J_3.

$$J_3 = \int \left(y_3^2 + \omega u_3^2\right) f(y_3, u_3|D(2))\ \mathrm{d}(y_3, u_3) =$$

$$= \int\int \left(y_3^2 + \omega u_3^2\right) f(y_3|u_3, D(2)) f(u_3|D(2))\mathrm{d}y_3\mathrm{d}u_3 =$$

$$= \int E\left[y_3^2 + \omega u_3^2|u_3, D(2)\right] f(u_3|D(2))\mathrm{d}u_3$$

kde $E[.|u_3, D(2)]$ je podmíněná střední hodnota, která je funkcí dat z podmínky, tedy u_3 a $D(2)$ tj. veličin $u_3, y_2, u_2, y_1, u_1, y_0, u_0, ..$

Poznámka: *Zmíněný integrál představuje střední hodnotu výrazu $y_3^2 + \omega u_3^2$ podle náhodné veličiny y_3. Pro hp $f(y_3|u_3, d(2))$ definovanou regresním modelem $y_3 = \theta^T \varphi_3 + e_3$ nebo jeho speciálním tvarem prvního řádu, který jsme uvažovali v Kapitole* **??**, *bude výsledek středování* *shodný (až na rozptyl šumu) s dosazením rovnice deterministického modelu za náhodnou veličinu y_3. Tento obrat jsme používali v Kapitole* **??** *k tomu, abychom veličinu y_3 predikovali (zbavili se jí), protože byla v daném okamžiku neznámá. V tomto bodě se tedy schází deterministická a stochastická optimalizace a pojmy v nich užívané: "dosazení modelu" a "středování". Oba dva pojmy představují predikci, na níž je řízení (podobně jako identifikace) 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 kriterium J. Protože z celého kriteria na $f(u_3|D(2))$ závisí pouze člen J_3, hledáme minimum tohoto výrazu. Toho dosáhneme tehdy, když nalezneme minimum střední hodnoty v integrandu a do něho soustředíme celou hp $f(u_3|D(2))$ ve formě Diracova impulzu (definovaného vztahem $\delta(x) = 0$ pro $x \neq 0$ a $\int \delta(x) \mathrm{d}x = 1$), tedy

$$f(u_3|D(2)) = \delta(u_3 - u_3^*),$$

kde

$$u_3^* = \arg\min_{u_3} E\left[y_3^2 + \omega u_3^2 | u_3, D(2)\right].$$

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 v integrandu střední hodnoty. Výsledek minimalizace je tedy

$$\min_{u_3} J_3 = E\left[y_3^2 + \omega (u_3^*)^2 | u_3^*, D(2)\right] = J_3^*(D(2)).$$

Tímto je ukončena minimalizace v čase $t = 3$. Výraz J_3^* je zbytek po této minimalizaci. Je funkcí dat do času 2. Dosadíme jej zpět do kriteria J a budeme pokračovat v jeho minimalizaci v čase $t = 2$

$$J = \min_{f_1, f_2} \int \left(\sum_{\tau=1}^{2} (y_\tau^2 + \omega u_\tau^2) + J_3^* \right) f(y_2, u_2, y_1, u_1 | D(0)) \mathrm{d}(y_2, u_2, y_1, u_1).$$

Postup minimalizace kriteria J pro čas 2 bude obdobný, jako pro čas 3

$$J = \min_{f_1, f_2} \int \int \left[\sum_{\tau=1}^{1} \left(y_\tau^2 + \omega u_\tau^2 \right) + \left(y_2^2 + \omega u_2^2 \right) + J_3^* \right] \cdot$$

$$\cdot f(y_2, u_2 | D(1))\ f(y_1, u_1 | D(0))\ \mathrm{d}(y_2, u_2) \mathrm{d}(y_1, u_1) =$$

$$= \min_{f_1} \int \left\{ y_1^2 + \omega u_1^2 + \min_{f_2} \int \left(y_2^2 + \omega u_2^2 + J_3^* \right) f(y_2, u_2 | D(1)) \mathrm{d}(y_2, u_2) \right\} \cdot$$

$$\cdot f(y_1, u_1 | D(0))\ \mathrm{d}(y_1, u_1).$$

Druhý sčítanec v integrandu označíme analogicky k předchozímu kroku a obdobně upravíme

$$J_2 = \int \int \left(y_2^2 + \omega u_2^2 + J_3^* \right) f(y_2 | u_2, D(1)) f(u_2 | D(1))\ \mathrm{d}y_2 \mathrm{d}u_2 =$$

$$= \int E\left[y_2^2 + \omega u_2^2 + J_3^* | u_2, D(1) \right] f(u_2 | D(1)) \mathrm{d}u_2.$$

Minima dosáhneme analogicky předchozímu kroku volbou $f(u_2|D(1)) = \delta(u_2 - u_2^*)$ a minimalizací střední hodnoty. Opět platí

$$u_2^* = \arg\min_{u_2} E\left[y_2^2 + \omega u_2^2 + J_3^*|u_2, D(1)\right]$$

$$\min_{u_2} J_2 = E\left[y_2^2 + \omega(u_2^*)^2 + J_3^*(D^*(2))|u_2^*, D(1)\right] = J_2^*(D(1)),$$

kde $D^*(2)$ označuje data do času 2, kde za akční veličinu u_2 je dosazena její optimální hodnota u_2^*.
Na rozdíl od předchozího kroku přibyl v argumentu střední hodnoty další člen. Je jím zbytek po předchozí minimalizaci J_3^*.

Po zpětném dosazení do kriteria J obdržíme

$$J = \min_{f_1} \int \left(\sum_{\tau=1}^{1}(y_\tau^2 + \omega u_\tau^2) + J_2^*\right) f(y_1, u_1|D0)\mathrm{d}(y_1, u_1),$$

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

$$J_1 = \int E\left[y_1^2 + \omega u_1^2 + J_2^*|u_1, D(0)\right] f(u_1|D(0))\mathrm{d}u_1,$$

$$f(u_1|D(0)) = \delta(u_1 - u_1^*), \qquad \text{kde} \quad u_1^* = \arg\min_{u_1} E\left[y_1^2 + \omega u_1^2 + J_2^*|u_1, D(0)\right]$$

$$\min_{u_1} J_1 = E\left[y_1^2 + \omega(u_1^*)^2 + J_2^*(D^*(1))|u_1, D(0)\right] = J_1^*(D(0)).$$

Výsledná minimální hodnota kriteria J je

$$J = J_1^*(D(0))$$

Obecný algoritmus minimalizace kriteria pro optimální řízení na horizontě délky N je

Algoritmus

Syntéza optimálního řízení

Start *počáteční hodnota kriteria*

$$J^*_{N+1} = 0$$

For $\tau = N$ **downto** 1 **do**

středování

$$E_\tau = E\left[y_\tau^2 + \omega u_\tau^2 + J^*_{\tau+1} | u_\tau, D(\tau-1)\right] =$$

$$= \int \left(y_\tau^2 + \omega u_\tau^2 + J^*_{\tau+1}\right) f(y_\tau | u_\tau, D(\tau-1)) \mathrm{d}y_\tau$$

minimalizace

$$J^*_\tau = \min_{u_\tau} E_\tau, \qquad u^*_\tau = \arg\min_{u_\tau} E_\tau$$

End for

Optimální strategie řízení $u_1^*, u_2^*, .., u_N^*$
Minimální hodnota kriteria $J = J_1^*$

Syntéza řízení pro pro statistiky

Předchozí algoritmus ukazuje rekurzivní postup minimalizace kriteria pro hp. Konkrétní tvar těchto hustot zatím nebyl zadán ani vyžadován. V každém kroku se přitom počítá střední hodnota

$$E\left[y_\tau^2 + \omega u_\tau^2 | u_\tau, D(\tau-1)\right]$$

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, přes které se středuje). Teprve když zmíněné hp parametrizujeme, tj. zavedeme parametrizovaný model, který tyto hustoty popisuje, můžeme převést tento algoritmus 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 $y_t = \theta^T\varphi_t + e_t$, kde $e_t \sim N(0, \sigma^2)$, můžeme spočítat naznačené středování kvadrátu výstupu y_t

$$E\left[y_t^2|u_t, D(t-1)\right] = E\left[(\theta^T\varphi_t + e_t)^2|u_t, D(t-1)\right] =$$

$$= E\left[(\theta^T\varphi_t)^2 + 2\theta^T\varphi_t e_t + e_t^2|u_t, D(t-1)\right] = \left(\theta^T\varphi_t\right)^2 + \sigma^2$$

neboť $E[e_t|u_t, D(t-1)] = 0$ a $E[e_t^2|u_t, D(t-1)] = \sigma^2$.

Z předchozího výsledku je vidět, že efekt středování je stejný, jako při dosazení 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 kriteria na horizontě délky N o $N\sigma^2$.

Algoritmus minimalizace kriteria pro statistiky hp je tedy formálně shodný s tím, který jsme odvodili v minulém díle pro deterministický model.

3.2 Odvození v operátorovém vyjádření

Algoritmus pro hp, který jsme v této kapitole odvodili, 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 podmínění střední hodnoty při rozvedení sdružené hp podle řetězového pravidla. Uvažujme střední hodnotu funkce $h(u, v)$ dvou náhodných veličin u a v se sdruženou hp $f(u, v)$

$$E[h(u,v)] = \int\int h(u,v)\cdot f(u,v)\mathrm{d}u\,\mathrm{d}v =$$

$$= \int\left(\int h(u,v)f(u|v)\mathrm{d}u\right)f(v)\mathrm{d}v = E\left[\,E[h(u,v)|v]\,\right].$$

S použitím tohoto pravidla vyjádříme jeden krok minimalizace kriteria. Budeme sledovat minimalizaci pro čas 2 z předchozího odvození, 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.

$$\begin{aligned}
J &= \min_{u_1,u_2,..,u_t} E\left[\sum_{\tau=1}^{t}(y_\tau^2 + \omega u_\tau^2) + J_{t+1}^*|D(0)\right] = \\
&= \min_{u_1,u_2,..,u_t} E\left[E\left[\sum_{\tau=1}^{t-1}(y_\tau^2 + \omega u_\tau^2) + y_t^2 + \omega u_t^2 + J_{t+1}^*|D(1..t-1)\right]|D(0)\right] = \\
&= \min_{u_1,u_2,..,u_{t-1}} E\left[\sum_{\tau=1}^{t-1}(y_\tau^2 + \omega u_\tau^2) + \min_{u_t} E\left[y_t^2 + \omega u_t^2 + J_{t+1}^*|u_t, D(1..t-1)\right]|D(0)\right]
\end{aligned}$$

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

$$J_t^* = \min_{u_t} J_t$$

$$u_t^* = \arg\min_{u_t} J_t.$$

Tvar kriteria se po dosazení minima J_t reprodukuje

$$J = \min_{u_1, u_2, .., u_{t-1}} E\left[\sum_{\tau=1}^{t-1}(y_\tau^2 + \omega u_\tau^2) + J_t^* | D(0)\right]$$

Výsledek, který jsme dostali je shodný s tím, který jsme odvodili pomocí integrálů a hp.

Kapitola 4

Řízení při neznámých parametrech modelu

Dosud jsme se zabývali úlohami identifikace a řízení více méně odděleně. Odvodili jsme algoritmy pro identifikaci parametrů modelu řízené soustavy. Odhady parametrů jsme vyjádřili pomocí podmíněné hustoty parametrů (aposteriorní hp). Dále můžeme s pomocí aposteriorní hp vypočítat bodové odhady parametrů. Jejich tvar závisí na kriteriu, podle kterého je určujeme. Pro nejčastěji uvažované kvadratické kriterium je bodovým odhadem parametrů jejich podmíněná střední hodnota.

♣ Odhady parametrů můžeme vyjádřit buď pomocí hp parametrů nebo jejich číselnou hodnotou jako bodové odhady.

Pro syntézu řízení jsme uvažovali parametry známé. Optimální řízení, minimalizující kvadratické kriterium jsme v každém kroku řízení obdrželi minimalizací podmíněné střední hodnoty součtu příslušné části kriteria a zbytku z minimalizace v předchozím kroku.

♣ Syntéza řízení je odvozena pro známé parametry.

Jsou-li parametry soustavy neznámé, nelze přesně prakticky použitelné adaptivní řízení rozumně odvodit. Takové řízení se nazývá duální a výpočty, které pro určení akční veličiny vyžaduje, nelze na současných počítačích prakticky realizovat.

♣ V případě optimálního řízení s modelem soustavy, jehož parametry nejsou známy, a mají být průběžně odhadovány, je třeba provést aproximaci, která identifikaci a řízení oddělí.

Nyní ukážeme dva nejčastěji používané postupy, které umožní provést prakticky realizovatelnou syntézu řízení. V obou případech budeme při získávání odhadu parametrů pro řízení a vlastní syntéze řízení postupovat takto (*metoda klouzavého horizontu*): Stanovíme určitý časový horizont, tj. počet kroků identifikace a řízení, které směřují od současného okamžiku dopředu. Pro jednotlivé okamžiky tohoto intervalu počítáme od konce intervalu řídící zákon. Ten se postupně ustaluje (jeho koeficienty se přestávají měnit). Vezmeme poslední řídící zákon a použijeme jej pro jeden krok řízení. Parametry modelu soustavy se odhadují průběžně, tj. s použitím všech dat, která jsme dosud naměřili. Pro výpočet řídícího 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 kriteria zbytky, které by přecházely do dalšího kroku minimalizace. V důsledku toho je řídící zákon ustálen již při prvním kroku a postačí zvolit horizont délky jedna. Tím se vyhneme nepříjemným iteracím na horizontě.*

4.1 Důvěřivá strategie řízení

Nejjednodušším přístupem 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 řídícího zákona se parametry nahradí jejich aktuálními bodovými odhady.

♣ Důvěřivá strategie znamená bezprostřední kombinaci identifikace a řízení se známými parametry, kdy za parametry modelu pro řízení použijeme bodové odhady, získané identifikací.

Tuto metodu stručně ilustrujeme na příkladě.

PŘÍKLAD 4.1 Řízení s důvěřivou strategií

Zadání příkladu

Pro ilustraci zvolíme nejjednodušší možný – statický model. Stejný model použijeme i pro další, tzv. opatrnou strategii, která je komplikovanější a příklad by se stal pro složitější model nepřehledný. Model soustavy bude

$$y_t = \theta u_t + e_t,$$

kde $e_t \sim N(0, \sigma^2)$, $\sigma^2 = \text{konst.}$ $\forall t$ a kriterium řízení

$$J = \sum_{t=1}^{N} (y_t - w_t)^2 + \omega u_t^2,$$

kde ω je penalizace akční veličiny a w_t je žádaná hodnota výstupu, to je hodnota, kterou má nabýt řízený výstup soustavy. (Řízení je nejlepší, je-li argument kvadrátu nula – potom je $y_t = w_t$.)

Bodové odhady parametru

Obecný vzorec pro aktualizaci informační matice nebo jejích submatic je dán v Kapitole 2. Jemu také odpovídá vyjádření věrohodnostní funkce (4.6) a (4.7), která v našem případě odpovídá aposteriorní hp. Jejím rekurentním výpočtem a maximalizací jsou dány bodové odhady parametru. Algoritmus výpočtu bodových odhadů lze vyjádřit buď podle (4.6) nebo podle (4.7). První algoritmus je jednodušší a průhlednější, druhý má lepší výpočetní vlastnosti.

▷ `Algoritmus I`
Informační matice se přepočítává rekurzivně a z ní se v každém kroku počítají odhady parametrů, případně další charakteristiky.

`Přepočet informační matice:`

$$\begin{aligned} S_y^2(t) &= S_y^2(t-1) + y_t^2 \\ S_{yu}^2(t) &= S_{yu}^2(t-1) + y_t u_t \\ S_u^2(t) &= S_u^2(t-1) + u_t^2 \end{aligned}$$

s počáteční podmínkou $S_y^2(0), S_{yu}(0), S_u^2(0)$ danými volbou apriorní hp.

`Výpočet odhadu parametrů:`

$$\hat{\theta}_t = \frac{S_{yu}}{S_u^2}$$

`Kovariance odhadu parametrů:`

$$C_t = \frac{1}{S_u^2}$$

`Zbytek kvadratické formy po minimalizaci:`

$$\Lambda_t = S_y^2 - \frac{(S_{yu})^2}{S_u^2}$$

▷ `Algoritmus II`
Rekurzivně se 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 pro dynamické modely, kdy S_u^2 odpovídá matice. Pro následující vzorce označíme

$$C_{t-1} u_t^2 = \zeta_t$$

Potom rekurze má tvar

$$\begin{aligned} \hat{\theta}_t &= \hat{\theta}_{t-1} + \frac{1}{1+\zeta_t} C_{t-1} u_t (y_t - \hat{\theta}_{t-1} u_t) \\ C_t &= C_{t-1} - \frac{1}{1+\zeta_t} (C_{t-1} u_t)^2 \\ \Lambda_t &= \Lambda_{t-1} + \frac{1}{1+\zeta_t} (y_t - \hat{\theta}_{t-1} u_t)^2 \end{aligned} \tag{4.1}$$

`Poznámka:` *Odvození vzorců Algoritmu II z Algoritmu I je uvedeno v Doplňku 5.3. V obecném případě vyžaduje odvození těchto vzorců speciality maticového počtu, ale postup je obdobný.*

Syntéza řízení

Pro určení akčních veličin máme minimalizovat podmíněnou střední hodnotu

$$\min_{u_t} E[(y_t - w_t)^2 + \omega u_t^2 + J^*_{t+1}|u_t, D(t-1)]$$

s koncovou podmínkou $J^*_{N+1} = 0$. Podmíněná hp, která je pro výpočet střední hodnoty zapotřebí, je dána modelem (předpokládáme známé parametry a dosadíme za ně bodové odhady)

$$f(y_t|u_t, D_t) \sim f(y_t|u_t, D_t, \Theta = \hat{\Theta}_{t-1}) = N(\hat{\theta}_{t-1}u_t; \sigma^2)$$

Minimalizované kriterium umocníme provedeme středování jednotlivých členů. Pro výpočet střední hodnoty kvadrátu použijeme vzorec $E[y^2] = D[y] + (E[y]^2)$. Dostaneme

$$\min_{u_t}\{\hat{\theta}^2_{t-1}u_t^2 + \sigma^2 - 2\hat{\theta}_{t-1}u_t w_t + w_t^2 + \omega u_t^2 + E[J^*_{t+1}|u_t, D(t-1)]\}$$

Hledáme minimum, např. pomocí derivace. Protože lze obecně ukázat, že pro statický model je zbytek kriteria 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

$$\hat{\theta}^2_{t-1}u_t^* - \hat{\theta}_{t-1}w_t + \omega u_t^* = 0$$

odkud plyne

$$u_t^* = \frac{\hat{\theta}_{t-1}}{\omega + \hat{\theta}^2_{t-1}} w_t$$

pro $t = 1, 2, .., N$.

Poznámka: *Výsledek, který jsme obdrželi lze srovnat s výsledky z Kapitoly* **??**. *Tam jsme uvažovali vícekrokové řízení na určitém horizontu, tady je horizont roven jedné. Žádaná hodnota, zde uvažovaná je rovněž nová.*

4.2 Opatrná strategie řízení

Je zřejmé, že jednoduchost předešlé strategie bude na úkor její obecnosti. Skutečně. Tato strategie bere 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 přesné. Proto je lépe reprezentovat odhadnuté parametry jejich aposteriorními hp. V nich je zahrnut nejen bodový odhad parametru, ale také jeho rozptyl. Syntéza potom počítá s tím, že odhady parametrů jsou sice takové, ale jsou to jen 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 identifikace a a algoritmu řízení se známými parametry. Naznačíme její odvození.

Základním krokem při postupné minimalizaci kriteria je výpočet střední hodnoty

$$\begin{aligned} J_t^* &= \min_{u_t} E[(y_t^2 + \omega u_t^2 + J^*_{t+1})|u_t, D(t-1)] \\ &= \min_{u_t} \int (y_t^2 + \omega u_t^2 + J^*_{t+1}) f(y_t|u_t, D(t-1)) \mathrm{d}y_t. \end{aligned} \tag{4.2}$$

Prediktivní hp

Pro výpočet střední hodnoty potřebujeme prediktivní hp $f(y_t|u_t, D(t-1))$. Tu však přímo známe jen pro model se známými parametry – potom prediktivní hp se přímo rovná hp definované modelem, neboť v ní parametry jako proměnné nevystupují. Pro model s neznámými parametry musíme prediktivní hp dopočítat (viz Kapitola 2)

$$\begin{aligned} f(y_t|u_t, D(t-1)) &= \int f(y_t|u_t, D(t-1), \Theta)\, f(\Theta|u_t, D(t-1))\mathrm{d}\Theta = \\ &= \int f(y_t|\varphi_t, \Theta)\, f(\Theta|D(t-1))\mathrm{d}\Theta, \end{aligned}$$

kde první hp na pravé straně je dána modelem ($y_t = \theta^T\varphi_t + e_t$) a druhá je hp parametrů, která je vyvíjena v identifikaci.

Ukážeme odvození vzorce pro off-line (jednorázový) výpočet prediktivní hp. Vezmeme-li v úvahu tvar hp parametrů tak, jak jej dostáváme z identifikace, jako normovaný součin hp modelů v naměřených datech a apriorní hp

$$f(\Theta|D(t-1)) = \frac{\prod_{\tau=1}^{t-1} f(y_\tau|\varphi_\tau, \Theta)\, f(\Theta|D(0))}{\int \prod_{\tau=1}^{t-1} f(y_\tau|\varphi_\tau, \Theta)\, f(\Theta|D(0))\mathrm{d}\Theta}, \tag{4.3}$$

můžeme vzorec pro prediktivní hp ještě upravit. Dosadíme předchozí vyjádření hp parametrů a dostaneme

$$\begin{aligned} f(y_t|u_t, D(t-1)) &= \int f(y_t|\varphi_t, \Theta) \frac{\prod_{\tau=1}^{t-1} f(y_\tau|\varphi_\tau, \Theta)\, f(\Theta|D(0))}{\int \prod_{\tau=1}^{t-1} f(y_\tau|\varphi_\tau, \vartheta)\, f(\vartheta|D(0))\mathrm{d}\vartheta}\mathrm{d}\Theta = \\ &= \frac{\int \prod_{\tau=1}^{t} f(y_\tau|\varphi_\tau, \Theta)\, f(\Theta|D(0))\mathrm{d}\Theta}{\int \prod_{\tau=1}^{t-1} f(y_\tau|\varphi_\tau, \vartheta)\, f(\vartheta|D(0))\mathrm{d}\vartheta} \end{aligned}$$

Integrál z věrohodnostní funkce násobené apriorní hp nazveme Bayesovým integrálem a označíme jej I_t

$$I_t(\Theta) = \int \prod_{\tau=1}^{t} f(y_\tau|\varphi_\tau, \Theta)\, f(\Theta|D(0))\mathrm{d}\Theta \tag{4.4}$$

prediktivní hp lze potom psát jako podíl Bayesových integrálů v časech t a $t-1$

$$f(y_t|u_t, D(t-1)) = \frac{I_t}{I_{t-1}}. \tag{4.5}$$

Poznámka: *Vztah pro prediktivní hp je přesně tím místem, kde se schází identifikace a řízení a vytváří základy adaptivního řízení. První hp představuje model se známými parametry, se kterým bychom mohli řídit, pokud bychom parametry znali. Druhá hp přináší informace o neznámých parametrech.*

Protože obecné odvození opatrné řídící 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 přes velikou jednoduchost použitého modelu, struktura odvození i jednotlivé vzorce budou dobře ukazovat na obecný případ.

PŘÍKLAD 4.2 Řízení s opatrnou strategií

Zadání příkladu

Pro ilustraci odvození opatrné strategie řízení zvolíme stejný model, jako jsme uvažovali pro důvěřivou strategii

$$y_t = \theta u_t + e_t,$$

kde opět $e_t \sim N(0, \sigma^2)$, $\sigma^2 =$ konst. $\forall t$ a stejné kriterium řízení

$$J = \sum_{t=1}^{N} (y_t - w_t)^2 + \omega u_t^2,$$

kde opět w_t je žádaná hodnota výstupu.

Přípravné výpočty

Společným prvkem jak identifikace tak i řízení je predikce výstupu. Podíváme se tedy nejdříve, co potřebujeme znát pro určení prediktivní hp $f(y_t|u_t, D(t-1))$. Dostaneme ji podle výše uvedeného vzorce, dosadíme-li hp modelu

$$\begin{aligned} f(y_t|u_t,\Theta) &= \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(y_t - \theta u_t)^2\right\} = \\ &= \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left\{\frac{1}{2\sigma^2} [-1, \quad \theta] \begin{bmatrix} y_t^2 & y_t u_t \\ y_t u_t & u_t^2 \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix}\right\} \end{aligned}$$

a hp parametrů $f(\Theta|D(t-1))$. Hp parametrů můžeme chápat jako aposteriorní z minulého kroku (tedy pro čas $t-1$). Tuto hp lze formálně zapsat jako součin hp modelu pro časy $t = 1, 2, .., N-1$ násobený apriorní hp $f(\Theta|D(0))$. Připravíme si tedy potřebné nástroje.

▷ `Věrohodnostní funkce`
Součin hp modelu pro naměřená data do času t se nazývá věrohodnostní funkce $L_t(\Theta)$. V našem případě bude

$$\begin{aligned} L_t(\Theta) &= \prod_{\tau=1}^{t} f(y_\tau|u_\tau,\Theta) = \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma^2}\right)^t \exp\left\{-\frac{1}{2\sigma^2} [-1, \quad \theta] \begin{bmatrix} \sum_{\tau=1}^{\tau} y_\tau^2 & \sum_{\tau=1}^{t} y_\tau u_\tau \\ \sum_{\tau=1}^{\tau} y_\tau u_\tau & \sum_{\tau=1}^{\tau} u_\tau^2 \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix}\right\} \end{aligned} \tag{4.6}$$

Označíme-li

$$S_y^2(t) = \sum_{\tau=1}^{\tau} y_\tau^2, \quad S_{yu}(t) = \sum_{\tau=1}^{\tau} y_\tau u_\tau, \quad S_u^2(t) = \sum_{\tau=1}^{\tau} u_\tau^2$$

bude věrohodnostní funkce

$$\begin{aligned} L_t(\Theta) &= \left(\frac{1}{\sqrt{2\pi}\sigma^2}\right)^t \exp\left\{-\frac{1}{2\sigma^2} [-1, \quad \theta] \begin{bmatrix} S_y^2(t) & S_{yu}(t) \\ S_{yu}(t) & S_u^2(t) \end{bmatrix} \begin{bmatrix} -1 \\ \theta \end{bmatrix}\right\} = \\ &= \left(\frac{1}{\sqrt{2\pi}\sigma^2}\right)^t \exp\left\{-\frac{1}{2\sigma^2}[S_u^2(t)(\theta - \hat{\theta}_t)^2 + \Lambda_t]\right\} \end{aligned} \tag{4.7}$$

kde

$$\hat{\theta}_t = \frac{S_{yu}(t)}{S_u^2(t)} \qquad \text{a} \qquad \Lambda_t = S_y^2(t) - \frac{(S_{yu}(t))^2}{S_u^2(t)}$$

Poslední úpravu jsme provedli roznásobením kvadratické formy v exponentu a doplněním na čtverec v proměnné θ.

▷ `Apriorní hp`
Pro jednoduchost budeme uvažovat rovnoměrné rozdělení apriorní hp. Potom aposteriorní hp, je úměrná přímo věrohodnostní funkci. Stačí ji jen normalizovat. To provedeme snadno tak, že vynecháme multiplikativní konstantu věrohodnostní funkce (vzhledem k proměnné θ) a místo ní napíšeme multiplikativní konstantu normálního rozdělení, které tato hp má.

▷ `Aposteriorní hp`
Normováním věrohodnostní funkce dostaneme aposteriorní hp

$$f(\Theta|D(t-1)) = \frac{1}{\sqrt{2\pi\sigma^2/S_u^2(t)}} \exp\left\{-\frac{1}{2\sigma^2/S_u^2(t)}(\theta - \hat{\theta}_t)^2\right\} \tag{4.8}$$

což je hp normálního rozdělení se střední hodnotou $\hat{\theta}_t$ a rozptylem $\sigma^2/S_u^2(t)$.

▷ `Bayesův integrál`
Bayesův integrál je integrál ze součinu věrohodnostní funkce a apriorní hp, což je vlastně nenormovaná aposteriorní hp, podle parametru Θ. Vzhledem k rovnoměrnosti apriorní hp bude integrand úměrný (s konstantou úměrnosti k) věrohodnostní funkci.

$$\begin{aligned}
I_t &= \int k \frac{1}{(2\pi\sigma^2)^{\frac{t}{2}}} \exp\left\{-\frac{1}{2\sigma^2}[S_u^2(t)(\theta-\hat{\theta}_t)^2 + \Lambda_t]\right\} \mathrm{d}\theta \\
&= k(2\pi\sigma^2)^{-\frac{t}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_t\right\} \int \exp\left\{-\frac{1}{2\sigma^2/S_u^2(t)}(\theta-\hat{\theta}_t)^2\right\} \mathrm{d}\theta \\
&= k(2\pi\sigma^2)^{-\frac{t}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_t\right\} \sqrt{2\pi\sigma^2/S_u^2(t)} \\
&= k(2\pi\sigma^2)^{-\frac{t-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_{t-1}\right\} (S_u(t))^{-1} \exp\left\{-\frac{1}{\sigma^2}\frac{S_u^2(t-1)}{S_u^2(t)}(y_t - \hat{\theta}_{t-1}u_t)^2\right\}
\end{aligned} \tag{4.9}$$

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 Doplňku na konci oddílu 5.3.

Tímto jsme připravili vše potřebné pro konstrukci základních vzorců adaptivního řízení. Zopakujeme explicitní vzorce pro odhad parametrů (i když pro vlastní řízení již nejsou potřeba – jsou zahrnuty v Bayesově integrálu ve formě aposteriorní hp), zkonstruujeme prediktivní hp podle (4.5) a určíme řídící zákon minimalizací podle (4.2).

Průběžná identifikace

Aposteriorní hp, která je nejobecnějším popisem neznámých parametrů a kterou my zde pro odhadování skutečně využijeme, je dána Bayesovým vzorcem (4.3) a v našem příkladě má tvar (4.8). Bodové odhady parametrů se v tomto případě neuplatní a identifikace sehraje svou úlohu pouze při výpočtu prediktivní hp, jak bude patrno v následujícím odstavci.

Prediktivní hp

Prediktivní hp je třeba, pro výpočet střední hodnoty podle (4.2), jejíž minimalizací získáme akční veličinu. Podle (4.5) je prediktivní hp dána podílem Bayesových integrálů

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

$$= \frac{k(2\pi\sigma^2)^{-\frac{t-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_{t-1}\right\} (S_u(t))^{-1} \exp\left\{-\frac{1}{\sigma^2}\frac{S_u^2(t-1)}{S_u^2(t)}(y_t - \hat{\theta}_{t-1}u_t)^2\right\}}{k(2\pi\sigma^2)^{-\frac{t-1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\Lambda_{t-1}\right\} (2\pi\sigma^2)^{\frac{1}{2}}(S_u(t-1))^{-1}} =$$

$$= \frac{1}{\sqrt{2\pi\sigma^2 S_u^2(t)/S_u^2(t-1)}} \exp\left\{-\frac{1}{2\sigma^2 S_u^2(t)/S_u^2(t-1)}(y_t - \hat{\theta}_{t-1}u_t)^2\right\} \tag{4.10}$$

Dostali jsme hp s normálním rozdělením se střední hodnotou $\hat{y}_t = \hat{\theta}_{t-1}u_t$ – což je bodová predikce výstupu, a rozptylem $\sigma^2 S_u^2(t)/S_u^2(t-1)$. Bodová predikce výstupu $\hat{y}_t$ určená podle kvadratického kriteria se rovná podmíněné střední hodnotě a tedy bude

$$\hat{y}_t = \hat{\theta}_{t-1}u_t \tag{4.11}$$

Je to vlastně rovnice modelu, kde za šum je dosazena jeho střední hodnota – nula. Bodová predikce je tedy stejná, jako výpočet budoucí hodnoty výstupu z deterministického modelu.

Poznámka: *Bodovou predikci výstupu jsme opět uvedli pouze pro úplnost. K odvození řídícího zákona podle opatrné strategie potřebujeme jen prediktivní hp.*

Syntéza řízení

Pro určení akčních veličin použijeme stejné obecné kriterium jako pro důvěřivou strategii ve formě střední hodnoty

$$\min_{u_t} E[(y_t - w_t)^2 + \omega u_t^2 + J_{t+1}^*|u_t, D(t-1)]$$

s koncovou podmínkou $J_{N+1}^* = 0$. Podmíněná hp, která je pro výpočet střední hodnoty zapotřebí, je nyní prediktivní hp (4.10). Po úpravě $S_u^2(t)/S_u^2(t-1) = 1 + C_{t-1}u_t^2$ ji dostaneme ve tvaru

$$f(y_t|u_t, D_t) = N(\hat{\theta}_{t-1}u_t;\ \sigma^2(1 + C_{t-1}u_t^2))$$

Podobnými úpravami jako v předchozím příkladě dostaneme minimalizaci

$$\min_{u_t}\{\hat{\theta}_{t-1}^2 u_t^2 + \sigma^2 + \sigma^2 C_{t-1}u_t^2 - 2\hat{\theta}_{t-1}u_t w_t + w_t^2 + \omega u_t^2 + E[J_{t+1}^*|u_t, D(t-1)]\}$$

Optimální řízení u_t^* je nyní určeno rovnicí

$$\hat{\theta}_{t-1}^2 u_t^* + \sigma^2 C_{t-1} - \hat{\theta}_{t-1}w_t + \omega u_t^* = 0$$

odkud plyne

$$u_t^* = \frac{\hat{\theta}_{t-1}}{\omega + \hat{\theta}_{t-1}^2 + \sigma^2 C_{t-1}} w_t$$

pro $t = 1, 2, .., N$. Výsledek je dosti podobný předešlému. 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 přesné (tak, jak se tváříme při důvěřivé strategii), tj. bude $C_{t-1} = 0$, přejde odvozené řízení na předešlý případ.*

`Poznámka:` *Jak v předešlém tak i v tomto příkladě řízení jsme obdrželi tzv. jednokrokové řízení (horizont, který jsme potřebovali pro ustálení řídícího 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, závislý na předchozích akčních veličinách. Ten by vstupoval do dalšího kroku středování a minimalizace a vnášel by dynamiku do výpočtu řídícího zákona. K jeho ustálení bychom museli použít delší horizont, což úlohu syntézy značně komplikuje. Takovou syntézu jsme ukázali v kap. 5.*

Kapitola 5

Doplněk

5.1 Objekty a vzorce v predikci a identifikaci

V této části práce hovoříme o predikci výstupu soustavy a identifikaci parametrů jejího modelu. Tyto dvě úlohy jsou navzájem propojeny a při výpočtu predikce a odhadu parametrů lze sledovat obecné schéma výpočtů. Konkrétní výsledky se potom liší podle toho, jaký model jsme použili. V této práci se zabýváme třemi základními modely spojitým modelem s normálním šumem, spojitým modelem s rovnoměrným šumem a diskrétním modelem. Abychom neztratili návaznost na obecné schéma při práci s jednotlivými typy modelů, uvedeme toto schéma ihned v začátku tohoto Doplňku. Ve schématu se vyskytují, nebo se ho dotýkají, následující hustoty pravděpodobnosti nebo funkce:

Model

$$f(y_t|\varphi_t, \Theta) \tag{5.1}$$

Součin modelů

$$\Pi(D(t), \Theta) = \prod_{\tau=1}^{t} f(y_\tau|\varphi_\tau, \Theta) \tag{5.2}$$

Věrohodnostní funkce

$$L_t(\Theta) = \prod_{\tau=1}^{t} f(y_\tau|\varphi_\tau, \Theta) \tag{5.3}$$

Apriorní hp

$$f(\Theta|D(0)) \propto \prod_{\tau=-\kappa_0}^{0} f(y_\tau|\varphi_\tau, \Theta) \tag{5.4}$$

Aposteriorní hp

$$f(\Theta|D(t)) \propto f(y_t|\varphi_t, \Theta) f(\Theta|D(t-1)) \propto L_t(\Theta) f(\Theta|D(0)) \tag{5.5}$$

Prediktivní hp

$$f(y_t|u_t, D(t-1)) \underbrace{=}_{\text{p.p.ř.}} \int f(y_t|u_t, D(t-1), \Theta) f(\Theta|D(t-1)) \quad ^1 \tag{5.6}$$

Bodový odhad

$$\hat{\Theta}_t = E[\Theta|D(t-1)] \tag{5.7}$$

Bodová predikce

$$\hat{y}_t = E[y_t|u_t, D(t-1)] \tag{5.8}$$

5.2 Bodová predikce výstupu soustavy

Počítáme optimální bodovou predikci $\hat{y}_t$ výstupu soustavy, která minimalizuje kvadratické kriterium optimality

$$J_{\text{pre}} = E[(\hat{y}_t - y_t)^2]$$

Střední hodnotu v tomto kriteriu je třeba provést přes všechny veličiny, které mají statistickou souvislost s náhodnou veličinou y_t (tj. nejsou na ni nezávislé). Podle našich předpokladů, se kterými jsme také konstruovali model, souvisí y_t s řízením u_t a minulými daty $D(t-1)$. Bude proto

$$\begin{aligned}
\min_{\hat{y}_t} J &= \min_{\hat{y}_t} \int (y_t - \hat{y}_t)^2 f(y_t, u_t, D(t-1)) \mathrm{d}(y_t, u_t, D(t-1)) = \\
&= \min_{\hat{y}_t} \int \int (y_t - \hat{y}_t)^2 f(y_t|u_t, D(t-1)) f(u_t, D(t-1)) \mathrm{d}y_t \mathrm{d}(u_t, D(t-1)) = \\
&= \int \min_{\hat{y}_t} \left[\int (y_t - \hat{y}_t)^2 f(y_t|u_t, D(t-1)) \mathrm{d}y_t \right] f(u_t, D(t-1)) \mathrm{d}(u_t, D(t-1)) = \\
&= E[\min_{\hat{y}_t} E[(y_t - \hat{y}_t)^2|u_t, D(t-1)]].
\end{aligned}$$

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.

$$E[\hat{y}_t^2 - 2\hat{y}_t y_t + y_t^2|u_t, D(t-1)] =$$

$$= \hat{y}_t^2 - 2\hat{y}_t E[y_t|u_t, D(t-1)] + E[y_t^2|u_t, D(t-1)] =$$

$$\begin{aligned}
= \hat{y}_t^2 - 2\hat{y}_t E[y_t|u_t, D(t-1)] + E[y_t|u_t, D(t-1)]^2 + \\
+ E[y_t^2|u_t, D(t-1)] - E[y_t^2|u_t, D(t-1)]^2 =
\end{aligned}$$

$$(\hat{y}_t - E[y_t|u_t, D(t-1)])^2 + D[y_t|u_t, D(t-1)],$$

kde $D[.|.]$ značí podmíněný rozptyl.

[1]Označení p.p.ř. připomíná platnost přirozených podmínek řízení.

Minima kriteria je dosaženo pro $\hat{y}_t = E[y_t|u_t, D(t-1)]$, kdy je kvadratický člen anulován a tedy minimální. Zbytek kriteria po minimalizaci, který již na argumentu minimalizace nezávisí, je $E[y_t^2|u_t, D(t-1)] - E[y_t|u_t, D(t-1)]^2 = D[y_t|u_t, D(t-1)] = D[y_t]$.

5.3 Identifikace s přepočtem 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.

▷ `Přepočet kovariance parametrů` C_t
Vyjdeme ze vztahu $S_u^2(t) = S_u^2(t-1) + u_t^2$, který plyne z přepočtu informační matice $V(t-1) \Rightarrow V(t)$. Protože $C_t = \frac{1}{S_u^2(t)}$, budeme se snažit uvedený přepočet 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

$$C_t = C_{t-1} + \frac{1}{S_u^2(t-1) + u_t^2} - \frac{1}{S_u^2(t-1)} = C_{t-1} - \frac{u_t^2}{(S_u^2(t-1) + u_t^2)S_u^2(t-1)},$$

kde v čitateli jsme využili vztah $S_u(t) = S_u^2(t-1) + u_t^2$. Dále vytkneme ze závorky ve jmenovateli $S_u^2(t-1)$ a dostaneme

$$C_t = C_{t-1} - \frac{1}{1 + \frac{u_t^2}{S_u^2(t-1)}} \left(\frac{u_t}{S_u^2(t-1)} \right)^2.$$

Protože $1/S_u^2(t-1) = C_{t-1}$ a $C_{t-1}u_t^2 = \zeta_t$, platí

$$C_t = C_{t-1} - \frac{1}{1+\zeta_t}(C_{t-1}u_t)^2,$$

což je požadovaný vztah pro C_t.

`P o z n á m k a:` *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.*

▷ `Přepočet 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 dále $S_{yu}(t) = S_{yu}(t-1) + y_t u_t$. Obě závorky, do nichž jsme dosadili, roznásobíme a upravíme. Postupně dostaneme

$$\hat{\theta}_t = \left[C_{t-1} - \frac{1}{1+\zeta_t}(C_{t-1}u_t)^2 \right] [S_{yu}(t-1) + y_t u_t]$$

$$= C_{t-1}S_{yu}(t-1) + C_{t-1}y_tu_t - \frac{1}{1+\zeta_t}[(C_{t-1}u_t)^2S_{yu}(t-1) + (C_{t-1}u_t)^2y_tu_t] =$$
$$= \hat{\theta}_{t-1} + \frac{1}{1+\zeta_t}[(1+\zeta_t)C_{t-1}y_tu_t - (C_{t-1}u_t)^2S_{yu}(t-1) - (C_{t-1}u_t)^2y_tu_t],$$

kde jako první člen jsme obdrželi minulé odhady parametrů, druhý člen jsme vsunuli do hranaté závorky. Roznásobením vsunutého členu dostaneme člen

$$\zeta_tC_{t-1}y_tu_t = C_{t-1}u_t^2y_tu_t$$

který se odečte s posledním členem hranaté závorky. Potom dostaneme

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{1+\zeta_t}[C_{t-1}y_tu_t - C_{t-1}^2u_t^2S_{yu}(t-1)].$$

Vytknutím členu $C_{t-1}u_t$ a označením $C_{t-1}S_{yu}(t-1) = \hat{\theta}_{t-1}$ dostaneme hledaný vztah

$$\hat{\theta}_t = \hat{\theta}_{t-1} + \frac{1}{1+\zeta_t}C_{t-1}u_t(y_t - \hat{\theta}_{t-1}u_t)$$

▷ **Přepočet zbytku kvadratické formy** Λ_t

Vyjdeme z 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

$$\Lambda_t = S_y^2(t) + y_t^2 - \left(C_{t-1} - \frac{1}{1+\zeta_t}(C_{t-1}u_t)^2\right)(S_{yu}(t-1) + y_tu_t)^2.$$

Protože je $\zeta_t = C_{t-1}u_t^2$ je prostřední závorka

$$C_{t-1} - \frac{1}{1+\zeta_t}(C_{t-1}u_t)^2 = \frac{C_{t-1}}{1+\zeta_t}.$$

Abychom dostali Λ_{t-1} ubereme a přidáme člen $C_{t-1}S_{yu}(t-1)$. 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

$$\Lambda_t = S_y^2(t-1) - C_{t-1}S_{yu}(t-1) + \frac{1}{1+\zeta_t}[y_t^2 - 2y_t\hat{\theta}_{t-1}u_t + (\hat{\theta}_{t-1}u_t)^2]$$

Odtud plyne dokazovaný vztah

$$\Lambda_t = \Lambda_{t-1} + \frac{1}{1+\zeta_t}(y_t - \hat{\theta}_{t-1}u_t)^2$$

5.4 Regresní model s rovnoměrným rozdělením šumu

V této části uvedeme stručná odvození některých výsledků z příkladu, který pojednává o identifikaci statického regresního modelu s rovnoměrně rozděleným šumem. Jedná se o Příklad 2.3 z Kapitoly 2.

Bodový odhad parametru θ

Podle obecného výsledku (2.7) je bodový odhad podmíněná střední hodnota

$$\hat{\theta}_t \underbrace{=}_{\text{definice}} \int_{-\infty}^{\infty} \theta f(\theta|y(t))\mathrm{d}\theta \underbrace{=}_{\text{dosazeno za pdf}} \int_{-\infty}^{\infty} \theta \frac{1}{B_t^U - B_t^L}\chi(\theta \in [B_t^L;\ B_t^U])\mathrm{d}\theta =$$

... použijeme funkci χ tak, že integrujeme jen přes její nosič ...

$$= \frac{1}{B_t^U - B_t^L}\int_{B_t^L}^{B_t^U} \theta \mathrm{d}\theta \underbrace{=}_{\text{integrace přes } \theta} \frac{1}{B_t^U - B_t^L}\left[\frac{1}{2}\theta^2\right]_{B_t^L}^{B_t^U} =$$

... po dosazení mezí a algebraické úpravě dostaneme výsledek ...

$$= \frac{1}{2}\frac{1}{B_t^U - B_t^L}\left[(B_t^U)^2 - (B_t^L)^2\right] = \frac{B_t^U - B_t^L}{2} = \hat{\theta}_t$$

Prediktivní hp

Podle (2.10) je prediktivní hp

$$f(y_t|y(t-1)) \underbrace{=}_{\text{úplná pravděpodobnost}} \int_{-\infty}^{\infty} f(y_t|\theta)f(\theta|y(t-1))\mathrm{d}\theta =$$

$$\underbrace{=}_{\text{dosazeno za hp}} \int_{-\infty}^{\infty} \frac{1}{2}\chi(y_t \in [\theta-1;\ \theta+1])\frac{1}{B_{t-1}^U - B_{t-1}^L}\chi(\theta \in [B_{t-1}^L;\ B_{t-1}^U])\mathrm{d}\theta =$$

... součin indikátorových funkcí χ v argumentu integrálu
je opět indikátorová funkce s mezemi aposteriorní hp – viz (2.49) ...

$$= \frac{1}{2(B_{t-1}^U - B_{t-1}^L)}\int_{-\infty}^{\infty} \chi(\theta \in [\max\{B_{t-1}^L;\ y_t-1\}; \min\{B_{t-1}^U;\ y_t+1\}])\mathrm{d}\theta =$$

$$= \frac{1}{2}\frac{\min\{B_{t-1}^U;\ y_t+1\} - \max\{B_{t-1}^L;\ y_t-1\}}{B_{t-1}^U - B_{t-1}^L} =$$

... rozepsáním pro jednotlivé intervaly dostaneme ...

$$= \begin{cases} \frac{1}{2}\frac{y_t+1-B_{t-1}^L}{B_{t-1}^U-B_{t-1}^L} & \text{pro} \quad y_t \in [B_{t-1}^L - 1;\ B_{t-1}^U - 1] \\ \frac{1}{2} & \text{pro} \quad y_t \in [B_{t-1}^U - 1;\ B_{t-1}^L + 1] \\ \frac{1}{2}\frac{B_{t-1}^U+1-y_t}{B_{t-1}^U-B_{t-1}^L} & \text{pro} \quad y_t \in [B_{t-1}^L + 1;\ B_{t-1}^U + 1] \\ 0 & \text{jinak.} \end{cases}$$

což je výsledek uvedený v (2.51).

P o z n á m k a: *Se stejným výsledkem bychom počítali i pomocí integrálů (2.11), 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 mezí aposteriorní hp (2.49), 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 predikce

Hodnota predikovaného výstupu soustavy je dána podmíněnou střední hodnotou výstupu

$$\hat{y}_t = \int y_t f(y_t|y(t-1))\mathrm{d}y_t$$

kde prediktivní hp $f(y_t|y(t-1))$ je (2.51). Výsledek uvedený v textu bychom získali naznačenou integrací přes jednotlivé úseky hp. Integrace je přímočará, ale poněkud zdlouhavá. Výsledek proto dokážeme jinak.

Z funkčního předpisu (2.51) je patrné, že prediktivní hp 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

$$\int_{-\infty}^{\infty}(x-m)f(x)\mathrm{d}x.$$

V integrálu provedeme substituci: $x - m = t,\ \mathrm{d}x = \mathrm{d}t$. Dostaneme

$$\int_{-\infty}^{\infty} tf(t+m)\mathrm{d}t.$$

Integrand je nyní lichou funkcí neboť je součinem funkce liché (t) a sudé ($f(t+m)$ – 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

$$\int_{-\infty}^{\infty}(x-m)f(x)\mathrm{d}x = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x - m = 0,$$

a tedy

$$\int_{-\infty}^{\infty}(x-m)f(x)\mathrm{d}x = m.$$

Střed symetrie m je tedy střední hodnotou. Konkrétně, pro prediktivní hp, je střední hodnotou, a tedy bodovou predikcí, hodnota $\hat{y}_t = \frac{B_{t-1}^L + B_{t-1}^U}{2}$

5.5 Regresní model s diskrétním rozdělením šumu

Bodový odhad parametru θ

Tento odhad je podle (2.7) dán jako podmíněná střední hodnota

$$\hat{\theta}_t = \int_0^1 \theta f(\theta|y(t))\mathrm{d}\theta =$$

... dosazeno za aposteriorní hp z (2.57) ...

$$= \frac{1}{B(\kappa_t^R, \kappa_t^L)} \int_0^1 \theta \cdot \theta^{\kappa_t^R - 1}(1-\theta)^{\kappa_t^L - 1}\mathrm{d}\theta = \frac{1}{B(\kappa_t^R, \kappa_t^L)} \int_0^1 \theta^{\kappa_t^R}(1-\theta)^{\kappa_t^L - 1}\mathrm{d}\theta =$$

... integrál je definicí beta funkce $B(\kappa_t^R + 1, \kappa_t^L)$
a podle vzorce $B(x+1,y)/B(x,y) = x/(x+y)$ pak dostaneme požadovaný výsledek ...

$$= \frac{B(\kappa_t^R + 1, \kappa_t^L)}{B(\kappa_t^R, \kappa_t^L)} = \frac{\kappa_t^R}{\kappa_t^R + \kappa_t^L}$$

Prediktivní hp

Podle (2.10) je prediktivní hp

$$f(y_t|y(t-1)) \underbrace{=}_{\text{definice}} \int f(y_t|\theta) f(\theta|y(t-1))\mathrm{d}\theta =$$

... dosadíme za model (2.54) a aposteriorní hp (2.57) v čase $t-1$ ...

$$= \int_0^1 \theta^{\delta(y_\tau,0)}(1-\theta)^{\delta(y_\tau,1)} \cdot \frac{1}{B(\kappa_{t-1}^R, \kappa_{t-1}^L)} \theta^{\kappa_{t-1}^R - 1}(1-\theta)^{\kappa_{t-1}^L - 1}\mathrm{d}\theta =$$

$$= \frac{1}{B(\kappa_{t-1}^R, \kappa_{t-1}^L)} \int_0^1 \theta^{\kappa_{t-1}^R + \delta(y_\tau,0) - 1}(1-\theta)^{\kappa_{t-1}^L + \delta(y_\tau,1) - 1}\mathrm{d}\theta =$$

... tuto pravděpodobnost vyčíslíme pro jednotlivé hodnoty y_t ...

$$= \begin{cases} \frac{1}{B(\kappa_{t-1}^R, \kappa_{t-1}^L)} \int_0^1 \theta^{\kappa_{t-1}^R - 1}(1-\theta)^{\kappa_{t-1}^L}\mathrm{d}\theta = \frac{\kappa_{t-1}^R}{\kappa_{t-1}^R + \kappa_{t-1}^L} & \text{pro } y_t = 0 \\ \frac{1}{B(\kappa_{t-1}^R, \kappa_{t-1}^L)} \int_0^1 \theta^{\kappa_{t-1}^R}(1-\theta)^{\kappa_{t-1}^L - 1}\mathrm{d}\theta = \frac{\kappa_{t-1}^L}{\kappa_{t-1}^R + \kappa_{t-1}^L} & \text{pro } y_t = 1 \end{cases}$$

Bodová predikce

Podobně jako bodový odhad parametru je podle (2.7) i bodová predikce výstupu dána jako podmíněná střední hodnota

$$\hat{y}_t = \sum_{y_t=0}^{1} y_t f(y_t|y(t-1)) = 0 \cdot \frac{\kappa_{t-1}^R}{\kappa_{t-1}^R + \kappa_{t-1}^L} + 1 \cdot \frac{\kappa_{t-1}^L}{\kappa_{t-1}^R + \kappa_{t-1}^L} = \frac{\kappa_{t-1}^L}{\kappa_{t-1}^R + \kappa_{t-1}^L}$$

Literatura

[1] J. Anděl. *Matematická statistika.* SNTL, Praha, 1978.

[2] J. Hátle, J. Likeš. *Základy počtu pravděpodobnosti a matematické statistiky.* SNTL, Praha, 1974.

[3] I. Nagy, M. Kárný, J. Novovičová, and M. Valečková. Mixture-model identification in traffic control problems. Technical Report 1973, ÚTIA, Prague, 1999.

[4] M. Kárný, I. Nagy. Dynamical bayesian decision-making: part I. Technical Report 1971, ÚTIA, Prague, 1999.

[5] J. Novovičová. Pravděpodobnost a matematická statistika. Skripta FD ČVUT, Praha, 1999.

[6] V. Peterka. Bayesian approach to system identification. In P. Eykhoff, editor, *Trends and Progress in System Identification*, pages 239–304. Pergamon Press, Oxford, 1981.

[7] R. C. Rao. *Lineární metody statistické indukce a jejich aplikace.* ACADEMIA, Praha, 1978.