O bayesovském učení

Ivan Nagy, Petr Nedoma, Miroslav Kárný, Lenka Pavelková, Pavel Ettler

Článek podává základní informace o bayesovském přístupu k identifikaci systémů, o tzv. bayesovském učení. Výklad je demonstrován na jednoduchém příkladu, kterým je hod potenciálně poškozenou mincí. Jsou zde vyzdviženy přednosti bayesovského učení ve srovnání s obvyklým učením metodami klasické statistiky. Tyto přednosti spočívají především v možnosti zahrnout apriorní informace do procesu odhadování a možnosti dát smysl i odhadům z velmi malého množství dat.

Klíčová slova:
Model systému, apriorní hustota pravděpodobnosti, aposteriorní hustota pravděpodobnosti, věrohodnostní funkce, bodové odhady parametrů, předpověď budoucích dat.

1. Úvod

Statistika slouží pro poznávání a předvídání jevů v reálných podmínkách neurčitosti. Její popisné i analytické nástroje jsou mimořádně bohaté ([1], [2], [3]) a jen specialisté jsou schopni je účinně využít. I ti jsou však často nuceni zkoušet, která z dostupných a potenciálně slibných metod je vhodná v uvažovaném konkrétním případě.

Bayesovská statistika je mnohdy chápána jako jedna varianta ze zmíněného arzenálu statistiky. Tento příspěvek, který je úvodem k dalším dvěma článkům zaměřeným na praxi, se pokouší ukázat bayesovskou statistiku jako způsob myšlení, který:

- umožňuje nerozporně využít teoretickou, experimentální a expertní znalost,
- poskytuje nejen odhady neznámých veličin, ale i informaci o jejich přesnostech, a to i v konečných časech pozorování,
- soustřeďuje pozornost uživatele na modelování jeho specifického problému a nikoliv na výběr statistické metody.

Význam pro praxi zvláště posledního rysu nelze přecenit; případné neúspěchy ve zpracování nejsou subjektivní chybou způsobenou špatným výběrem metody, ale buď jsou dány objektivně, nebo mohou být zlepšeny lepším modelováním.

V tomto článku budeme zmíněné vlastnosti ilustrovat na „školním“ příkladu odhadování vlastností hodů s poškozenou mincí. Tím uvedeme specialisty z jiných oblastí do myšlenkové struktury bayesovského zpracování informací, aniž bychom text přetížili technickými detaily. Příspěvek slouží jako úvod ke zmiňovaným článkům a může být využi i jako úvod systematičtějšího studia, pro které je možné doporučit [4], [5].

2. Parametrický model pozorovaných dat

Házíme opakovaně mincí a sledujeme, zda padne rub, označený číslem 0, či líc, označený číslem 1.

2.1 Pozorovaná data

Pozorovaná data tvoří posloupnost

$$y(T) = [y_1; y_2 \dots y_T] \qquad (1)$$

v níž $y_t \in \{0; 1\}$

kde y_t označuje výsledek t-tého pokusu.

Pozn.: V souvislosti se zavedeným značením posloupnosti dat používáme toto značení: $y(t) = [y(t-1); y_t]$. To je třeba chápat v souladu s konvencí jazyka Matlab, kdy $[a_1; [a_2; a_3]] = [a_1; a_2; a_3]$.

Modelovaný proces je náhodný, neboť nejsme schopni zajistit trvale stejné podmínky házení. Jsme tedy schopni nejvýše stanovit stupeň očekávání různých možných výsledků, stanovit pravděpodobnosti $f(y(T))$ všech možných konkrétních výsledků $y(T)$. Jako příklad jedné takové posloupnosti výsledků uveďme

$$y(10) = [1; 0; 1; 1; 1; 0; 1; 0; 1; 1] \qquad (2)$$

2.2 Teoretický model

Pro zpracování a využití výsledků pokusu je třeba tyto pravděpodobnosti $f(y(T))$ popsat, je třeba vytvořit model procesu házení. Proto uvažujeme možné chování procesu, vytváříme jeho teoretický model. Zde např. předpokládáme, že výsledek t-tého hodu není ovlivněn ani výsledky hodů minulých ani budoucích. Dále předpokládáme, že pravděpodobnost líce Θ se nemění s číslem hodu, tj. s „časem“ t. Protože nevíme, zda mince je poškozená či nikoliv, je teoreticky motivovaný parametrizovaný model procesu v jednom časovém okamžiku t

$$f(y_t|\Theta) = \Theta^{y_t}(1-\Theta)^{1-y_t}$$
$$y_t \in \{0; 1\} \qquad (3)$$
$$\Theta \in \langle 0; 1 \rangle$$

2.3 Věrohodnostní funkce

Tento model je znám až na neznámý časově neproměnný parametr Θ. Označíme-li $\nu_{1;T}$ počet, kolikrát se v posloupnosti $y(T)$ vyskytl líc, a $\nu_{0;T}$ kolikrát rub, tj.

$$\nu_{1;T} = \sum_{t=1}^{T} y_t$$

$$\nu_{0;T} = \sum_{t=1}^{T} (1-y_t)$$

lze model celé soustavy zapsat ve tvaru

$$f(y(T)|\Theta) = \prod_{t=1}^{T} f(y_t|\Theta) =$$
$$= \prod_{t=1}^{T} \Theta^{y_t}(1-\Theta)^{1-y_t} = \Theta^{\nu_{1;T}}(1-\Theta)^{\nu_{0;T}} \qquad (4)$$

V první rovnosti jsme využili podmínku nezávislosti hodů.

Pravděpodobnost pozorovaných dat s konkrétně dosazeným pozorováním a nahlížená jako funkce neznámého parametru se nazývá věrohodnostní funkce. Pro zdůraznění této závislosti je označována

$$L_T(\Theta) = f(y(T) = \text{měřená data} \mid \Theta)$$

V našem případě

$$L_T(\Theta) = \Theta^{\nu_{1;T}}(1-\Theta)^{\nu_{0;T}}$$

s pevnými hodnotami $\nu_{1;T}$ a $\nu_{0;T}$. Pro vzorek (2) je $L_{10} = \Theta^7(1-\Theta)^3$.

Pozn.: Zde použité předpoklady nejsou univerzální. Například výsledek t-tého pokusu může dynamicky záviset na několika či všech předchozích pokusech. Pak by bylo nutné zvolit jinou parametrizaci a věrohodnostní funkce by mohla dostat obecnější tvar

$$f(y(T)|\Theta) = \prod_{t=1}^{T} f(y_t|y(t-1), \Theta)$$

Ani uvažovaná data nejsou jednoznačně dána. Například bychom jako vstup u_t do procesu házení mohli uvažovat, zda před t-tým hodem položíme na dlaň minci navrch lícem ($u_t = 1$) či rubem ($u_t = 0$). Pak by potřebný parametrizovaný model dosáhl svého (téměř) nejobecnějšího tvaru

$$f(y(T), u(T)|\Theta) =$$
$$= \prod_{t=1}^{T} f(y_t|y(t-1), u(t), \Theta)$$
$$f(u_t|y(t-1), u(t-1), \Theta)$$

modelující obecný dynamický řízený proces.

Slovo “téměř” znamená, že máme ještě možnost připustit změny neznámého parametru Θ a modelovat měnící se vlastnosti mince, která může např. padat do bláta a měnit s každým hodem své fyzikální vlastnosti.

Tyto možnosti zde dále nerozvíjíme. Uvádíme je však proto, abychom zdůraznili, že volba společně zpracovávaných dat a parametrizovaného modelu je hlavním nástrojem při řešení konkrétních problémů.

3. Problémy učení

Na uvažovaném příkladu ukážeme, jak bayesovskou metodikou řešit problémy učení, tj. odhadování parametru modelu a předpovídání budoucích hodnot dat. Lze ukázat, že výsledky učení jsou základními prvky potřebnými pro řešení rozhodovacích úloh, jako je testování hypotéz o alternativních modelech, návrh systému sázení na budoucí hody, návrhy diagnostických systémů, zpětnovazebního řízení či poradních systémů pro operátory ([5], [6], [7]).

Zde se omezíme na odhadování, tj. na kvantitativní výroky o neznámém parametru založené na pozorovaných datech, a na předpovídání, tj. na kvantitativní výroky o budoucích datech založené na datech pozorovaných.

4. Obvyklé učení

Pro pochopení možností bayesovského učení načrtneme standardní (klasický, fisherovský) přístup k problémům učení.

Věrohodnostní funkce $L_T(\Theta)$ říká, jak je pravděpodobný pozorovaný datový vzorek pro různé hodnoty parametru. Lze tedy očekávat, že nejlepší odhad $\hat{\Theta}_T$ parametru Θ maximalizuje věrohodnostní funkci. Tento maximálně věrohodný odhad se vskutku hojně používá. Odhad, jako funkce náhodných dat, je náhodný a je možné studovat jeho rozložení i asymptotické vlastnosti, tj. chování pro $T \to \infty$. Odvodit rozložení maximálně věrohodného odhadu je obecně obtížné. Byla však dokázána řada obecných dobrých, asymptoticky zaručených vlastností ([1], [2]).

Pro házení mincí je maximálně věrohodný odhad dán vztahem

$$\hat{\Theta}_T = \frac{\nu_{1;T}}{\nu_{1;T} + \nu_{0;T}} = \frac{\nu_{1;T}}{T} \qquad (5)$$

kde $\frac{\nu_{1;T}}{T}$ je relativní četnost padlých líců v celém vzorku. Pro konkrétní výběr (2) je $\hat{\Theta}_{10} = 0{,}7$.

Odpovídající předpověď další hodnoty dat $\hat{y}$ je možné zkonstruovat na základě parametrizovaného modelu (3) $f(y_T{+}1|\hat{\Theta})$, přičemž neznámá hodnota parametru Θ se nahradí jeho odhadem $\hat{\Theta}_T$. Budoucím hodnotám y_{T+1} přiřazujeme pravděpodobnosti $f(y_{T+1}|\hat{\Theta}_T)$.

Pro konkrétní výběr (2) očekáváme, že v jedenáctém hodu padne líc s pravděpodobností $\hat{\Theta}_{10} = 0{,}7$ a rub s pravděpodobností $1 - \hat{\Theta}_{10} = 0{,}3$.

Protože klasické postupy vycházejí pouze z naměřeného datového vzorku, výsledky mohou být pro velmi krátké časové horizonty nerozumné. Ze stejného důvodu nelze do procesu odhadování zavést expertní znalost. Například není formální prostor pro znalost vycházející z hodnocení vzhledu a stupně poškození mince.

4.1 Příklad krátkého horizontu při klasickém učení

Nerozumnost odhadu s krátkým horizontem budeme demonstrovat pro jediný pozorovaný hod s $T = 1$. Podle (5) je odhadem $\hat{\Theta}_1$ parametru Θ relativní četnost padlých líců při jednom hodu, tj. $\hat{\Theta}_1 = 1$, jestliže padl líc, a $\hat{\Theta}_1 = 0$ v případě, že padnul rub. Tudíž naše předpověď je $f(y_2|\hat{\Theta}_1) = \hat{\Theta}_1$, že s jistotou v následujícím hodu padne to, co padlo právě teď.

5. Bayesovské učení

5.1 Bayesův vzorec

Parametry odhadujeme podle Bayesova vzorce. V něm vystupují hustoty pravděpodobnosti (zkratkou hp) dvou typů objektů. Jsou to data $y(T)$, nesoucí informaci o neznámých parametrech, a tyto parametry Θ. Bayesův vzorec s těmito objekty má tvar

$$f(\Theta|y(T)) = \frac{f(y(T)|\Theta)f(\Theta)}{f(y(T))} \propto f(y(T)|\Theta)f(\Theta) \qquad (6)$$

kde jsou

$f(\Theta) = f(\Theta|y(0))$ apriorní hp, vyjadřující očekávání různých hodnot Θ přiřazených expertem ještě před začátkem odhadování,

$f(y(T)|\Theta)$ věrohodnostní funkce (4),

$f(\Theta|y(T))$ aposteriorní hp, která vyjadřuje expertní stupeň očekávání různých hodnot Θ, korigovaný pozorovanými daty $y(T)$, $\Theta \in \langle 0; 1 \rangle$.

Pozn.: Aposteriorní hustota pravděpodobnosti $f(\Theta|y(T))$ je funkcí Θ a data $y(T)$ zde vystupují jako konstanty. Faktor $1/f(y(T))$ je proto jen pouhá normalizační konstanta, kterou jsme vypustili a nahradili znaménkem úměrnosti $\propto$. Normalizační konstantu lze kdykoliv dopočítat tak, aby integrál z hp byl jednotkový.

Dosadíme-li za věrohodnostní funkci podle (4), dostaneme odhad parametru Θ ve tvaru

$$f(\Theta|y(T)) \propto \left[\prod_{t=1}^{T} f(y_t|\Theta)\right] f(\Theta|y(0)) = \Theta^{\nu_{1;T}}(1-\Theta)^{\nu_{0;T}} f(\Theta|y(0)) \qquad (7)$$

Pozn.: Apriorní hp nese výchozí informaci o parametru Θ (např. pohledem na minci zjistíme, že není tak příliš poškozená, a že jako výchozí lze uvažovat hodnotu 0,5). Tato informace může být buď „expertní“, tj. zadaná na základě zkušenosti, nebo může být získána na základě tzv. apriorních dat, tj. dat změřených do začátku odhadování.

Kombinací expertní znalosti s apriorními daty je metoda tzv. fiktivních dat, kdy expert sestavuje fiktivní pokus tak, aby data z něho odpovídala jeho apriorním představám. Pomocí nich je potom vytvořena apriorní hp $f(\Theta|y(0))$.

V Bayesově vztahu vystupuje věrohodnostní funkce a apriorní hp v součinu. Z výpočetních důvodů je vhodné, aby obě hp měly strukturálně shodný tvar.

5.2 Apriorní hustota pravděpodobnosti pro hod mincí

Apriorní hp zavedeme v obdobném tvaru, jako má věrohodnostní funkce

$$f(\Theta|y(0)) \propto \Theta^{n_{1;0}-1}(1-\Theta)^{n_{0;0}-1} \qquad (8)$$

kde $n_{1;0}$ a $n_{0;0}$ jsou apriorní statistiky, blíže popsané v kap. 5.4, a –1 v exponentu je vyčleněna z formálně výpočetních důvodů.

5.3 Aposteriorní hustota pravděpodobnosti pro hod mincí

Uvedenou volbou apriorní hp zajistíme stejný tvar aposteriorní hp, tj.

$$f(\Theta|y(T)) \propto \Theta^{n_{1;T}-1}(1-\Theta)^{n_{0;T}-1} \qquad (9)$$

5.4 Přepočet statistik pro hod mincí

Rekursi pro statistiky dostaneme dosazením (8) a (9) do Bayesova vztahu (7) a porovnáním exponentů. Tak dostaneme

$$n_{1;T} = n_{1;0} + \nu_{1;T}$$
$$n_{0;T} = n_{0;0} + \nu_{0;T} \qquad (10)$$

kde

$$\nu_{0;T} = \sum_{\tau=1}^{T}(1-y_\tau)$$

je počet padlých rubů a

$$\nu_{1;T} = \sum_{\tau=1}^{T} y_\tau$$

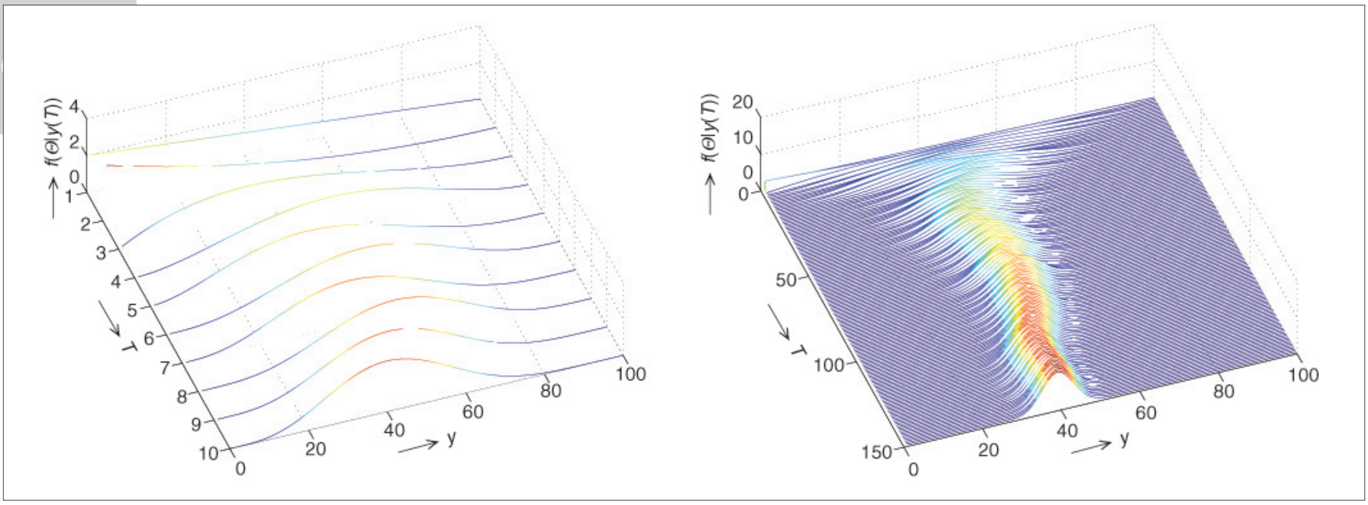
je počet padlých líců v T hodech.

Pro odhad tedy stačí kromě apriorních statistik $n_{1;0}$ a $n_{0;0}$ pamatovat si jen dvě čísla $\nu_{1;T}$ a $\nu_{0;T}$, ve kterých je uložena informace získaná z dat. Odtud je také vidět význam proměnných $n_{1;0}$ a $n_{0;0}$. Lze je chápat jako počty dat ve fiktivním experimentu.

5.5 Výsledky pro hod mincí

Bayesovským odhadem je obecně aposteriorní hp, která udává rozdělení pravděpodobnosti pro všechny možné hodnoty neznámého parametru.

Bodovým odhadem Θ, minimalizujícím kvadratickou vzdálenost odhadu od skutečné



Obr. 1. Odhadování s různou délkou datového vzorku a nulovou apriorní informací

hodnoty parametru po T měřeních, je podmíněná střední hodnota aposteriorní hp

$$\hat{\Theta}_T = E\left[\Theta|y(T)\right] = \int_0^1 \Theta\, f\left(\Theta|y(T)\right) \mathrm{d}\Theta = \frac{n_{1;T}}{n_{1;T}+n_{0;T}} \qquad (11)$$

Je vidět, že tento výraz je formálně shodný s výsledkem (5), který jsme obdrželi metodou maximální věrohodnosti. Statistiky $n_{1;T}$ a $n_{0;T}$ jsou ale součtem četnosti příslušné strany $\nu_{1;T}$ a $\nu_{0;T}$ a apriorních statistik $n_{1;0}$ a $n_{0;0}$. Apriorní informace jej tak může výrazně ovlivnit.

Lze ukázat, že rozptyl D odhadu Θ, vyjadřující neurčitost odhadu $\hat{\Theta}_T$, je

$$D\left(\Theta|y(T)\right) \approx \frac{1}{T}\hat{\Theta}_T\left(1-\hat{\Theta}_T\right) \qquad (12)$$

odkud je vidět, že odhad parametru se průběžně zpřesňuje v tom smyslu, že neurčitost aposteriorní hp je T-krát menší než apriorní hp.

Předpověď budoucí hodnoty dat y_{T+1} je udána prediktivní hp

$$f\left(y_{T+1}|y(T)\right) = \int_0^1 f\left(y_{T+1}|\Theta\right) f\left(\Theta|y(T)\right) \mathrm{d}\Theta \qquad (13)$$

$$\begin{aligned} y_{T+1} = 1 &\Rightarrow f\left(y_{T+1}|y(T)\right) = \hat{\Theta}_T \\ y_{T+1} = 0 &\Rightarrow f\left(y_{T+1}|y(T)\right) = 1-\hat{\Theta}_T \end{aligned} \qquad (14)$$

Pro bodový odhad $\hat{y}_{T+1}$ tedy platí formálně stejný vztah jako při klasickém učení. Jeho význam je však poněkud jiný, zvláště v případě krátkého horizontu.

Uvedené skutečnosti budeme ilustrovat v následujících příkladech.

5.6 Příklad krátkého horizontu při bayesovském učení

Budeme sledovat příklad (4.1) krátkého horizontu z kapitoly 4.1 o klasickém učení a ukážeme jeho bayesovské řešení. Budeme opět uvažovat odhad parametrů a predikci budoucího hodu y_2 při jediném pozorovaném hodu v $T = 1$. Přitom chceme zajistit stejné podmínky experimentu, tj. v případě bayesovského učení chceme uvažovat co nejslabší apriorní informaci. Tu vyjádříme pomocí rovnoměrného rozdělení apriorní hp. Podle (8) takovou hp dostaneme v případě volby $n_{1;0} = n_{0;0} = 1$. Statistiky v čase $T = 1$ budou podle (10)

$$n_{1;1} = 1 + y_1$$
$$n_{0;1} = 1 + \left(1 - y_1\right)$$

Odtud plyne

$$\hat{\Theta}_1 = \frac{n_{1;1}}{n_{1;1}+n_{0;1}}$$

$$y_1 = 1 \Rightarrow \hat{\Theta}_1 = 2/3$$
$$y_1 = 0 \Rightarrow \hat{\Theta}_1 = 1/3$$

Předpověď výsledku budoucího hodu (např. předpověď líce) je $f(y_2 = 1|y_1) = \hat{\Theta}_1$ rovna 2/3 pro případ, že v prvního hodu padl líc, a 1/3 pro případ, že v prvním hodu padl rub. Předpověď budoucího výstupu není jednoznačná 1 nebo 0, ale připouští vždy s pravděpodobností 1/3 i druhou alternativu.

5.7 Příklad bodových odhadů při učení

Uvažujeme příklad hodu poškozenou mincí, kdy z deseti hodů padl sedmkrát líc (viz datový soubor (2)). Ukážeme výpočet odhadu parametru Θ pomocí obvyklého a bayesovského učení.

5.7.1 Obvyklé učení

Počet líců je 7, počet hodů 10. Odhad parametru Θ je

$$\hat{\Theta}_{10} = \frac{7}{10} = 0{,}7$$

5.7.2 Bayesovské učení se slabší apriorní informací

Předpokládejme, že poškození mince je zanedbatelné, a tedy obě pravděpodobnosti by měly být stejné a rovny 0,5. Touto skutečností si ale nejsme příliš jisti. Zvolíme proto apriorní statistiky $n_{1;0} = 1$ a $n_{0;0} = 1$, které odpovídají „dvěma apriorním datům“. Apriorní hp potom odpovídá odhadu 1/(1+1) = 0,5, konečné statistiky jsou $n_{1;10} = 1 + 7 = 8$ a $n_{0;10} = 1 + 3 = 4$ a konečný odhad je

$$\hat{\Theta}_{10} = \frac{8}{12} = 0{,}\overline{6}$$

5.7.3 Bayesovské učení se silnější apriorní informací

Uvažujeme stejnou situaci jako v předchozím případě, ale již jsme se z dřívějších pokusů přesvědčili, že tak malé poškození, jaké je u naší mince, nemá téměř žádný vliv. Statistiky proto volíme $n_{1;0} = 100$ a $n_{0;0} = 100$, odpovídající „dvěma stům apriorních dat“. Apriorní odhad je stejný: 100/(100+100) = 0,5, konečné statistiky jsou $n_{1;10} = 100 + 7 = 107$ a $n_{0;10} = 100 + 3 = 103$. Konečný odhad je

$$\hat{\Theta}_{10} = \frac{107}{210} = 0{,}51$$

Vidíme, že obě apriorní informace „přitáhly“ výsledek k a priori předpokládané hodnotě 0,5, avšak větší hodnoty statistik mají proti informaci z dat větší vliv. Je to pochopitelné. První apriorní informace odpovídá jednou a priori změřenému líci a jednomu rubu. Druhá odpovídá 100 lícům a 100 rubům. Proti nim vždy stojí deset změřených dat a odhad je optimálním kompromisem mezi apriorní informací a informací získanou z dat.

Úplným popisem parametru Θ je ale aposteriorní hp. Na dalším příkladu ukážeme její vývoj během rekursivního odhadu a její výsledný tvar pro různou apriorní informaci.

5.7.4 Příklad aposteriorní hustoty pravděpodobnosti při bayesovském učení

Uvažujeme stejný hod mincí jako v předchozím příkladu, ale budeme sledovat celou aposteriorní hp neznámého parametru Θ, a to pro různý počet měřených dat a různou apriorní informaci. Data budeme získávat na simulovaném experimentu, kde volíme skutečnou, tj. simulovanou, hodnotu parametru $\Theta = 0{,}4$. Pro ilustraci budeme v jednotlivých případech uvádět trojrozměrný graf, ve kterém budou za sebou řazeny hp parametrů tak, jak se měnily s přibývajícím počtem zpracovaných dat. Nejstarší průběhy hp jsou v grafu vzadu, směrem dopředu postupují novější.

Zdrojový kód simulace v jazyce Matlab je uveden v tab. 1 a lze si jej stáhnout z adresy http://www.automa.cz/download/0588.txt

5.7.5 Srovnání krátkého a dlouhého datového vzorku

Se zvětšujícím se počtem zpracovaných dat se zpřesňuje odhad. To se projeví snižováním rozptylu aposteriorní hp. Na *obr. 1* jsou dva experimenty s parametrem $\Theta = 0{,}4$. Levý je s deseti daty, pravý se 150 daty.

Z levé části *obr. 1* je patrné, že aposteriorní hp z malého datového vzorku je dosti neurčitá (má velký rozptyl – šířku). Pravá část *obr. 1* ukazuje, jak se s přibývajícím počtem zpracovaných dat aposteriorní hp postupně zpřesňuje (rozptyl klesá). Konečný průběh (nejvíce vpředu) je již dosti přesný.

Začátky na obou částech *obr. 1* (tj. průběhy úplně vzadu) jsou velmi „nejisté“ a „roztěkané“. Je to dáno tím, že oba starty jsou bez jakékoliv apriorní informace a snaží se přesně respektovat veškerou informaci, kterou jednotlivá data přinášejí.

5.7.6 Vliv slabé apriorní informace

Tento experiment (*obr. 2*) odpovídá předchozímu se 150 daty, na začátku je ale aplikována apriorní informace. Tváříme se, jako bychom neznali skutečnou hodnotu parametru Θ, a zadáme apriorní informaci odpovídající nepoškozené minci. Nejdříve použijeme „slabou informaci“ s hodnotami apriorních statistik $n_{1;0} = n_{0;0} = 10$. Levý obrázek ukazuje vývoj aposteriorní hp, pravý vývoj bodového odhadu.

Z *obr. 2* je patrné, že počáteční vývoj aposteriorní hp se hezky „uklidnil“. Navíc hodnota koeficientu byla určena správně, protože na konci experimentu byla ne zcela správná apriorní informace „přebita“ měřenými daty a hodnota parametru Θ se odhadla správně: $\hat{\Theta}_T = 0{,}4$.

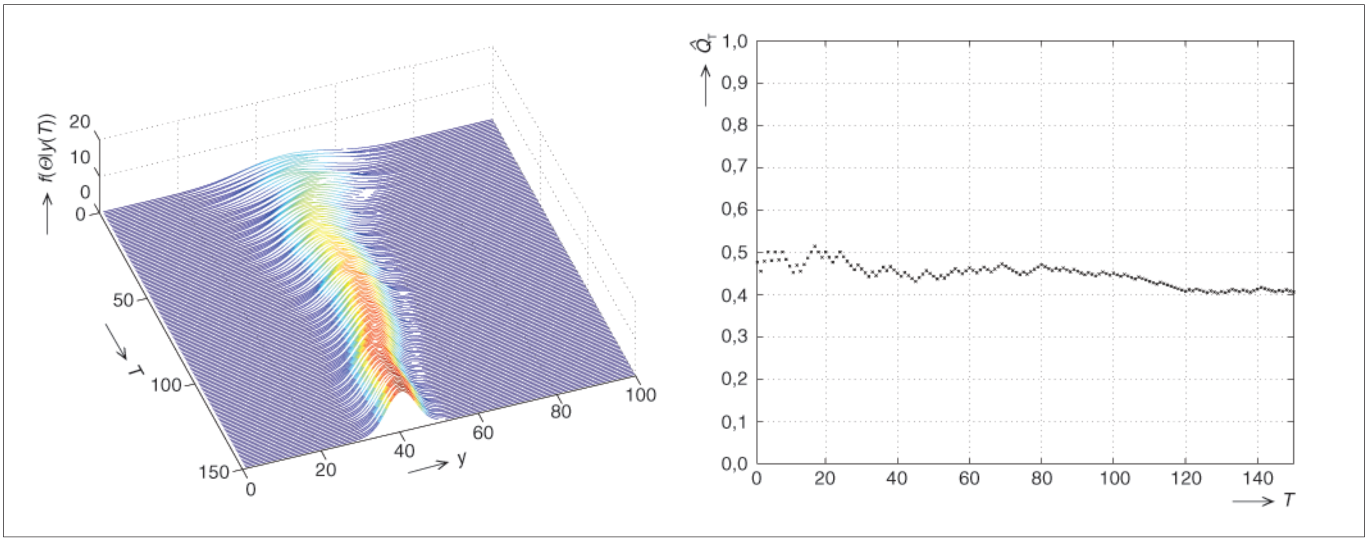
5.7.7 Vliv silné apriorní informace

Jde o stejný experiment, jako je předchozí, ale s apriorními statistikami $n_{1;0} = n_{0;0} = 100$. Výsledek ukážeme na obdobných obrázcích, jako jsou předchozí.

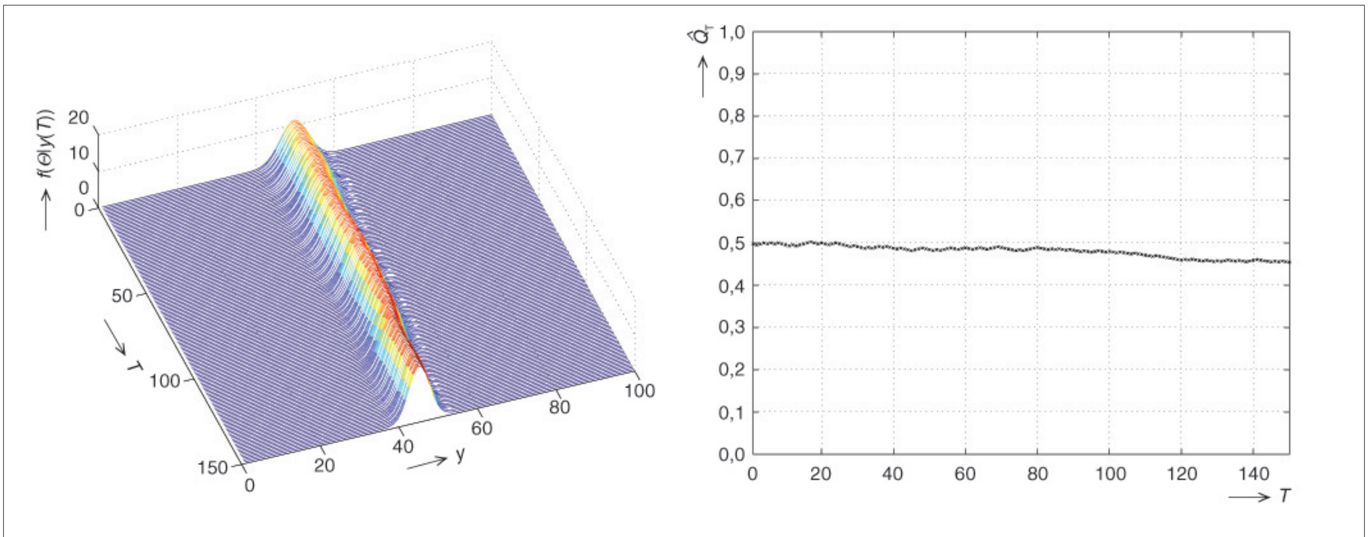
Z *obr. 3* je vidět, že použitá apriorní informace byla příliš silná – nedovolila, aby se prosadila správná hodnota parametru, která je 0,4.

6. Závěr

Článek je úvodem do problematiky bayesovského učení. Po něm budou následovat další



Obr. 2. Odhadování se slabou apriorní informací



Obr. 3. Odhadování se silnou apriorní informací

dva články, zabývající se teorií a aplikací bayesovského přístupu k odhadování směsí v praxi. Ty se používají k modelování složitých systémů pro podporu operátorů. Metody bayesovského učení jsou předkládány v jednoduché podobě na nejjednodušším statistickém příkladu, kterým je hod potenciálně poškozenou mincí. Na tomto příkladu jsou ilustrovány základní principy bayesovského učení, jsou vyzdviženy jejich přednosti a provedeno srovnání s obvyklým učením metodami klasické statistiky. V závěru příspěvku jsou uvedeny výsledky experimentů pro simulovaný pokus hodu mincí a demonstrován vliv apriorní informace při bayesovském učení.

Příklad je simulován v jazyce Matlab a jeho zdrojový kód je pro případné zájemce rovněž uveden nebo si jej lze stáhnout na adrese www.automa.cz/download/0588.txt. Všem, kdo si s ním zkusí pohrát, přejeme příjemnou zábavu.

Tento výzkum byl částečně podporován grantem EU IST-1999-12058 a grantem GA ČR č. 102/99/1564.

Literatura:

[1] HÁTLE, J. – LIKEŠ, J.: Základy počtu pravděpodobnosti a matematické statistiky. Praha, SNTL 1974.
[2] ANDĚL, J.: Matematická statistika. Praha, SNTL1978.
[3] RAO, R. C.: Lineární metody statistické indukce a jejich aplikace. Praha, Academia 1978.
[4] PETERKA, V.: Bayesian approach to system identification. In Trends and Progress in System Identification. P. Eykhoff (Ed.). Oxford, Pergamon Press 1981. Pp. 239-304.
[5] BEREC, L.: Algorithm for Determination of Model Structure of Predicted and/or Controlled Process. Tech. Rep., 1842, Praha, ÚTIA AV ČR 1995.
[6] KÁRNÝ, M. – NAGY, I. – NOVOVIČOVÁ J.: Quasi-bayes approach to multi-model fault detection and isolation. In: Adaptive Control and Signal Processing. John Willey and Sons, vol. 16, 2002, no. 1, pp. 61-83.
[7] KÁRNÝ, M. – BÖHM, J. – GUY, T. V. – JIRSA, L. – KANOURAS, A. – NAGY, I. – NEDOMA, P. – QUINN, A. – TESAŘ, L. – PARRY, D. – TICHÝ, M.: Prodactool background – theory, algorithms and software. Tech. Rep., 2001 (draft of the report).

Tab. 1 Výpis programu

```
% Vývoj aposteriorní hp pri hodu mincí s P(líc)=Ths
% s apriorní hp zadanou pomocí statistik n1 a n0
clc, clear all, clf, rand('seed',125)
% Zadání vstupních údaju
h=.01; % krok diskretizace
T=150; % pocet kroku simulace
Ths=.4;                                    % simulovaná pravdepodobnost lícu
n1=20; n0=20;                              % pocátecní hodnoty statistik

Thv=h:h:1;                                 % diskretizace hodnot parametru
yt=fix(rand(1,T)+Ths);                     % simulace mince
Thm=[]; Thh=[];

% Cyklus pro cas (zpracování dat)
for t=1:length(yt)
   n1=n1+yt(t); n0=n0+1-yt(t);             % prepocet statistik
   Tht=[];
   for Th=Thv
      fTh=Th^(n1)*(1-Th)^(n0);             % konstrukce aposteriorní hp
      Tht=[Tht fTh];
   end
   Tht=Tht/sum(Tht)/h; Thm=[Tht; Thm]; % normování hp
   The=n1/(n1+n0);     Thh=[Thh The];  % bodové odhady
end

% Kreslení výsledku
figure(1)
subplot(211),waterfall(Thm), grid on, view(-20,75)
subplot(212),plot(Thh,'x','markersize',3),grid  on,axis([1,T,0,1])
```

Ivan Nagy,
Fakulta dopravní ČVUT a Ústav teorie informace a automatizace AV ČR
(nagy@fd.cvut.cz),

Petr Nedoma,
Miroslav Kárný,
Lenka Pavelková,
Ústav teorie informace a automatizace AV ČR
({nedoma, school, pavelkov}@utia.cas.cz),
Pavel Ettler,
Compureg Plzeň
(ettler@compureg.cz)

Lektoroval: Prof. Ing. Vladimír Mařík, DrSc., katedra kybernetiky FEL ČVUT Praha (marik@labe.felk.cvut.cz)

Thomas Bayes – geniální outsider

Z četných matematiků a všestranných vědců, kteří v 18. a 19. století vytvořili základní práce v oblasti teorie pravděpodobnosti a statistiky, jsou nejčastěji uváděni bratři Bernoulliové, Euler, Laplace, de Moivre, Gauss a další. Členem této vybrané společnosti světem uznávaných vědců se stal i anglický duchovní, „geniální outsider“ – Thomas Bayes.

Pocházel z rodiny duchovního, který patřil ke skupině prvních šesti veřejně vysvěcených nonkonformistických duchovních v Anglii. (*Pozn. red.: Nonkonformisté se v sedmnáctém století odštěpili od státní anglikánské církve. Požadovali nezávislost anglikánské církve na státu, mj. to, aby jmenování duchovních nepodléhalo souhlasu státních orgánů a aby anglický panovník nestál automaticky v čele církve. Odmítali i některé církevní ceremoniály.*) Stejně jako otec, byl i Thomas vysvěcen nonkonformistickým duchovním a zpočátku působil jako otcův pomocník ve staršovstvu presbyteriánské církve v Holbornu. V roce 1720 se stal duchovním v presbyteriánské kapli v Tunbridge Wells, 60 kilometrů jižně od Londýna. Tuto funkci vykonával až do roku 1752, kdy odešel do penze, ale ve svém dosavadním působišti žil nadále. Stal se zámožným starým mládencem, a přestože žil v provinciálních podmínkách, udržoval kontakty s mnoha vzdělanými přáteli.

Bayes byl roku 1742 zvolen členem Royal Society, navzdory tomu, že v oblasti matematiky publikoval za svého života pouze jednu práci, a to navíc anonymně. Šlo o *Introduction to the Doctrine of Fluxions* (1736 – Úvod do nauky o infinitesimálním počtu; na vysvětlenou uveďme, že termínem „*fluxional calculus*“ I. Newton označoval infinitesimální počet). Bayes touto svou prací ostře nesouhlasně reagoval na spis *The Analyst* (1736), ve kterém irský teolog a filozof, biskup George Berkeley (1685 – 1753), napadl logické základy Newtonova infinitesimálního počtu. Druhým Bayesovým spisem, jenž vyšel za jeho života, byl *Divine Benevolence* (1731 – Božská dobrotivost).

Bayesova teorie pravděpodobnosti byla uvedena v jeho práci *Essay towards solving a problem in the doctrine of chances* (Esej směrovaná k řešení problému ve vědě o náhodě), publikované ve *Philosophical Transactions of the Royal Society of London* (Filozofická pojednání londýnské Královské společnosti) v roce 1746 – tedy již po Bayesově smrti. Bayesovu práci zaslal do Royal Society jeho přítel Richard Price, který ji našel v písemnostech pozůstalosti.

Bayes se touto svou prací stal jedním ze zakladatelů induktivní statistiky (statistická analýza opírající se o vzorky tvořící malé podíly základního statistického souboru). V době, kdy počet pravděpodobnosti byl víceméně ve svých počátcích, Bayes se již zabýval problémem, jak by bylo možné uvádět pozdější zkušenosti do souladu s apriorními předpoklady, resp. jak takové zkušenosti změní následné hodnocení situace proti apriorním předpokladům. Šlo tedy do jisté míry o dynamické ověřování hypotéz, včetně korektury těchto hypotéz.

Klasická matematická statistika Bayesovu teorii – Bayesova věta říká, jak změnit apriorní pravděpodobnost dodatečnou informací – odmítá. Bezpochyby k tomu přispěl i zmatek v tom, co Bayes skutečně řekl, co myslel a co bylo pozdějším výkladem do jeho úvah vloženo.

Přestože klasická matematická statistika uvrhla Bayesovu větu do klatby, objevili se její důslední obhájci. Patří mezi ně např. i jeden z nejvýznamnějších statistiků moderní doby Leonard Savage. Bayesova teorie tak právě v posledních letech prožívá své znovuzrození.

(Josef Heřmann)

Leonard J. Savage – znovuobjevitel bayesovského učení

Leonard Jimmie Savage se narodil 20. listopadu 1917 v Detroitu (USA). Vystudoval matematiku na Michiganské univerzitě. Habilitoval prací na téma metrika a diferenciální geometrie. Léta 1941–1942 strávil v Princetonu v Institute for Advanced Study. Zde se setkal s významným matematikem von Neumannem a pod jeho vlivem se v roce 1944 stal členem Statistical Research Group na Kolumbijské univerzitě (USA).

V roce 1954 vydal knihu *The Foundations of Statistic* (Základy statistiky). To je jeho zřejmě nejvýznamnější dílo. V knize jsou popsány principy subjektivní statistiky a účelových funkcí. Speciální případy těchto funkcí byly objeveny von Neumannem a Morgensternem v jejich pracích o teorii her.

Další Savegeho prací je *How to gamble if you must: Inequalities for stochastic processes* (1965), napsaná společně s L. Dubinsem.

Další Savageho publikace a články se týkají statistické dedukce. Uvedl v nich mj. testy baysovských hypotéz a postup pro bayesovské odhady. Jeho bayesovský přístup je v rozporu s klasickou statistikou, reprezentovanou v moderní době především matematikem Fisherem.

Ne nevýznamné jsou i Savageho práce o filozofických základech statistiky.

(Bk)