O bayesovském učení

Ivan Nagy[a,b], Petr Nedoma[b], Miroslav Kárný[b],
Lenka Pavelková[b], Pavel Ettler[c]

[a] Fakulta dopravní ČVUT,
Na Florenci 25, 110 00 Praha 1, nagy@fd.cvut.cz
[b] Ústav teorie informace a automatizace AV ČR,
POB 18, 18208 Praha 8,
{school,nedoma,pavelkov}@utia.cas.cz
[c] Compureg Plzeň, s.r.o., POB Plzeň,
ettler@compureg.cz

9. září 2002

Abstrakt

Článek podává základní informace o bayesovském přístupu k identifikaci systémů, o tzv. bayesovském učení. Výklad je demonstrován na jednoduchém příkladě, 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 zahrnutí 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.

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

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 je schopni účinně využít. I ti jsou však často nuceni zkoušet, která z dostupných a potenciálně slibných metod se hodí na uvažovaný konkrétní případ.

Bayesovská statistika je často chápána jako jedna varianta ze zmíněného arzenálu statistiky. Tento článek, který je úvodem k dalším dvěma prakticky zaměřeným článkům, 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.

Praktický význam 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 se pokusíme výše uvedené vlastnosti ilustrovat na "školním" příkladu odhadování vlastností hodů s poškozenou mincí. Tím uvedeme specialisty z jiných oblastí do myšlenkové struktury bayesovského zpracování informaci aniž bychom přetížili text technickými detaily. Text slouží jako úvod zmiňovaným článkům a může sloužit jako úvod systematičtějšího studia, pro které můžeme doporučit [4], [5].

2 Parametrický model pozorovaných dat

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

Pozorovaná data

Pozorovaná data tvoří posloupnost

$$y(T) = [y_1, y_2, \cdots, y_T], \quad \text{v níž} \quad y_t \in \{0, 1\}, \tag{1}$$

kde

y_t označuje výsledek t-tého pokusu,
$y(T)$ značí celou posloupnost dat až do okamžiku T, tj. $[y_1, y_2, \ldots, y_T]$.

POZNÁMKA: *V souvislosti se zavedeným značením posloupnosti dat používáme následující značení $y(t) = [y(t-1), y_t]$. Tu 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 jeho 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 takové posloupnosti výsledků uveďme

$$y(10) = [1, 0, 1, 1, 1, 0, 1, 0, 1, 1], \quad T = 10. \tag{2}$$

Teoretický model

Pro zpracování a využití výsledků pokusu je třeba tyto *pravděpodobnosti* $f(y(T))$ popsat, je třeba vytvořit *model procesu* házení. Pro to uvažujeme o možném chování procesu, vytváříme jeho *teoretický model*. Zde například 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

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

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. $\nu_{1;T} = \sum_{t=1}^{T} y_t$ a $\nu_{0;T} = \sum_{t=1}^{T}(1-y_t)$, pak lze model celé posloupnosti dat $y(T)$ zapsat ve tvaru

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

kde 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 $L_T(\Theta) = f(y(T) =$ měřená data$|\Theta)$. V našem případě $L_T(\Theta) = \Theta^{\nu_{1;T}}(1-\Theta)^{\nu_{0;T}}$ s pevnými hodnotami $\nu_{1;T}\,\nu_{0;T}$. Pro vzorek (2) je $L_{10}(\Theta) = \Theta^7(1-\Theta)^3$.

POZNÁMKA: *Zde použité předpoklady nejsou univerzální. Například výsledek t-tého pokusu může dynamicky záviset na několika či všech předchozích pokusech. Pak by bylo nutno 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 svůj (téměř) nejobecnější tvar

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

modelujícího obecný dynamický řízený proces.

Slovo "téměř" znamená, že máme ještě možnost připustit změny neznámého parametru Θ a modelovat měnící se vlastnosti mince, která může například 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 jsou hlavními nástroji při řešení konkrétních problémů.

3 Problémy učení

Na uvažovaném příkladě ukážeme, jak bayesovskou metodikou řešit problémy učení, tj. *odhadování parametru* modelu a *předpovídání* budoucích hodnot dat. Dá se 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

$$\hat{\Theta}_T = \frac{\nu_{1;T}}{\nu_{1;T} + \nu_{0;T}} = \frac{\nu_{1;T}}{T}, \tag{5}$$

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 možno zkonstruovat na základě parametrizovaného modelu (3) $f(y_{T+1}|\Theta)$ s tím, že 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 11. hodu padne líc s pravděpodobností $\hat{\Theta}_{10} = 0.7$, a rub s pravděpodobností $1 - \hat{\Theta}_{10} = 0.3$.

Protože klasické postupy vychází 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 poškozenosti 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 padnutí rubu. 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

Bayesův vzorec

Odhadování parametru Θ provádíme podle Bayesova vzorce. V něm vystupují hustoty pravděpodobnosti (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

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

kde

$f(\Theta) = f(\Theta|y(0))$ je apriorní hp, vyjadřující očekávání různých hodnot $\Theta \in \langle 0; 1\rangle$ přiřazenou expertem ještě před začátkem odhadování,

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

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

POZNÁMKA: *Hp $f(\Theta|y(T))$ 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 kdykoli dopočítat tak, aby integrál z hp byl jednotkový.*

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

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

POZNÁMKA: *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 jeho zkušenosti, nebo může být získána na základě tzv. apriorních dat, tj. dat změřených do začátku odhadování.*

Jakousi 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.

Apriorní hp pro hod mincí

Apriorní hp zavedeme v obdobném tvaru, jako má věrohodnostní funkce

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

kde

$n_{1;0}, n_{0;0}$ jsou apriorní statistiky a -1 v exponentu je ze statistiky vyčleněna formálně z výpočetních důvodů.

Aposteriorní hp pro hod mincí

Uvedenou volbou apriorní hp zajistíme stejný tvar aposteriorní, tj.

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

Přepočet statistik pro hod mincí

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

$$\begin{aligned} n_{1;T} &= n_{1;0} + \nu_{1;T}, \\ n_{0;T} &= n_{0;0} + \nu_{0;T}, \end{aligned} \tag{10}$$

kde $\nu_{1;T} = \sum_{\tau=1}^{T} y_\tau$ je počet padlých líců a $\nu_{0;T} = \sum_{\tau=1}^{T}(1-y_\tau)$ je počet padlých rubů v T hodech.

Pro odhad stačí tedy kromě apriorních statistik $n_{1;0}$ a $n_{0;0}$ pamatovat 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.

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

$$\hat{\Theta}_T = E[\Theta|y(T)] = \int_0^1 \Theta \cdot f(\Theta|y(T))d\Theta = \frac{n_{1;T}}{n_{1;T} + n_{0;T}}. \tag{11}$$

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* Θ, vyjadřující neurčitost odhadu $\hat{\Theta}_T$, je

$$cov(\Theta|y(T)) \approx \frac{1}{T}\hat{\Theta}_T(1-\hat{\Theta}_T), \tag{12}$$

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

$$f(y_{T+1}|y(T)) = \int_0^1 f(y_{T+1}|\Theta)f(\Theta|y(T))d\Theta = \tag{13}$$

$$= \begin{cases} \hat{\Theta}_T & \text{pro } y_{T+1} = 1 \\ (1-\hat{\Theta}_T) & \text{pro } y_{T+1} = 0. \end{cases} \tag{14}$$

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.

5.1 Příklad krátkého horizontu při bayesovském učení

Budeme sledovat příklad (4.1) krátkého horizontu z kapitoly o klasickém učení a ukážeme jeho bayesovské řešení. Budeme opět uvažovat odhad parametru Θ a predikci budoucího hodu y_2 při jediném pozorovaném hodu v $T = 1$. Přitom chceme zajistit stejné podmínky experimentu, tj. v případě byesovského 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)

$$n_{1;1} = 1 + y_1, \qquad n_{0;1} = 1 + (1 - y_1).$$

Odtud plyne

$$\hat{\Theta}_1 = \frac{n_{1;1}}{n_{1;1} + n_{0;1}} = \begin{cases} 2/3 & \text{pro } y_1 = 1, \\ 1/3 & \text{pro } y_1 = 0. \end{cases}$$

Předpověď výsledku budoucího hodu (např. předpověď lícu) je

$$f(y_2 = 1|y_1) = \hat{\Theta}_1 = \begin{cases} 2/3 & \text{pro } y_1 = 1, \\ 1/3 & \text{pro } y_1 = 0. \end{cases} \tag{15}$$

Předpověď budoucího výstupu není jednoznačná 1 nebo 0, ale připouští vždy s pravděpodobností 1/3 i druhou alternativu.

5.2 Příklad bodových odhadů při učení

Uvažujeme příklad hodu poškozenou mincí, kdy z 10 hodů padl 7 krát líc – viz datový soubor (2). Ukážeme výpočet odhadu parametru Θ pomocí obvyklého a bayesovského učení.

Obvyklé učení

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

$$\hat{\Theta}_{10} = \frac{7}{10} = 0.7$$

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 $\frac{1}{1+1} = 0.5$, konečné statistiky jsou $n_{1;10} = 1 + 7 = 8$ a $n_{0;10} = 1 + 3 = 4$ a konečný odhad je

$$\hat{\Theta}_{10} = \frac{8}{12} = \frac{2}{3} = 0.\bar{6}.$$

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 jako je u naší mince nemá prakticky žádný vliv. Statistiky proto volíme $n_1(0) = 100$ a $n_0(0) = 100$, odpovídající "dvěma stům apriorních dat". Apriorní odhad je stejný $\frac{100}{100+100} = 0.5$, konečné statistiky jsou $n_1(10) = 100 + 7 = 107$ a $n_0(10) = 100 + 3 = 103$ a konečný odhad je

$$\hat{\Theta}_{10} = \frac{107}{210} = 0.51.$$

Vidíme, že obě apriorní informace "přitáhly" výsledek k apriori předpokládané hodnotě 0.5, avšak větší hodnoty statistik mají proti informaci z dat větší vliv. Je to pochopitelné. První apriorní informace odpovídá jednou apriori změřenému lícu a jednomu rubu. Druhá odpovídá 100 lícům a 100 rubům. Proti nim vždy stojí 10 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 příkladě ukážeme její vývoj během rekurzivního odhadu a její výsledný tvar pro různou apriorní informaci.

5.3 Příklad aposteriorní hp při bayesovském učení

Uvažujeme stejný hod mincí jako v předchozím příkladě, 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

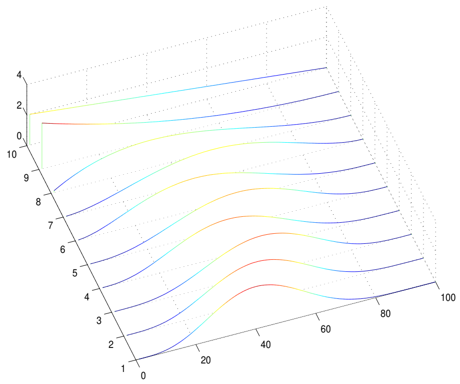
```
% Vývoj aposteriorní hp při hodu mincí s P(líc)=Ths
%  s apriorní hp zadanou pomocí statistik n1 a n0
```

```
clc, clear all, clf, rand('seed',125)
% Zadání vstupních údajů
h=.01;                      % krok diskretizace
T=150;                      % počet kroků simulace
Ths=.4;                     % simulovaná pravděpodobnost lícu
n1=20; n0=20;               % počáteční hodnoty statistik

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

% Cyklus pro čas (zpracování dat)
for t=1:length(yt)
  n1=n1+yt(t); n0=n0+1-yt(t);          % přepočet 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 výsledků
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])
```

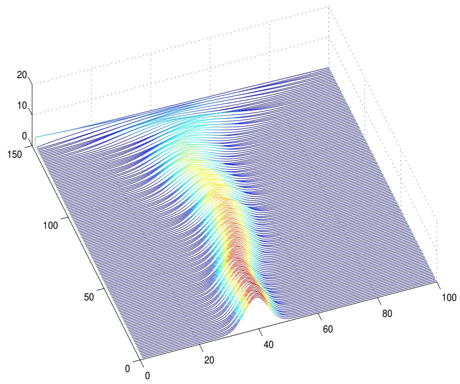
Srovnání krátkého a dlouhého datového vzorku

Se zvětšujícím se počtem zpracovaných dat dochází ke zpřesňování odhadu. To se projeví snižováním rozptylu aposteriorní hp. Na následujících obrázcích jsou dva experimenty s parametrem $\Theta = 0.4$. Levý s 10 daty, pravý se 150 daty.

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



Datový vzorek délky 10



Datový vzorek délky 150

Obrázek 1: Odhadování s různou délkou datového vzorku a nulovou apriorní informací

Začátek na obou částech obrázku 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.

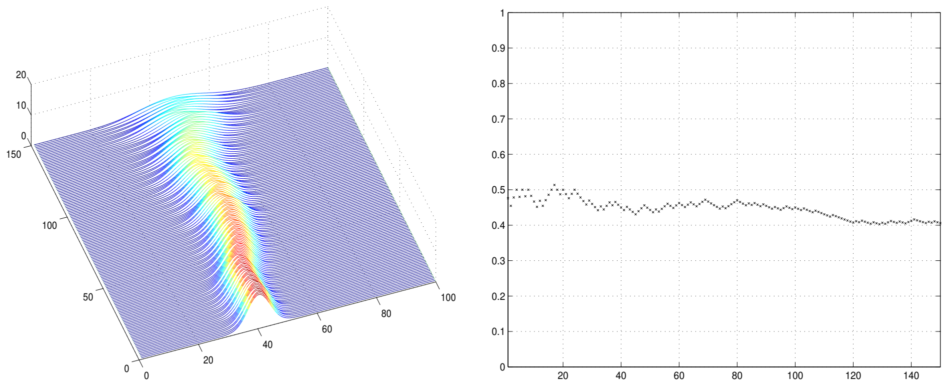
Vliv slabé apriorní informace

Tento experiment 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 obrázku 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, $\hat{\Theta}_T = 0.4$.

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



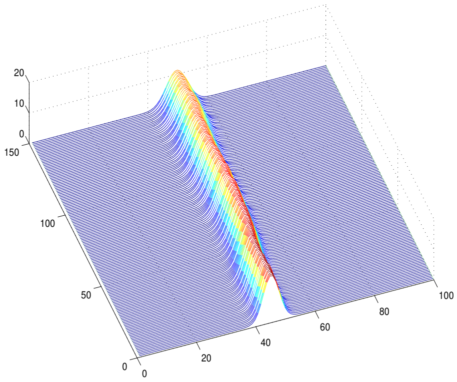
Vývoj aposteriorní hp　　Vývoj bodového odhadu parametru

Obrázek 2: Odhadování se slabou apriorní informací apriorní informací

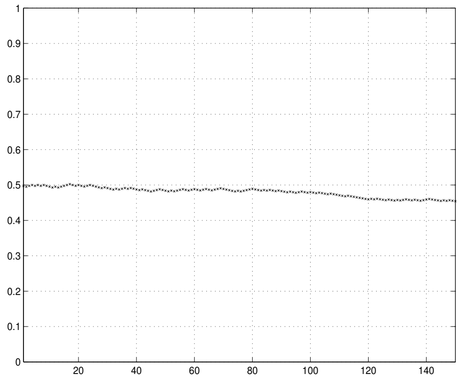
Z obrázku 3 je vidět, že použitá apriorní informace byla příliš silná – nedovolila, aby se prosadila správná hodnota parametru, která je 0.4.

6 Závěr

Článek představuje úvod do problematiky bayesovského učení. Po něm budou následovat další dva články, zabývající se teorií a praktickou aplikací bayesovského přístupu k odhadování směsí. Ty se používají k modelování složitých systémů pro podporu operátorů. Metody bayesovského učení jsou předkládány v jednoduché podobě na nejjednodušším statistickém příkladě, kterým je hod potenciálně poškozenou mincí. Na tomto příkladě jsou ilustrovány základní principy bayesovského učení, vyzdviženy jejich přednosti a provedeno srovnání s obvyklým učením metodami klasické statistiky. V závěru článku 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. Všem, kdo si s ním zkusí pohrát, přejeme příjemnou zábavu.



Vývoj aposteriorní hp



Vývoj bodového odhadu parametru

Obrázek 3: Odhadování se silnou apriorní informací apriorní informací

Reference

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

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

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

[4] V. Peterka, "Bayesian approach to system identification", in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.

[5] L. Berec, "Algorithm for Determination of Model Structure of Predicted and/or Controlled Process", Tech. Rep. 1842, ÚTIA AVČR, P.O.Box 18, 182 08 Prague, Czech Republic, 1995.

[6] M. Kárný, I. Nagy, and Jana Novovičová, "Quasi-bayes approach to multi-model fault detection and isolation", *Adaptive Control and Signal Processing, , John Willey and Sons*, vol. 16, no. 1, pp. 61–83, 2002.

[7] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, A. Kanouras, I. Nagy, P. Nedoma, A. Quinn, L. Tesař, D. Parry, and M. Tichý, "Prodactool background – theory, algorithms and software", Tech. Rep., 2001, Draft of the report, 313 pp.