

Bayesovská štatistika

M. Grendár

Abstrakt—Základy bayesovskej štatistiky sú ilustrované na príklade hádzania mincou.

I. ÚVOD

BAYESIÁNCOM je každý, kto robí štatistické úvahy pomocou Bayesovej vety. Bayesovská štatistika je koncepcne veľmi jednoduchá. Na príklade hádzania mincou sa pokúsime ilustrovať jej základné prvky. Posudzovanie modelu a bayesovské priemerovanie modelov budú predstavené prostredníctvom regresného modelu. Výpočtová stránka a niektoré ďalšie aspekty bayesovskej štatistiky sú spomenuté len okrajovo.

Začnime s nebayesovskou analýzou výsledkov hádzania mincou.

II. NEBAYESOVSKÁ ŠTATISTIKA: ANALÝZA HODOV MINCOU

Modelujme výsledok hodu mincou pomocou náhodnej premennej X ktorá nadobúda hodnoty z dvojprvkovej množiny $\mathcal{X} = \{0, 1\}$ s pravdepodobnosťou $P(X; \theta) = \theta^x(1 - \theta)^{1-x}$, kde parameter θ je pravdepodobnosť úspechu $P(X = 1)$. Rozdelenie $P(X; \theta)$ je bernoullijské a náhodný výber $X_1^n = X_1, X_2, \dots, X_n$ je známy aj ako bernoullijské schéma. Nech sa v n -tici hodov mincou vyskytoval úspech n_1 krát. Na základe tejto informácie chceme v rámci nebayesovskej štatistiky urobiť bodový a intervalový odhad hodnoty parametra θ , test hypotézy o θ a predpoveď výsledku budúceho hodu.

A. Odhad

V nebayesovskej štatistike existuje veľké množstvo metód na odhadovanie parametrov: metóda najväčšej vierohodnosti, momentová metóda, zovšeobecnená momentová metóda, metóda empirickej vierohodnosti, robustné odhadovanie, atď. Najčastejšie používaná je asi metóda najväčšej vierohodnosti (angl., *Maximum Likelihood*, ML), v ktorej sa za odhad $\hat{\theta}_{ML}$ parametra θ berie tá hodnota parametra, pri ktorej by daná realizácia výberu z danej parametrickej triedy rozdelení mala najväčšiu hodnotu vierohodnosti:

$$\hat{\theta}_{ML} = \arg \sup_{\theta \in \Theta} L(\theta; x_1^n),$$

kde, v tomto prípade, vierohodnostná funkcia (čiže pravdepodobnosť realizácie výberu, chápaná ako funkcia parametra) $L(\theta; x_1^n) = \binom{n}{n_1} \theta^{n_1} (1 - \theta)^{n - n_1}$ a parametrický priestor $\Theta = [0, 1]$. Lahko sa zistí, že $\hat{\theta}_{ML} = \frac{n_1}{n}$. V prípade tohto jednoduchého zadania by asi všetky nebayesovské odhadovacie metódy viedli na tento odhad.

Katedra matematiky, FPV UMB, Tajovského 40, 974 01 Banská Bystrica. Ústav matematiky a informatiky, SAV a UMB, Banská Bystrica. Ústav merania SAV, Bratislava. Email: marian.grendar@savba.sk. Podporené VEGA grantom 1/0077/09. Dátum: 14. september 2009.

Príklad: Nech $n = 20$, $n_1 = 5$. Potom $\hat{\theta}_{ML} = \frac{n_1}{n} = \frac{5}{20} = 0.25$. Teda, na základe 20-tich hodov mincou, v ktorých sa strana identifikovaná s $X = 1$ vyskytovala 5 krát, zoberieme ako odhad skutočnej, nám neznámej hodnoty parametra θ hodnotu 0.25. Tento bodový odhad by teda naznačoval, že minca je nevyvážená.

Ako prípravu na bayesovské úvahy poznamenajme, že mince zvyknú mať homogénne zloženie, bývajú vyvážené, takže by 0 a 1 mali padať s rovnakou pravdepodobnosťou. Otázkou je, ako zahrnúť túto mimodátovú informáciu do štatistických úvah, konkrétne do odhadovania parametra θ ? Nebayesovská štatistika na to nemá odpoveď. \diamond

B. Inferencia

Nebayesovský bodový odhad je v prípade bernoullijskej schémy veľmi jednoduchý. To isté sa už nedá povedať o inferenciách, teda o konfidenčných intervaloch a testoch. Tvorba konfidenčných intervalov pre θ stále zamestnáva nebayesovských štatistikov, ako o tom svedčí napr. prehľadový článok [21]. Jeden z najobľúbenejších konfidenčných intervalov pre θ je waldovský 95%-ný interval: $\hat{\theta}_{ML} \pm 1.96 \sqrt{\hat{\theta}_{ML}(1 - \hat{\theta}_{ML})/n}$, ktorý je založený na Centrálnej limitnej vete (alebo, v tomto prípade, ekvivalentne, asymptotickej normalite ML odhadu). Jeho pravdepodobnosť pokrytia je u výberov o veľkosti 20 iba 80%-ná¹. Preto bolo navrhnutých viacero korekcií waldovského intervalu. Iný populárny interval je Clopperov Pearsonov interval, zviazaný dualitou s rovnomenným testom.

Príklad (pokr.): Clopperov Pearsonov test dáva pre naše dáta p -hodnotu 0.0414, teda v teste nulovej hypotézy $H_0 : \theta = 0.5$ oproti alternatíve $H_1 : \theta \neq 0.5$, nulovú hypotézu, na konvenčnej 5%-tnej hladine významnosti zamietame. A 95%-ný konfidenčný interval je (0.086, 0.491). Asymptotický 95%-ný konfidenčný interval je (0.060, 0.440)². Tieto nebayesovské inferencie vedú k záveru, že minca je nevyvážená. \diamond

Konfidenčný interval je jeden z mnohých *keby*-konceptov (angl., *pre-data concept*) nebayesovskej štatistiky. *Keby* sme realizovali veľký (nekonečný) počet 20-tíc hodov peniazom, tak by daný $100(1 - \alpha)\%$ konfidenčný interval pokryl neznámu skutočnú hodnotu parametra v $100(1 - \alpha)\%$ ách prípadov. Či v prípade konkrétnej realizácie 20-tich hodov mincou daný interval pokryl alebo nepokryl skutočnú hodnotu parametra sa ale nedozvieme: buď pokryl, alebo nepokryl. Nie vždy je toto *keby*-uvažovanie zmysluplné, navrviac o tom, že v praxi je obyčajne dôležitejšie *ked'*-uvažovanie (angl., *post-data reasoning*): keď máme túto konkrétnu realizáciu, aká je

¹Ďalšou črtou tohto intervalu je, že ak $\hat{\theta}_{ML} = 0$, tak interval obsahuje, bez ohľadu na n , jediný bod: nulu.

²Vypočítané v R. Kód: `library(Hmisc); binconf(5, 20, method = 'all')`. R-kovský zdrojový kód ku všetkým výpočtom z tohto článku sa dá nájsť na www.savbb.sk/~grendar.

pravdepodobnosť, že skutočná hodnota parametra leží v tomto konkrétnom intervale? Táto otázka má zmysel v bayesovskej štatistike.

Testovanie hypotéz je notoricky problematické, viď napr. [2], [24], [29], [9].

C. Predikcie

V prípade náhodného výberu je predpovedanie veľmi jednoduché.

Príklad (pokr.): Pravdepodobnosť $P(X = 1 | x_1^n)$, že výsledok budúceho hodu mincou bude $X = 1$ je rovná θ . Aby sme ju odhadli, je prirodzené nahradiť neznáme θ jeho odhadom $\hat{\theta}_{ML}$, čo je v tomto prípade 0.25. Teda, podľa nášho odhadu padne v nasledujúcom hode $X = 1$ s pravdepodobnosťou 1/4. \diamond

III. BAYESOVSKÝ RÁMEC

Rámec, v ktorom sa deje bayesovská štatistika je daný vetou Bayesa³.

A. Bayesova veta

Bayesovský štatistik musí mimodátovú informáciu o parametre $\theta \in \Theta$ dáta-generujúceho rozdelenia pravdepodobnosti $p(X_1^n | \theta)$ vyjadriť vo forme apriórneho rozdelenia pravdepodobnosti $p(\theta)$. Prior, ako sa apriórne rozdelenie stručnejšie nazýva, vyjadruje štatistikovu neistotu o hodnote parametra θ . Po tom ako je získaná realizácia x_1^n výberu X_1^n , modifikuje bayesiánc svoju apriórnu informáciu pomocou Bayesovej vety

$$p(\theta | x_1^n) = \frac{p(x_1^n | \theta)p(\theta)}{\int_{\Theta} p(x_1^n | \theta)p(\theta) d\theta}, \quad (1)$$

a získa tak aposteriórne rozdelenie $p(\theta | x_1^n)$ parametra θ , pri daných dátach x_1^n . Posteriórne rozdelenie (alebo posterior) vyjadruje štatistikovu neistotu o θ po tom, ako boli pozorované dáta. Posterior samozrejme závisí aj od modelu – teda od apriórneho rozdelenia $p(\theta)$ a dáta-generujúceho rozdelenia $p(x_1^n | \theta)$.

Niektoré bayesovské úvahy sa dajú robiť aj bez menovateľa v Bayesovej vete, ktorý sa zvykne nazývať evidencia (angl., *evidence*), alebo aj marginálna vierohodnosť, alebo prediktívna hustota. V takom prípade je zvykom písať Bayesovu vetu v stručnejšom tvare:

$$p(\theta | x_1^n) \propto p(x_1^n | \theta)p(\theta),$$

alebo, neformálne, posterior \propto vierohodnosť \times prior.

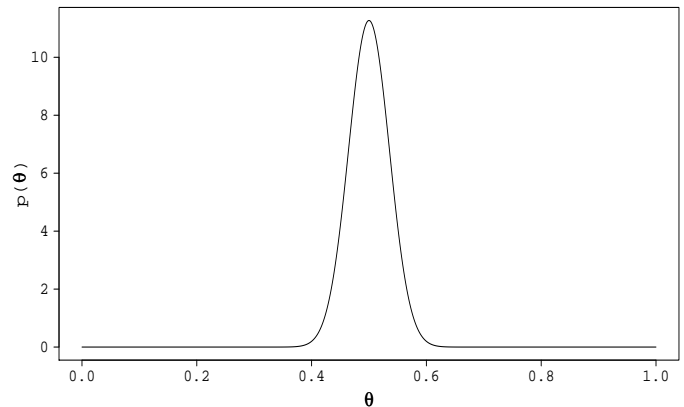
Príklad (pokr.): Neistota o parametre θ je v našom prípade dosť malá: mince sú väčšinou vyvážené. Ináč povedané, je dosť pravdepodobné, že náhodne vybratá minca, ktorou ideme hádzať, je vyvážená. Mieru neistoty o θ môžeme kvantifikovať napríklad nasledovne: $P(0.44 < \theta < 0.56) = 0.9$. Teda, apriórne sme si takmer istí (naša miera presvedčenia je 0.9), že θ (t.j., pravdepodobnosť padnutia tej strany mince, ktorú sme zviazali s hodnotou náhodnej premennej $X = 1$), je

v intervale (0.44, 0.56) – teda, že minca sa nevelmi odlišuje od vyvázenej. Ako sme spomínali, bayesiánc musí svoje apriórne presvedčenie vyjadriť pomocou prioru. Jeden možný prior, ktorý súhlasí s týmto apriórnym presvedčením je $\text{Beta}(\alpha = 100, \beta = 100)$.

Pripomeňme:

$$\text{Beta}(\theta; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1},$$

príčom $\alpha, \beta > 0$, $\theta \in [0, 1]$. Momenty: $E\theta = \frac{\alpha}{\alpha + \beta}$, $\text{Var}\theta = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$. A mód(θ) = $\frac{\alpha - 1}{\alpha + \beta - 2}$.



Obr. 1. Beta(100, 100) prior.

Skombinujeme

- Beta($\alpha = 100, \beta = 100$) prior $p(\theta)$
- s vierohodnosťou $p(n_1 | \theta) \propto \theta^{n_1} (1 - \theta)^{n - n_1}$ dát x_1^n v ktorých sa $X = 1$ vyskytovalo n_1 krát,
- pomocou Bayesovej vety,
- a dostaneme posterior $p(x_1^n | \theta)$. Dá sa uvidieť, že posteriórne rozdelenie je $\text{Beta}(\alpha + n_1 = 105, \beta + n - n_1 = 115)$.

Posterior ktorý sme obdržali patrí do tej istej triedy rozdelení ako prior. Prior môže mať akýkoľvek tvar. Prior, ktorý po bayesovskom skombinovaní s vierohodnostnou funkciou vedie na posterior patriaci do tej istej triedy rozdelení čo prior, sa nazýva konjugovaný (angl., *conjugate prior*); [9]. Pretože sú bayesovské výpočty s konjugovanými priormi oveľa ľahšie, zvyknú sa takéto priory tiež nazývať pohodlnými priormi (angl., *convenience priors*). Mimo pohodlnosti, neexistujú vo všeobecnosti dôvody prečo by mala byť apriórna informácia vyjadrená konjugovaným priorom.

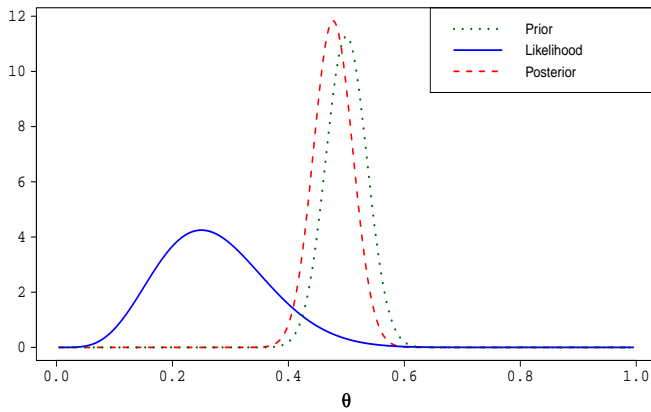
Vierohodnostná funkcia, prior a posterior sú zobrazené⁴ na obr. 2. Maximum vierohodnosti sa dosahuje v 0.25 čo je $\hat{\theta}_{ML}$. Prior je koncentrovaný okolo bodu 0.5. Dát je málo ($n = 20$), preto je aj posterior nimi len minimálne ovplyvnený. \diamond

B. Bodové charakteristiky posterioru

Posteriórne rozdelenie obsahuje všetku dostupnú informáciu: sumarizuje model a dáta.

³Zaujímavá diskusia Bayesovej vety, ako aj základná charakterizácia odlišnosti bayesovskej a nebayesovskej štatistiky je v práci [30], ktorej znalosť predpokladáme.

⁴`library(LearnBayes); prior = c(100, 100) # params of Beta prior; data = c(5, 15) # 5 successes and 15 failures; triplot(prior, data)`



Obr. 2. Triplot pre Beta(100, 100) prior.

V prípade, že je ale nutné zhrnúť posterior do jediného 'bodu', dá sa tak urobiť v rámci bayesovskej teórie rozhodovania; [9]. Rozhodovanie sa deje za neurčitosti. Nesprávne rozhodnutie má za následok stratu. Je nutné špecifikovať stratovú funkciu $L(\theta, a)$, ktorá udáva závislosť straty od toho ako sa líši bayesovský odhad $a \in \Theta$ od θ . Po tom ako bola špecifikovaná stratová funkcia je prirodzené hľadať také bodové zhrnutie a posteriôru (tzv. bayesovský estimátor/odhad), ktoré minimalizuje priemernú posteriornú stratu

$$\int_{\Theta} L(\theta, a) p(\theta | x_1^n) d\theta.$$

Niektoré základné stratové funkcie a príslušné bayesovské odhady:

- posteriorná stredná hodnota, minimalizuje kvadratickú stratovú funkciu $L(\theta, a) = (a - \theta)^2$,
- posteriorný medián, minimalizuje absolútnu stratovú funkciu $L(\theta, a) = |a - \theta|$,
- posteriorný móď, minimalizuje všetko-alebo-nič (angl., *zero-one*) stratovú funkciu.

Nie vždy je zmysluplné hľadať bodovú charakteristiku v rámci bayesovskej teórie rozhodovania. V mnohých prípadoch sa za bodový odhad berie posteriorná stredná hodnota. Ak má ale posterior viac než jeden móď, je zrejme, že v tomto prípade posteriorná stredná hodnota nie je dobrou bodovou charakteristikou.

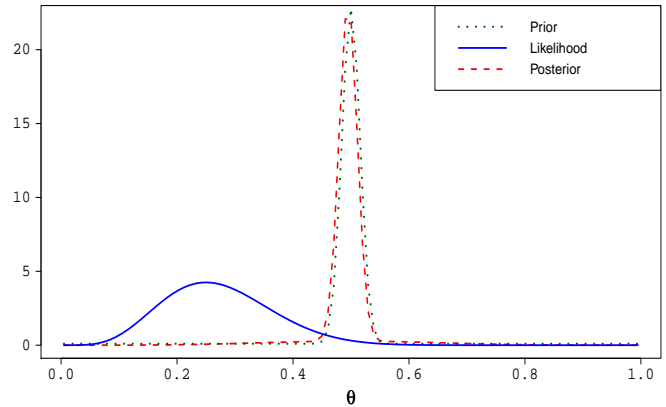
Príklad (pokr): Nakoľko posterior je Beta(105, 115), tak posteriorná stredná hodnota je $\frac{105}{105+115} = 0.47727$. Medián posteriôru je⁵ 0.47720, a posteriorný móď je 0.47706. \diamond

C. Bayesovská robustnosť

Do akej miery závisí posterior na voľbe prioru? Tejto otázke je v bayesovských výskumoch venovaná veľká pozornosť. Tvoria náplň toho, čo sa zvykne nazývať bayesovská robustnosť alebo analýza citlivosti (angl., *sensitivity analysis*), viď [17]. Navrhnuté boli viaceré techniky na zisťovanie citlivosti posteriorných úvah na priore, ako aj kvantitatívne miery robustnosti, viď napr. [2].

⁵qbeta(0.5, 105, 115)

Príklad (pokr): Obmedzíme sa len na veľmi primitívnu formu analýzy citlivosti. Skúsime zobrazíť iný prior, ktorý by vyhovoval nášmu apriórnemu presvedčeniu $P(0.44 < \theta < 0.56) = 0.9$. Napríklad prior daný ako nasledovná zmes hustôt $p(\theta) = 0.9 \text{Beta}(500, 500) + 0.1 \text{Beta}(1, 1)$ je taký. Aj posterior je potom zmes, a to $p(\theta | x_1^n) = 0.9 \text{Beta}(505, 515) + 0.1 \text{Beta}(6, 16)$. Na obrázku 3 je triplot⁶ prioru, vierohodnostnej funkcie a posteriôru.



Obr. 3. Triplot pre zmesový prior.

Posteriorná stredná hodnota je 0.472861, a teda je len o čosi menšia než hodnota 0.4772727, ktorú sme dostali pri Beta(100, 100) priore. \diamond

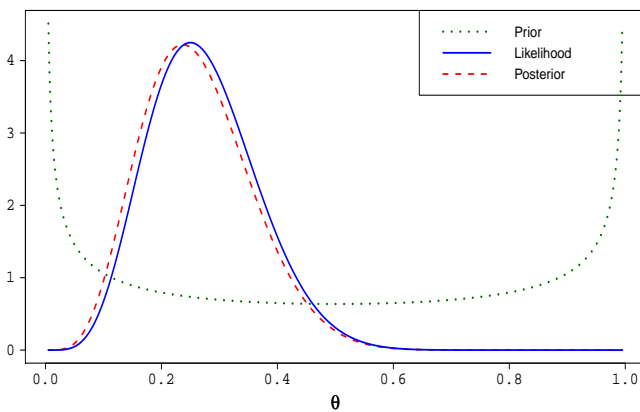
D. Objektívne priory

Nie vždy je k dispozícii taká jasná apriórna informácia ako v prípade hádzania mincou. Značná časť bayesovských výskumov je venovaná tomu, ako pretransformovať žiadne alebo veľmi slabé apriórne presvedčenie o skutočnej hodnote parametra do podoby apriórneho rozdelenia. Tento problém je taký zásadný, že rozdeľuje bayesiáncov na dve veľké skupiny: subjektivistov a objektivistov. So značnou dávkou zjednodušenia sa dá povedať, že subjektivisti sú presvedčení, že nemá zmysel hľadať akési automatické (nesubjektívne) metódy na konštrukciu priorov. Naopak, objektivisti aktívne navrhujú metódy na prekladanie neznalosti (angl., *ignorance*) do podoby apriórneho rozdelenia. Jeden z prominentných predstaviteľov objektívnej bayesovskej štatistiky, James Berger, vraví: "V objektívnej bayesovskej analýze sa priory volia tak, aby predstavovali 'neutrálne' znalosti o neznámých". Väčšina bayesiáncov zastáva pragmatický postoj: v prípade modelu s mnohými parametrami sú pre nepodstatné parametre a parametre o ktorých hodnotách sa apriórne toho veľa nevie použité neinformatívne priory, a pre ostatné parametre sa subjektívne špecifikujú informatívne priory.

⁶x = seq(0,1,0.001); curve(0.9*dbeta(x,505, 515) + 0.1*dbeta(x, 6,6), col = 'red') # posterior; curve(0.9*dbeta(x,500, 500) + 0.1*dbeta(x, 1,1), col = 'darkgreen', add = TRUE) # prior; curve(dbeta(x, 6, 16), col = 'blue', add = TRUE) # like; legend('topright', c("Prior", "Likelihood", "Posterior"), col = c("darkgreen", "blue", "red"))

Existuje veľké množstvo techník na konštrukciu objektívnych priorov. Asi najznámejšou je technika navrhnutá Haroldom Jeffreysov. Jeffreysov prior má tvar $p(\theta) \propto \sqrt{I(\theta)}$, kde $I(\theta)$ je Fisherova informácia pre parameter θ . Takto skonštruovaný prior spĺňa požiadavku invariantnosti: pri transformácii θ na $g(\theta)$ sa Jeffreysova apriórna hustota transformuje tak, ako sa má hustota pravdepodobnosti transformovať, viď napr. [9]. Jeffreysové priory sú zvyčajne neznormálnozovateľné, nazývajú sa preto pseudo-priory (angl., *improper priors*). Vie sa, že pre mnohé jeffreysovské priory býva výsledný posterior už normalizovateľný, teda riadny (angl., *proper prior*). Neznormálnozovateľnosť mnohých jeffreysovských priorov ale spôsobuje problémy pri bayesovskom porovnávaní hypotéz (pozri G). Iné veľmi populárne objektívne priory sú tzv. referenčné priory, viď [19]. V posledných rokoch sa rozbehlo štúdium a aplikácie tzv. slabo informatívnych priorov (angl., *weakly informative priors*).

Príklad (pokr.): Predpokladajme, že nemáme k dispozícii žiadnu apriórnu informáciu o minciach, netušíme, že mince sa razia strojovo zo značne homogénnych plechov kovu. Ináč povedané, apriórne nevieme o parametre θ bernoulliovského rozdelenia nič. Vyjadríme našu neznalosť jeffreysovským priorom: $p(\theta) \propto \theta^{-1/2}(1-\theta)^{-1/2}$, čo je Beta(0.5, 0.5) prior. Posterior je potom Beta(5.5, 15.5). Z obrázku 4 je vidieť, že jeffreysovský prior naozaj vnáša do úvah minimum mimodávajovej informácie.



Obr. 4. Triplot pre Jeffreysov prior.

Stredná hodnota posterioru je 0.262. Mód posterioru je 0.237. Obe hodnoty sú blízke k ML odhadu. \diamond

E. Intervalové charakteristiky posterioru

Existuje viacero možných intervalových charakteristík posterioru. V prípade, že posterior je približne symetrický, je rozumnou charakteristikou jeho rozpätia $100(1-\alpha)\%$ -ný posteriorný interval, mimo ktorého leží pod každým chvostom posterioru $100(\alpha/2)\%$ masy rozdelenia. Takýto bayesovský konfidenčný interval sa nazýva (rovnako)chvostový (angl., *equal-tail*). V opačnom prípade je lepšou intervalovou charakteristikou vrcholový posteriorný interval (angl., *highest posterior density interval*, HPD), ktorý pozostáva z takej oblasti $s(x_1^n)$ paramet-

rického priestoru Θ , pre ktorú

$$1) P(\theta \in s(x_1^n) | X_1^n = x_1^n) = 1 - \alpha,$$

$$2) \text{ ak } \theta_a \in s(x_1^n) \text{ a } \theta_b \notin s(x_1^n),$$

$$\text{potom } p(\theta_a | X_1^n = x_1^n) > p(\theta_b | X_1^n = x_1^n).$$

Samozrejme, možné sú aj iné konštrukcie intervalových charakteristík posterioru.

Príklad (pokr.): V prípade Beta(100, 100) prioru a našich dát je posterior (t.j., Beta(105, 115)) v podstate symetrický. 95%-ný chvostový posteriorný interval je (0.412, 0.543)⁷. Pri danom priore a dátach môžeme tvrdiť, že s 95%-nou pravdepodobnosťou leží neznáma hodnota parametra θ v intervale (0.412, 0.543).

Laplace, už v roku 1774 analyzoval bernoulliovskú schému pomocou rovnomerného prioru (ktorý je totožný s Beta(1, 1)). Posterior je v takom prípade Beta($n_1 + 1, n - n_1 + 1$). Pre naše dáta by v takomto prípade 95%-ný posteriorný chvostový interval bol (0.132, 0.437). \diamond

F. Testovanie v teórii rozhodovania

Testovanie hypotéz je možné robiť v rámci bayesovskej teórie rozhodovania. Formuluje sa nulová hypotéza $H_0 : \theta \in \Theta_0$ a alternatívna hypotéza $H_1 : \theta \in \Theta_0^c$. Možné sú dve akcie: akcia a_0 – prijatie H_0 , akcia a_1 – prijatie H_1 . Je nutné špecifikovať stratovú funkciu, ktorá vyjadruje veľkosť straty v prípade chyby prvého a chyby druhého druhu, keď sa vykoná akcia a .

Najbežnejšou stratovou funkciou je zovšeobecnená 0-1 strata:

$$L(\theta, a_0) = \begin{cases} c_{II} & \text{ak } \theta \in \Theta_0^c, \\ 0 & \text{ináč,} \end{cases}$$

$$L(\theta, a_1) = \begin{cases} c_I & \text{ak } \theta \in \Theta_0, \\ 0 & \text{ináč.} \end{cases}$$

Zovšeobecnená je v tom, že strata c_{II} v prípade chyby druhého druhu nemusí byť rovnaká ako strata c_I v prípade chyby prvého druhu. Bayesovský test spočíva v tom, že sa vyberie tá z hypotéz, ktorá aposteriórne spôsobí v priemere menšiu stratu. Posteriórna priemerná strata sa zvykne tiež nazývať riziko (angl., *risk*). Riziko má v prípade zovšeobecenej 0-1 straty tvar:

$$r(a_0 | x_1^n) = c_{II}[1 - P(H_0 \text{ true} | x_1^n)],$$

$$r(a_1 | x_1^n) = c_I P(H_0 \text{ true} | x_1^n).$$

Aby sme zamietli nulovú hypotézu musí byť teda $r(a_1 | x_1^n) < r(a_0 | x_1^n)$. To je ekvivalentné tvrdeniu, že

$$\text{zamietni } H_0 : \theta \in \Theta_0, \text{ ak } P(\theta \in \Theta_0 | x_1^n) < \frac{c_{II}}{c_I + c_{II}}.$$

V opačnom prípade sa H_0 prijíma. Ak sú straty u oboch chybných rozhodnutí rovnaké, tak test zamietla nulovú hypotézu ak je jej posteriórna pravdepodobnosť menšia než 1/2.

Príklad (pokr.): Chceme testovať hypotézu, že θ je z $\Theta_0 = (0.47, 0.53)$. Použijeme Beta(100, 100) prior $p(\theta)$. Prior určuje apriórnu pravdepodobnosť nulovej hypotézy $p_0 = \int_{\Theta_0} p(\theta) d\theta$,

⁷qbeta(0.025, 105, 115); qbeta(0.975, 105, 115)

čo je 0.604⁸. Apriórne sme teda o čosi viac presvedčení o platnosti nulovej hypotézy, než o alternatíve. Predpokladajme že straty sú rovnaké u oboch chybných rozhodnutí. Nakoľko $P(\theta \in \Theta_0 | x_1^n) = 0.526^9$, tak nulovú hypotézu nezamietame. Všimnime si, že aposteriórne, po tom ako sme obdržali dáta, naše presvedčenie o platnosti nulovej hypotézy, v porovnaní s apriórnym presvedčením, kleslo. \diamond

Za zmienku stojí, že ak je nulová hypotéza bodová, potom je bayesovské testovanie problematické, pretože si vyžaduje priradenie pravdepodobnostnej miery jednoprvkovej množine. Keď už bayesiánci musia testovať bodovú nulovú hypotézu, robia tak pomocou prioru obsahujúceho diracovskú mieru vo vyšetřovanom bode.

G. Porovnávanie hypotéz

Testovanie v rámci teórie rozhodovania je síce principiálne, ale zato veľakrát príliš zväzujúce. V prípadoch, keď nie je možné alebo žiaduce formulovať stratovú funkciu, má bayesiánc možnosť namiesto testovania, hypotézy H_0 a H_1 porovnať. Robí sa to pomocou veľmi prirodzeného nástroja, ktorý sa nazýva bayesovský faktor (angl., *Bayes factor*). Ak je prior riadny, potom je možné definovať apriórny pomer šancí (angl., *prior odds ratio*):

$$\frac{p_0}{p_1} = \frac{\int_{\Theta_0} p(\theta) d\theta}{\int_{\Theta_1} p(\theta) d\theta},$$

a aposteriórny pomer šancí (angl., *posterior odds ratio*):

$$\frac{p_{n,0}}{p_{n,1}} = \frac{\int_{\Theta_0} p(\theta | x_1^n) d\theta}{\int_{\Theta_1} p(\theta | x_1^n) d\theta}.$$

Bayesovský faktor v prospech H_0 je

$$BF_{01} = \frac{\text{posterior odds}}{\text{prior odds}} = \frac{p_{n,0}/p_{n,1}}{p_0/p_1}.$$

Nakoľko nás skôr zaujíma či nulovú hypotézu zamietame, než jej potvrdenie, zvykne sa robiť s bayesovským faktorom BF_{10} proti H_0 . Už Jeffreys navrhol kalibráciu BF_{10} podľa ktorej sa určuje, aký silný je dôkaz o neplatnosti H_0 . Pred nedávnom bola táto kalibrácia mierne modifikovaná v práci [27]. Uvedená je v Tabuľke 1.

BF_{10}	dôkaz o neplatnosti H_0
1 až 3	takmer žiadny
3 až 20	mierny
20 až 150	silný
> 150	veľmi silný

Tabuľka I
KALIBRÁCIA BAYESOVSKÉHO FAKTORA BF_{10} .

Príklad (pokr.): Prior nech je $\text{Beta}(100, 100)$; Potom BF_{10} proti $H_0 : \theta \in (0.47, 0.53)$ je 1.374^{10} , teda nie je takmer žiadny dôkaz toho, že by H_0 bola neplatná. \diamond

⁸ $u = \text{pbeta}(0.53, 100, 100)$; $l = \text{pbeta}(0.47, 100, 100)$; $p_0 = u - l$

⁹ $u_p = \text{pbeta}(0.53, 105, 115)$; $l_p = \text{pbeta}(0.47, 105, 115)$; $pn_0 = u_p - l_p$

¹⁰ $p_1 = 1 - p_0$; $pn_1 = 1 - pn_0$; $BF_{10} = pn_1/pn_0 / (p_1/p_0)$

Mnohí bayesovskí štatistickí považujú hypotézy za neúčinnú formu štatistického usudzovania. Napr. v [3] sa hypotézy spomínajú len v súvislosti s kritikou nebayesovskej štatistiky.

H. Predikcie

Modelovanie sa často robí za účelom predpovedania, robenia inferencií o nových pozorovaniach \tilde{x} , teda, prediktívnych inferencií. Bayesiáncom na to slúži posteriórna prediktívna hustota (angl., *posterior predictive density/distribution*)

$$p(\tilde{x} | x_1^n) = \int p(\tilde{x} | \theta, x_1^n) p(\theta | x_1^n) d\theta. \quad (2)$$

Tento dôležitý koncept sa nepoužíva len na predikcie ale aj na posúdenie (angl., *validation*) modelu (t.j., prioru a dátá-generujúceho rozdelenia).

Podobne sa zavádza aj apriórna prediktívna hustota (angl., *prior predictive distribution*)

$$p(\tilde{x}) = \int p(\theta) p(\tilde{x} | \theta) d\theta,$$

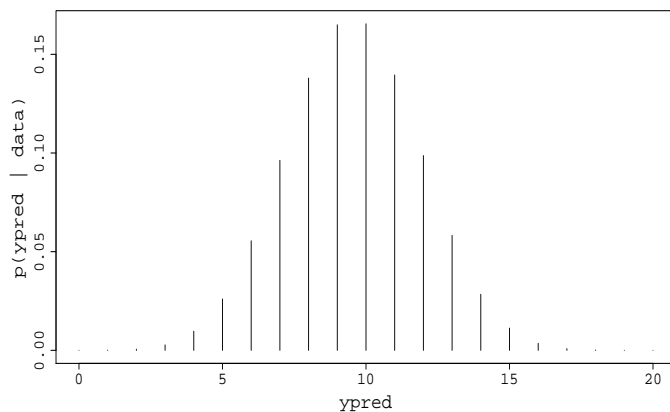
nazývaná tiež marginálna hustota (angl., *marginal density*).

Príklad (pokr.): Posteriórna prediktívna pravdepodobnosť toho, že v budúcom výbere o veľkosti m bude \tilde{y} výskytov $X = 1$ je (viď [1])

$$p(\tilde{y} | x_1^n) = \binom{m}{\tilde{y}} \frac{B(\alpha + \tilde{y}, \beta + m - \tilde{y})}{B(\alpha, \beta)}, \quad (3)$$

kde $\tilde{y} = 0, 1, \dots, m$ a $B(\cdot, \cdot)$ je beta funkcia; α, β sú parametre posterioru.

Prior nech je $\text{Beta}(100, 100)$, a nech $m = 20$. Posteriórne prediktívne rozdelenie¹¹ je na obr. 5.



Obr. 5. Posteriórne prediktívne rozdelenie $p(\tilde{y} | x_1^n)$.

92%-ný posteriórny prediktívny interval pre \tilde{y} , pri daných dátach a priore je¹² [6, 13]. Teda, s 92%-nou pravdepodobnosťou bude v budúcnosti 20-tich hodoch mincou počet úspechov ($X = 1$) niekde medzi šesť až trinásť.

V prípade, že by nás zaujímala pravdepodobnosť toho, že v budúcom jedinom hode nastane úspech, potom by sme

¹¹`library(LearnBayes); ab = c(105, 115) # posterior; m = 20; ys = 0:20; pred = pbetap(ab, m, ys); plot(ys, pred, type = 'h')`

¹²`discint(cbind(ys, pred), 0.9)`

podľa (3) dostali $p(X = 1 | x_1^n) = \frac{B(\alpha+1, \beta)}{B(\alpha, \beta)} = \frac{\alpha}{\alpha+\beta}$, čo je priemer posterioru. V prípade rovnomerného prioru by sme dostali známe Laplaceovo pravidlo postupu (angl., *rule of succession*): $p(X = 1 | x_1^n) = \frac{n_1+1}{n+2}$.

Pre Beta(100, 100) prior je pravdepodobnosť, že po sekvencii 20-tich hodov, v ktorých došlo k úspechu 5 krát, nastane ďalší úspech, rovná 0.477 – priemer posterioru. Pripomeňme, že nebayesiánc by túto pravdepodobnosť odhadol ML odhadom, ktorý je 0.25. Pre rovnomerný prior je hľadaná pravdepodobnosť rovná 0.273. \diamond

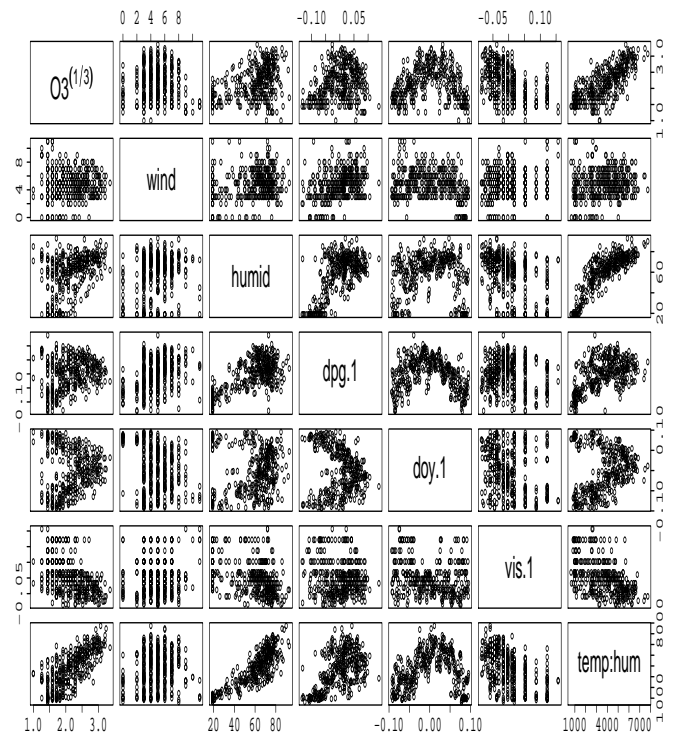
Ako už bolo spomenuté, posteriorna prediktívna pravdepodobnosť sa používa aj na posteriornú prediktívnu kontrolu modelu, na posúdenie, zhodnotenie modelu. Voľne povedané, ak je pozorovaný údaj v strede posteriorného prediktívneho rozdelenia, potom je v súlade s fitom modelu. Ak ale nameraná hodnota leží na chvoste prediktívneho posterioru, potom ju model nevystihuje. Na formálne vyhodnotenie kvality modelu sa používa posteriorna prediktívna p-hodnota, vid' [3].

Príklad (pokr.): Posteriorna prediktívna pravdepodobnosť $p(\tilde{y} \leq 5 | x_1^n)$ toho, že v nasledovnej 20-tici hodov sa bude $X = 1$ vyskytovať 5-krát alebo menej je¹³ 0.0392. To by naznačovalo, že náš model (t.j., prior a dáta-generujúce rozdelenie) nedostatočne vystihuje pozorované dáta, vid' tiež obr. 5. Nakoľko dáta-generujúce rozdelenie je v našom prípade sotva spochybniteľné, mýliť sa môže len v priore. Keďže máme ale dočinenia s malým výberom ($n = 20$) a apriórna informácia je dostatočne silná, nie je v tomto prípade dôvod meniť model. \diamond

Použitie prediktívneho posterioru na posúdenie modelu si zaslúži ešte jednu ilustráciu.

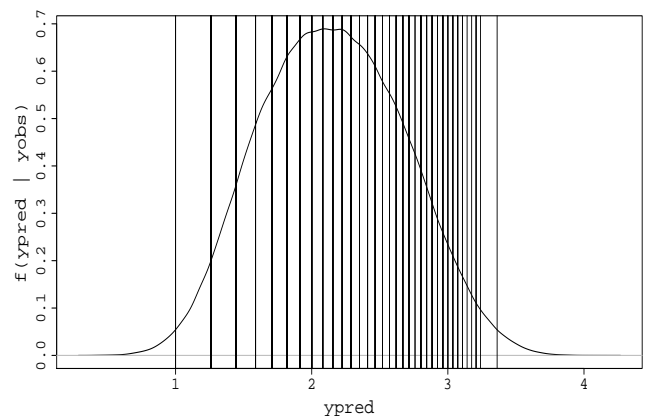
Príklad: V R-kovej knižnici *faraway* sa nachádzajú dáta ozonu obsahujúce merania hladiny ozónu ako aj deviatich ďalších premenných. Vyberieme z nich šesť (wind, humidity, temp, dpq, vis, doy), a s ich pomocou budeme chcieť v rámci regresného modelu modelovať hladinu ozónu O3. Po fáze budovania regresného modelu, skončíme napríklad pri modeli $O3^{1/3} \sim \text{wind} + \text{humidity} + \text{poly}(\text{dpq}, 2) + \text{poly}(\text{doy}, 2) + \text{poly}(\text{vis}, 2) + \text{temp}:\text{humidity}$. Párový graf je na obr. 6.

Na parametre modelu β, σ^2 položíme jeffreysovský prior. Nájdeme posterior: $\beta | \sigma^2, y \sim n(\hat{\beta}, \sigma^2 V_\beta)$ a $\sigma^2 | y \sim \text{Inv-}\chi^2(n - k, s^2)$, kde $\hat{\beta} = (X'X)^{-1}X'y$, $V_\beta = (X'X)^{-1}$, k je počet prediktorov, n počet pozorovaní, s^2 odhad variancie chýb. Hoci je v tomto prípade aj posteriorna prediktívna hustota $p(\tilde{y}_1^n | y_1^n)$ vyjadriteľná analyticky (vid' napr. [3]), stojí za zmienku, že predstava o prediktívnom posteriore $p(\tilde{y}_1^n | y_1^n)$ budúcich dát \tilde{y}_1^n , sa dá získať aj bez rátania integrálu v (2). Stačí ak vieme generovať nové \tilde{y} z rozdelenia $p(\tilde{y}, \theta | y_1^n)$. Odhad hľadaného prediktívneho posterioru sa dá dostať napríklad jadrovým vyhladením vygenerovaných dát \tilde{y} . Nechajme si z posterioru vygenerovať 1000 nových dvojíc (β, σ^2) , a pre každú následne vygenerujeme pri danej matici plánu X , n -tícu nových \tilde{y} . Jadrovo vyhladíme všetkých 1000 n -tíc a do výsledného odhadu prediktívneho posterioru ešte zakreslíme



Obr. 6. Párový graf dát.

pomocou zvislých čiar pozície nameraných y_1^n . Výsledok¹⁴ je na obr. 7. Z obrázka je zrejmé, že model v hrubých rosoch vystihuje dáta, ale zďaleka nie uspokojivo, nakoľko sa väčšina pozorovaných dát nachádza na pravom chvoste prediktívneho posterioru, a nie v strede.



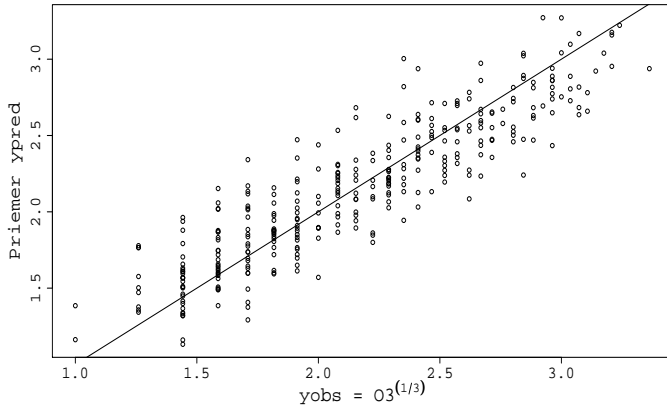
Obr. 7. Prediktívna aposteriorna hustota $p(\tilde{y}_1^n | y_1^n)$ a pozície nameraných y_1^n .

V inej forme je tá istá vada modelu viditeľná na obr. 8, zobrazujúcom graf závislosti priemerov z \tilde{y} -ov nasimulovaných

¹⁴ `parsim = blinreg((oz$O3)^(1/3), as.matrix(Xm), m = 1000); predy = blinregpred(as.matrix(Xm), parsim); plot(density(predy)) for (i in 1:330){abline(v = ((oz$O3)[i])^(1/3))}`

¹³ `sum(pred[1:6])`

z prediktívneho posterioru oproti pozorovaným y_1^n , ako aj referenčnú, 45° čiaru¹⁵.



Obr. 8. Graf priemerov z \tilde{y} -ov nasimulovaných z prediktívneho posterioru oproti pozorovaným y_1^n .

Z oboch obrázkov vyplýva, že by bolo potrebné modifikovať model tak, aby sa stred prediktívneho posterioru posunul doprava. \diamond

I. Priemerovanie modelov

Aby sme mohli popísať ďalší užitočný bayesovský koncept – priemerovanie modelov, budeme musieť opustiť hádzanie mincou. Ideálna na to bude opäť regresia. Predpokladajme, že modelujeme regresným modelom náhodnú premennú y pomocou nejakej množiny prediktorov. Býva zvykom hľadať najlepší podmodel, teda podmnožinu množiny prediktorov, ktorá by v nejakom zmysle najlepšie vystihovala chovanie závislej premennej. Obyčajne sa kvalita posudzuje nejakým kritériom (napr. Akaikeho kritériom, viď [6] alebo [8]), ktoré zaručuje parsimóniu, teda rovnováhu medzi počtom prediktorov a tesnosťou fitu. V bayesovskej štatistike na tento účel slúži Bayesovské informačné kritérium (angl., *Bayesian Information Criterion*, BIC) (viď [6]), ktoré je aproximáciou bayesovského faktora.

Snaha o výber 'naj' modelu je ale problematická, pretože výberom jediného modelu sa ignoruje neistota, ktorú máme o jednotlivých modeloch (angl., *model uncertainty*), a táto neistota veľmi často prevažuje všetky ostatné zdroje neistoty. Bayesovské priemerovanie modelov (angl., *Bayesian Model Averaging*, BMA) poskytuje koherentný spôsob ako zobrať neistotu o modeli do úvahy. Vychádza sa pri tom z apriórneho (zvyčajne rovnomerného) rozdelenia $p(\cdot)$ na množine modelov¹⁶ $M = \{M_1, \dots, M_k\}$. Prior sa následne bayesovsky koriguje dátami X_1^n a obdrží sa posterior

$$p(M_j | x_1^n) = \frac{p(x_1^n | M_j)p(M_j)}{\sum_{l=1}^k p(x_1^n | M_l)p(M_l)}, \quad (4)$$

¹⁵ `plot((oz$O3)^(1/3), colMeans(predy)); abline(0,1)`

¹⁶Množina M môže byť vo všeobecnosti spojitá.

kde $p(x_1^n | M_j)$ je integrovaná vierohodnosť (angl., *integrated likelihood*)

$$p(x_1^n | M_j) = \int p(x_1^n | \theta_j, M_j)p(\theta_j | M_j) d\theta_j, \quad (5)$$

a θ_j je parameter modelu M_j . Posteriórne inferencie a predikcie sú teda založené na množine modelov, a nie na jedinom, akokoľvek optimálnom modeli.

Nech ϕ je parameter, ktorý nás zaujíma. Posteriórna modelovo-spriemerovaná hustota $p(\phi | x_1^n)$ je

$$p(\phi | x_1^n) = \sum_{j=1}^k p(\phi | x_1^n, M_j)p(M_j | x_1^n).$$

No a jej stredná hodnota a variancia sú modelovo-spriemerné posteriórne charakteristiky parametra ϕ .

Príklad: V R-kovej knižnici MASS sa nachádzajú dáta UScrime obsahujúce údaje o 47 amerických štátoch, za rok 1960. Údaje boli zhromaždené za cieľom modelovania kriminality. Prediktorov je 15. Po logaritmickej transformácii prediktorov ako aj závislej premennej použijeme knižnicu BMA¹⁷ (viď [31]) na nájdenie modelovo-spriemerovaných posteriórnych hustôt parametrov modelu¹⁸.

Z nasledovného numerického zhrnutia sa môžeme dozvedieť, že 5 modelov s najvyššou posteriórnou pravdepodobnosťou má kumulatívnu pravdepodobnosť iba 0.3. Stĺpec $p!=0$ obsahuje údaje, pre štyri z prediktorov, o tom, aká je (percentuálna) pravdepodobnosť, že daný prediktor je v modeli. V stĺpci EV sa nachádzajú stredné hodnoty modelovo-spriemernej posteriórnej hustoty, a SD obsahuje smerodajné odchýlky, pre každú premennú. V ďalších stĺpcoch sú bayesovské odhady parametrov, v troch najlepších modeloch.

51 models were selected

Best 5 models have cumulative post. probab. 0.3

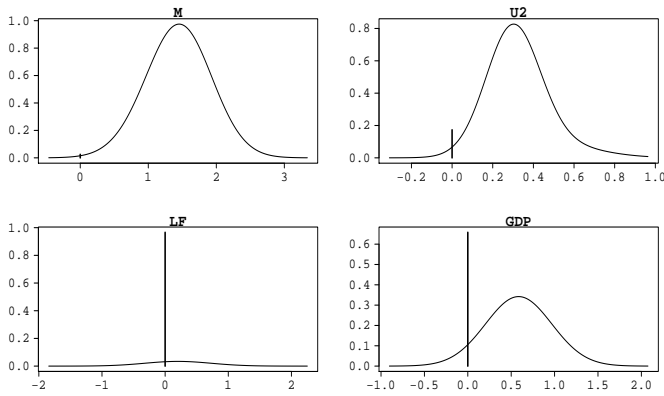
	p!=0	EV	SD	modell1	modell2	modell3
M	97.5	1.4014	0.532	1.478	1.514	1.605
U2	82.7	0.2709	0.194	0.289	0.322	0.274
LF	3.4	0.0069	0.103	.	.	.
GDP	34.2	0.2006	0.357	.	.	0.541

Na obr. 9 sú modelovo-spriemerné posteriórne hustoty zobrazené pre 4 vybrané prediktory. Veľkosť Diracovho impulzu v bode 0 udáva relatívny počet podmodelov (z celkového počtu 51 podmodelov s najvyššou posteriórnou pravdepodobnosťou), v ktorých sa daný prediktor nevyskytoval. Napríklad u prediktora M (percento mužov vo veku 14-24 rokov) je táto pravdepodobnosť (relatívna početnosť) veľmi malá. Krivka zobrazuje modelovo-spriemernú posteriórnú hustotu, cez podmodely v ktorých sa daný prediktor nachádza.

Predstava o tom, ktorý prediktor je ako významný sa dá rýchlo získať aj z BMA koberca (angl., *image plot*). Na obr. 10 sú

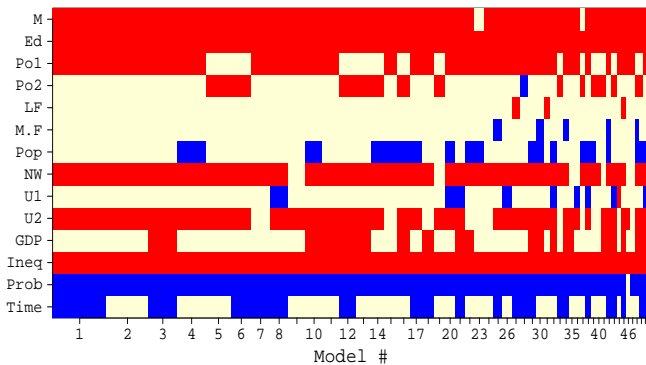
¹⁷V knižnici je implementovaná aproximácia integrovanej vierohodnosti pomocou BICu, čo značne znižuje komplexitu výpočtov. V knižnici BAS je na výpočet použité samplovanie bez navracania, z posteriórnej hustoty.

¹⁸`library(MASS); library(BMA); data(UScrime); xcrime = UScrime[,-16]; xcrime = log(xcrime[,-2]); ycrime = log(UScrime[,16]); breg = bicreg(xcrime, ycrime); summary(breg, digits = 2); plot(breg, mfrow = c(2,2), include = c(1,10,5,11), include.inter = F); imageplot.bma(breg)`



Obr. 9. Modelovo-spriemerný posterior pre štyri parametre modelu.

na x -ovej osi zoradené modely, zostupne, podľa posteriórnej pravdepodobnosti. Šírka binu vyjadruje veľkosť posteriórnej pravdepodobnosti daného modelu. Farebne je indikovaná posteriórna pravdepodobnosť (v %), že prediktor (na osi y) je v modeli. Červená¹⁹ označuje hodnoty vyššie než 90%, modrá hodnoty²⁰ v intervale (80, 90)%.



Obr. 10. BMA koberec

Pre porovnanie²¹, AIC (angl., *Akaike Information Criterion*) by do najlepšieho podmodelu vybralo prediktory M, Ed, Po1, M.F, Pop, NW, U2, GDP, Ineq, Prob, Time. Mnohé z nich sa v bayesovsky spriemerovanom modeli nenachádzajú. Naj-AIC model ani nepatrí medzi päť aposteriórne najpravdepodobnejších modelov. \diamond

Dôležitosť priemerovania modelov si vďaka bayesiáncom uvedomili aj nebayesiánci, viď [6].

J. A ďalšie

1) Inferencie – založené na (1), predikcie a posúdenie modelu – založené na (2), priemerovanie modelov – založené

¹⁹V tlačenej podobe tmavo sivá.

²⁰V tlačenej podobe čierna.

²¹`d = cbind(ycrime, xcrime); m = lm(ycrime ~ ., data = d); summary(m); stepAIC(m)`

na (4) a (5), ako aj ďalšie bayesovské operácie si vyžadujú výpočet integrálov. Bayesiánci sú spolu so štatistickými fyzikmi, od ktorých sa pred pár desaťročiami naučili integrovať mnoho-rozmerné integrály pomocou markovovského monte carlo simulovania (angl., *Markov chain Monte Carlo*, MCMC), známi ako 'agresívni' výpočtári. Hoci sú základné MCMC algoritmy (Metropolisov, Metropolisov-Hastingsov algoritmus, Gibbsov generátor (angl., *Gibbs sampler*)) pomerne ľahko popísateľné (viď napr. [9]), ich praktické zvládnutie si vyžaduje značnú dávku kumštu.

2) Bayesiánci sú majstri v modelovaní pomocou hierarchických modelov. S týmito modelmi úzko súvisí aj zaujímavá herézia, známa ako empirický bayes. Jednoduchým príkladom hierarchického bayesovského modelu je: $X_i | \theta_i \sim n(\theta_i, \sigma^2)$, kde X_i ($i = 1, 2, \dots, n$) sú nezávislé; a $\theta_i \sim n(\mu, \tau^2)$, kde θ_i ($i = 1, 2, \dots, n$) sú zameniteľné. Dá sa ukázať, že marginálne rozdelenie X_i je $n(\mu, \sigma^2 + \tau^2)$. Ak je σ^2 známe, môžeme teda parametre μ , τ^2 prioru (nazývajú sa tiež hyperparametrami) odhadnúť z dát. Vďaka marginálnemu rozdeleniu je tak možné určiť hyperparametre prioru z dát. V takomto empirickom bayesovskom prístupe sú teda dáta použité dva razy: na určenie hyperparametrov prioru a aj na jeho aktualizáciu (angl., *updating*). V hierarchickom bayesovskom modelovaní je podstatný predpoklad zameniteľnosti²² (angl., *exchangeability*) parametrov θ_i . Parametre $\theta_1, \theta_2, \dots, \theta_k$ sú zameniteľné ak ich rozdelenie je invariantné na ich permutáciu. So zameniteľnosťou úzko súvisí aj zásadná de Finettiho veta.

3) Odhliadnuc od fundamentálnej odlišnosti bayesovského prístupu od všetkých nebayesovských prístupov, je možné chápať bayesovanie ako metódu, pomocou ktorej sa dajú (principiálnym spôsobom) regularizovať ML odhady. Napríklad známy hrebeňový odhad (angl., *ridge estimator*) parametrov gaussovského regresného modelu, používaný nebayesiánmi v prípade že matica plánu je takmer singularná, sa dá bayesovsky obdržať pomocou gaussovského prioru, viď napr. [9].

4) V posledných rokoch sa prudko rozvíja neparametrická bayesovská štatistika. Pravdepodobnostné modely parametrizované konečným počtom parametrov sú veľakrát príliš neflexibilné. V neparametrickom bayesovaní (viď napr. [5]) sa kladie prior na množinu všetkých rozdelení.

5) ABC (angl., *Approximate Bayesian Computation*) je zaujímavá technika, pomocou ktorej sa dajú robiť bayesovské úvahy v prípade, že vierohodnostná funkcia je numericky nezvládnuteľná. Na to, aby sa získala vzorka hodnôt z posteriuoru sa v ABC postupuje podľa tohto receptu:

- zvoľte pracovnú (angl., *proposal*) apriórnu hustotu $p(\theta)$,
- vygenerujte kandidáta θ^* z $p(\theta)$,
- nasimulujte dáta Y_1^n z $p(Y_1^n | \theta^*)$,
- vypočítajte vhodnú sumárnu štatistiku $S(y_1^n)$ pre nasimulované dáta, ako aj pre napozorované dáta $S(x_1^n)$,
- akceptujte θ^* v prípade, že vhodná miera vzdialenosti $\delta(\cdot, \cdot)$ spĺňa $\delta(S(x_1^n), S(y_1^n)) < \epsilon$.

ABC sa objavilo v populačnej genetike a rozšírilo sa aj do ďalších disciplín.

²²Zaujímavá diskusia o nezávislosti a zameniteľnosti je v [30].

IV. PĚLE-MÊLE

Pre mnohých nebayesovských štatistikov je bayesovský rámec príliš zväzujúci, rigidný. Z pohľadu bayesovskej štatistiky sa zas nebayesovská štatistika javí ako súbor 'ad hoc' procedúr, 'keby'-uvažovania, a paradoxov. Tento desaťročia sa vyvíjajúci bayesovsko-nebayesovský ping-pong vniesol mnoho poznania do základov štatistiky, aj do praktickej analýzy dát, ale schizma trvá. Viacerí štatistickí hľadajú kompromis, viď napr. nedávnu prácu [29].

Vnútrobayesovským sporom (subjektivistí vs. objektivistí) je venovaná veľká časť tretieho zväzku prvého ročníka časopisu *Bayesian analysis*. V nedávnom čísle (Vol. 3, No. 3, 2008) toho istého časopisu je zas zaujímavá debata, vyprovokovaná Gelmanovým článkom, ktorý zhrňa výhrady voči bayesovskej štatistike.

Thomas Bayes žil v rokoch 1702 až 1761. Jeho slávna 'Esej o riešení jedného problému v rámci doktríny šancí' vyšla v roku 1763. Pomenovanie 'bayesiánci' vyvoláva v mnohých dojem akejsi sekty, viď [32]. Keby sa bayesiánci volali napríklad posterioristi hneď by to znelo menej sektársky!

K posterioristom majú spomedzi tých iných najbližšie asi stúpenci prístupu k štatistike, ktorý je založený na vierohodnostnej funkcii²³. Mimo zásadného rozdielu medzi likelihood-wallahovským 'keby' a posteriórskym 'ked' uvažovaním, sa oba prístupy líšia aj po technickej stránke: posterioristi integrujú, wallahovia maximalizujú.

Dva citáty, ako pripomienka na časy 'studenej vojny'. Maurice Kendall: 'Život by bol oveľa jednoduchší keby bayesiánci nasledovali svojho majstra a publikovali svoje práce posmrtné'. Dennis V. Lindley: 'Vo vnútri každého nebayesiánca je ukrytý bayesiánec, snažiaci sa predrat' von.' Keď sa dvaja bijú, môže zmocniť niekto tretí. Dolovanie v dátach (angl., *data mining*), strojové učenie (angl., *machine learning*) sú prístupy k analýze dát, v ktorých sa preferuje algoritmické získavanie informácie z dát (viď [20]), oproti štatistickému prístupu, ktorý je – nebayesovský, rovnako ako bayesovský – založený na pravdepodobnostnom modeli.

V. POĎAKOVANIE

Ďakujem Martinovi Kalinovi a Petrovi Volaufovi za podnet, prof. Andrejovi Pázmanovi a Viktorovi Witkovskému za diskusie o bayesovaní, Lukášovi Lafférsovi za pripomienky a otázky, ktoré prispeli k vylepšeniu textu.

LITERATÚRA

Učebnice bayesovskej štatistiky

- [1] Albert, J. (2009). *Bayesian computation with R*. New York:Springer-Verlag. 2nd ed.
- [2] Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York:Springer-Verlag.
- [3] Gelman, A., Karlin, B., Stern, H. and Rubin, D. (2004). *Bayesian Data Analysis*. London:Chapman& Hall. 2nd ed.
- [4] Hoff, P. (2009). *A First Course in Bayesian Statistical Methods*. New York:Springer-Verlag.
- [5] Ghosh, J. K., and Ramamoorthi, R. V. (2003). *Bayesian Nonparametrics*. New York:Springer-Verlag.
- [6] Hjort, N. L. and Claeskens, G. (2008). *Model Selection and Model Averaging*. Cambridge:CUP.

- [7] Koop, G. (2003). *Bayesian Econometrics*. Chichester:Wiley.
- [8] Mittelhammer, R., Judge, G. and Miller, D. (2000). *Econometric Foundations*. Cambridge:CUP.
- [9] Pázman, A. (2003). *Bayesovská štatistika*. Bratislava:Univerzita Komenského.
- [10] Roberts, Ch. (2001). *The Bayesian Choice: from Decision-Theoretic Motivations to Computational Implementation*. New York:Springer-Verlag. 2001, 2nd ed.

Výpočtová stránka bayesovskej štatistiky

- [11] Gilks, W., Richardson, S. and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. New York:Chapman & Hall.
- [12] I. Ntzoufras (2009). *Bayesian modeling using WinBUGS*. New Jersey:Wiley.
- [13] Marin, J.-M. and Robert, Ch. P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics*. New York:Springer-Verlag.

Články

- [14] Aldrich, J. (2008). R. A. Fisher on Bayes and Bayes' Theorem. *Bayesian Analysis*. 3/1:161-170.
- [15] Basu, D. (1975). Statistical information and likelihood. *Sankhya*. 37/1:1-71.
- [16] Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Royal Soc.* 53: 370-41. Facsimile available at <http://www.york.ac.uk/depts/maths/histstat/essay.pdf>.
- [17] Berger, J. (1990). Robust Bayesian analysis: sensitivity to the prior. *J. Statist. Plann. Inference*, 25:303-328.
- [18] Berger, J. (2006). The case for objective Bayesian analysis. *Bayesian Analysis*, 1/3:385-402. With discussion.
- [19] Berger, J., Bernardo, J. and Sun, D. (2009). The formal definition of reference priors. *Ann. Stat.*, 37:905-938.
- [20] Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statist. Sci.*, 16/3:199-231.
- [21] Brown, L. D., Cai, T. T. and Dasgupta, A. (2001). Interval Estimation for a Binomial Proportion. *Stat. Sci.*, 16:101-133.
- [22] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *Am. Stat.*, 49:327-225.
- [23] Clide, B. A. (1999). Bayesian model averaging and model search strategies. In *Bayesian Statistics 6*, J. Bernardo et al. (eds.), pp. 157-185.
- [24] Christensen, R. (2005). Testing Fisher, Neyman, Pearson and Bayes, *Am. Stat.*, 59:121-126.
- [25] Gelman, A. (2008). Objections to Bayesian statistics. *Bayesian Analysis* 3/3:445-467. With discussion by J. M. Bernardo, J. B. Kadane, S. Senn, and L. Wasserman.
- [26] Goldstein, M. (2006). Subjective Bayesian analysis: principles and practice. *Bayesian Analysis*, 1/3:403-420. With discussion.
- [27] Kass, R. and Raftery, A. (1995). Bayes factors. *J. Am. Stat. Assoc.*, 90:773-795.
- [28] Kass, R. E. and Wasserman, L. A. (1996). The selection of prior distributions by formal rules. *J. Am. Stat. Assoc.* 91:1343-1370.
- [29] Little, R. (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *Am. Stat.*, 60/3:1-11.
- [30] Pázman, A. (2004). Bayesovská štatistika (Ako alternatíva pre stredné školy). *Obzory mat. fyz. info.* 33/1:1-14.
- [31] Raftery, A. E., Painter, I. S. and Volinsky, Ch. T. (2005). BMA: An R package for Bayesian model averaging. *R news*. 5/2:2-8.
- [32] Robert, Ch. (2008). Misconceptions on Bayesianism. *The ISBA Bulletin*, 15/4:2-3.

²³Basu, v poučnej práci [15] ich označuje slovom 'likelihood-wallah'.