BAYESOVSKÉ ODHADY: PŘIROZENÝ NÁSTROJ PRO VYUŽITÍ APRIORNÍ INFORMACE

BAYESIAN ESTIMATES: TOOL FOR PROCESSING PRIOR INFORMATION

Jan Kalina[1,2], Lubomír Soukup[3]

Adresa: [1]Ústav informatiky AV ČR, Pod Vodárenskou věží 2, 182 07 Praha 8, [2]Univerzita Karlova, Matematicko-fyzikální fakulta, Sokolovská 83, 186 75 Praha 8, [3]Ústav teorie informace a automatizace AV ČR, Pod Vodárenskou věží 4, 182 00 Praha 8

E-mail: kalina@cs.cas.cz

Abstrakt: Tento článek studuje některé základní statistické modely a zamýšlí se nad situací, zda a jak bayesovské odhady jejich parametrů odpovídají intuici v případě, že se kombinují naměřená data s výsledky předchozích měření prováděných za stejných podmínek. Konkrétně se věnujeme bayesovským odhadům parametrů pro normální nebo binomické rozdělení, lineární regresi, ale i regularizačním sítím z oblasti strojového učení.

Klíčová slova: Bayesovské odhady, apriorní informace, předchozí měření, regularizace.

Abstract: This paper considers some fundamental statistical models and investigates whether Bayesian estimates of their parameters correspond to intuition in the situation, when observed data are combined with results of previous (prior) measurements obtained under the same conditions. Particularly, the paper considers Bayesian estimates of parameters for the normal or binomial distributions, linear regression, or regularization networks from the field of machine learning.

Keywords: Bayesian estimation, prior information, previous measurements, regularization.

1. Bayesovské odhady

Bayesovské bodové odhady jsou v souladu s intuicí, a to aspoň v některých poměrně jednoduchých situacích, kdy je třeba zkombinovat dostupná měření (či výsledky experimentů) s výsledky předchozích měření. V učebních textech bayesovské statistiky, a to ani v těch fundovaných jako [12] nebo kap. XVI knihy [1], však není prostor na podrobné rozepsání vzorců a intuitivní interpretování bayesovských odhadů; těmto otázkám se věnuje tento článek.

V příkladech v následujících kapitolách se bayesovský odhad typicky určí jako střední hodnota aposteriorního rozdělení, kterou zde nebudeme přímo odvozovat. Postup pro odvození pro různé situace lze najít v široké nabídce dostupné literatury (např. [18]), a tak se raději zaměříme na interpretaci výsledků v situaci, kdy je předchozích R měření (pro $R \geq 1$) prováděno za stejných podmínek jako stávající (nová) data. Vysvětlíme, zda bayesovské odhady opravdu považují (např. fyzikální) konstanty za náhodné veličiny. Hlubší filozofickou obhajobu bayesovského uvažování lze nalézt např. v knize [14], která prezentuje (bayesovské) pravděpodobnostní uvažování jako přirozenou nadstavbu matematické logiky a zdravého rozumu.

2. Střední hodnota normálního rozdělení

2.1. Jednorozměrný případ

Uvažujeme náhodný výběr $X_1, \ldots, X_n$, kde X_i pochází z normálního rozdělení $X_i \sim \mathsf{N}(\theta, \sigma^2)$ pro $i = 1, \ldots, n$. Může jít např. o výsledky měření rychlosti světla, kterou zde tím pádem označujeme jako θ. Normalita hodnot X_i může být poměrně přirozená[1], pokud se každé jednotlivé měření X_i realizuje jako průměr několika nezávislých měření a pokud data neobsahují odlehlé hodnoty[2]. Jako $\overline{X}$ označíme průměr pozorovaných dat, pro nějž samozřejmě[3] platí $\mathsf{var}\,\overline{X} = \sigma^2/n$.

Nezávisle na naměřených datech mějme k dispozici výsledky předchozích měření. Odhady θ získané z těchto apriorních měření označíme jako $t_1, \ldots, t_R$. Dejme tomu, že každý z nich byl získán také jako realizace normálně rozdělené náhodné veličiny. Označme $\overline{t} = \sum_{r=1}^{R} t_r/R$.

Nyní postupujme velmi pomalu. Znalost $t_1, \ldots, t_R$ nám snižuje nejistotu (neurčitost) o neznámé hodnotě θ. Očekáváme, že hodnota $\overline{t}$ bude blízká skutečné, neznámé hodnotě veličiny θ, neboť považujeme za nedůvěryhodné, aby skutečná hodnota θ byla výrazně odchýlená od experimentálně určeného průměru $\overline{t}$. Neznáme příčinu tohoto odchýlení, proto jej považujeme za náhodné. Věříme, že odchylka určité velikosti je stejně pravděpodobná, ať nastane na jednu nebo na druhou (opačnou) stranu od skutečné hodnoty

[1] Simon Newcomb (1835–1909) provedl v roce 1882 celkem 66 měření rychlosti světla. Data jsou k dispozici v knihovně `BayesDA` softwaru R a historii jejich zpracování shrnul Stigler [20]. Už Newcomb přiřazoval malé váhy „diskordantním“ pozorováním; když se odlehlé hodnoty vynechají, normální rozdělení je vhodným modelem ([21], str. 1070).

[2] Ostatně již Carl Friedrich Gauss (1777–1855), otec teorie chyb měření, modeloval chyby měření pomocí normálního rozdělení [15].

[3] Daný vztah lze zdůvodnit také z hlediska bayesovského přístupu za předpokladu neinformativního apriorního rozdělení.

Mezi více než sedmdesáti účastníky, kterým zelená na slovenském koronavirovém semaforu[1] umožnila účast, mně slovenští kolegové coby jedinému zástupci ČStS přisoudili roli čestného hosta z bratrské krajiny a pozorně naslouchali mému příspěvku na téma *COVID a statistika – co jsme se (ne)naučili*.

Konference byla doplněna příjemným společenským programem, mj. organizovanou prohlídkou historického jádra stredoslovenské metropole zalité podzimním (jesenným) sluncem.

[1] Semafor se rozsvěcuje vždy ve čtvrtek a je „platný“ pro celý týden od následujícího pondělí. Barva semaforu signalizuje míru restrikcí. Barvy jsou uváděny pro jednotlivé okresy. V době zahájení konference, tedy ve čtvrtek 9. září 2021 byl sice Banskobystrický okres stále ještě zelený, ale na další týden už se rozsvěcovala oranžová.

Členové pracovní skupiny FENStatS se na vzniku dokumentu podíleli. FENStatS byl jedním z prvních subjektů podporujících vznik dokumentu, viz `https://www.stifterverband.org/data-literacy-charter`.

4. Standard pro gramotnost v oblasti dat a umělé inteligence

Ve spolupráci s IEEE SA (Institute of Electrical and Electronic Institute' Standard Association, viz `https://standards.ieee.org/`) bude v následujících dvou letech vznikat celosvětový standard pro gramotnost v oblasti dat a umělé inteligence. Standard by měl vytvořit společný operační rámec tvořící základnu pro navrhování cílených politických zásahů, sledování jejich vývoje a empirické vyhodnocení jejich výsledků a tak koordinovat celosvětové úsilí o zvyšování gramotnosti v oblasti dat a umělé inteligence. Standard bude mj. obsahovat základní sadu definic a terminologie.

K dnešnímu dni je podána žádost o autorizaci projektu (očekává se v polovině září) a pracovní skupina IEEE SA (předsedkyně: Katharina Schüller) byla schválena Komisí pro informatiku a AI standardy.

ZPRÁVA Z VÝROČNÍ DVOJKONFERENCE SŠDS
REPORT FROM THE ANNUAL SŠDS CONFERENCE

Ondřej Vencálek

E-mail: `ondrej.vencalek@upol.cz`

Ve dnech 9. a 10. září 2021 se v Banské Bystrici uskutečnila výroční „dvojkonference“ Slovenské štatistické a demografické společnosti (SŠDS), která spojila 20. slovenskou statistickou konferenci a 18. slovenskou demografickou konferenci.

Účastníky této konference přivítala předsedkyně SŠDS Iveta Stankovičová a předseda Štatistického úradu SR Alexander Ballek. Odborný program, viz také: `http://ssds.sk/casopis/konferencie/20SSK_18SDK_program.pdf`, byl rozdělen do čtyř tématických částí – dvě z nich byly věnovány demografii, jedna letošnímu sčítání lidu a jedna statistice, kde šlo především o aplikace statistiky v ekonomii. Všechny prezentace jsou k dispozici na stránce konference, `http://ssds.sk/sk/zoznamkonferencii/74/20ssk/`.

veličiny θ. Zahrnutí pojmů nahodilost, pravděpodobnost do naší úvahy znamená, že nejistota ohledně hodnoty θ může být vyjádřena pojmem náhodná veličina. Tím ovšem netvrdíme, že θ je objektivně náhodná veličina, protože to by byl pouhý sebeklam[4]. Vzhledem ke stejné pravděpodobnosti odchýlení na obě strany můžeme považovat rozdělení příslušné nejistotě ohledně hodnoty θ za symetrické. Zde jej považujme za normální

$$\theta \sim \mathsf{N}(\bar{t}, \gamma^2), \tag{1}$$

kde $\gamma^2 = \mathsf{var}\,\theta$ je vypočteno pomocí vzorce pro rozptyl aritmetického průměru

$$\gamma^2 = \frac{\beta^2}{R}. \tag{2}$$

Parametr β udává míru přesnosti (standardní odchylku), se kterou byly naměřeny hodnoty $t_1, \ldots, t_R$.

Obecně je možné Bayesovu větu (viz např. [1, 12]) interpretovat dvěma způsoby, které jsou matematicky shodně vyjádřeny:

(I) Apriorní rozdělení odpovídá **nejistotě**, kterou máme o odhadovaném parametru.[5]

(II) Odhadovaný parametr je **náhodná** veličina, o jejímž rozdělení nás informuje náhodný výběr $X_1, \ldots, X_n$ (viz [1], od str. 279).

Střední hodnota aposteriorního rozdělení, tedy bayesovský odhad parametru θ, má pak tvar (viz [1], str. 287; [12], str. 20)

$$\begin{aligned} \widehat{\theta} &= \frac{\sigma^{-2}\sum_{i=1}^{n} X_i + \gamma^{-2}\bar{t}}{n\sigma^{-2} + \gamma^{-2}} \\ &= \frac{n\gamma^2\overline{X} + \sigma^2\bar{t}}{n\gamma^2 + \sigma^2}. \end{aligned} \tag{3}$$

Snadno můžeme (3) upravit do tvaru

$$\widehat{\theta} = (1-\delta)\overline{X} + \delta\bar{t}, \quad \text{kde} \quad \delta = \frac{\gamma^{-2}}{n\sigma^{-2} + \gamma^{-2}}. \tag{4}$$

[4] Čtenář se smyslem pro humor může v kontextu rychlosti světla prohlásit, že její nenáhodný charakter je nad slunce jasnější.

[5] V přístupu (I), kterého se držíme, se pracuje s nejistotou, se kterou operoval už Pierre-Simon Laplace (1749–1827), propagátor bayesovského uvažování. Ten jistě nepovažoval fyzikální konstanty za náhodné už jen kvůli tomu, že vnímal svět a vesmír do značné míry deterministicky. Obecně můžeme nejistotu vnímat buď objektivně ve smyslu teorie informace, anebo subjektivně jako stupeň víry (přesvědčení).

Uvažujme speciální situaci, kdy je každá z hodnot $t_1, \ldots, t_R$ (tak jako hodnota $\overline{X}$ spočtená z měření) získána z celkového počtu n měření s rozptylem σ^2. Protože $\mathsf{var}\, t_r = \sigma^2/n$ pro $r = 1, \ldots, R$, platí $\mathsf{var}\, \bar{t} = \sigma^2/(nR)$. Obvykle se v literatuře jen stroze uvádí, že γ lze odhadnout z předchozích měření; zde je nicméně přirozené pro $\mathsf{var}\, \theta$ konkrétně zvolit $\gamma^2 = \sigma^2/(nR)$. Pak získáváme

$$\delta = \frac{R}{R+1} \quad \text{a} \quad \widehat{\theta} = \frac{1}{R+1}\overline{X} + \frac{R}{R+1}\bar{t}. \tag{5}$$

Výsledný odhad je vlastně průměr všech dostupných měření, který se získá zahrnutím předchozích (předběžných) měření mezi ostatní, později naměřené hodnoty. V praxi často bývají předběžná měření méně přesná než aktuální měření. Ovšem i v situaci, kdy nemáme k dispozici předchozí měření $t_1, \ldots, t_R$ a je zafixován poměr $\sigma^2/(n\gamma^2)$, platí výsledek obdobný vztahu (5).

2.2. Mnohorozměrný případ

Uvažujme p-rozměrný náhodný výběr $\mathbf{X}_1, \ldots, \mathbf{X}_n$, kde $\mathbf{X}_i$ pochází z normálního rozdělení $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ pro $i = 1, \ldots, n$. Postupujme obdobně jako v kapitole 2.1, ale již stručněji. Z předchozích měření jsou k dispozici odhady parametru $\boldsymbol{\mu}$, které označíme jako $\mathbf{t}_1, \ldots, \mathbf{t}_R$. Označíme výběrové průměry jako

$$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i \quad \text{a} \quad \bar{\mathbf{t}} = \frac{1}{R}\sum_{r=1}^{R} \mathbf{t}_r. \tag{6}$$

Pokud předběžnou znalost vektorového parametru $\boldsymbol{\mu}$ modelujeme normálním rozdělením $\boldsymbol{\mu} \sim \mathsf{N}_p(\bar{\mathbf{t}}, \boldsymbol{\eta})$, pak je střední hodnota aposteriorního rozdělení vektoru $\boldsymbol{\mu}$ vyjádřena v [5] vztahem

$$\widehat{\boldsymbol{\mu}} = (n\boldsymbol{\Sigma}^{-1} + \boldsymbol{\eta}^{-1})^{-1}(n\boldsymbol{\Sigma}^{-1}\overline{\mathbf{X}} + \boldsymbol{\eta}^{-1}\bar{\mathbf{t}}). \tag{7}$$

Speciálně uvažujme, že $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{\mathcal{I}}$. Předpokládejme, že přesnost všech složek vektoru $\boldsymbol{\mu}$ je stejná a že předběžnou znalost vektorového parametru $\boldsymbol{\mu}$ lze modelovat normálním rozdělením

$$\boldsymbol{\mu} \sim \mathsf{N}_p(\bar{\mathbf{t}}, \mathsf{diag}(\gamma^2, \ldots, \gamma^2)). \tag{8}$$

Střední hodnota (7) aposteriorního rozdělení vektoru $\boldsymbol{\mu}$ pak odpovídá vztahu

$$\widehat{\boldsymbol{\mu}} = \frac{n\gamma^2\overline{\mathbf{X}} + \sigma^2\bar{\mathbf{t}}}{n\gamma^2 + \sigma^2}, \tag{9}$$

ZPRÁVA O ČINNOSTI MEZINÁRODNÍ PRACOVNÍ SKUPINY COVID-19 (VE SPOLUPRÁCI SE SKUPINOU PRO VZDĚLÁVÁNÍ)

COVID-19 INTERNATIONAL WORKING GROUP REPORT (IN COOPERATION WITH LITERACY GROUP)

Katharina Schüller

E-mail: katharina.schueller@stat-up.com

1. Webové stránky

Na stránce FENStatS je nově sekce *Guides & Resources*. Pracovní skupina COVID-19 shromáždila materiály (data, dashboards, reports) vztahující se k pandemii COVID-19 na úrovni národní i mezinárodní. Prostřednictvím nové sekce jsou tyto informace zpřístupněny veřejnosti. Jsou k dispozici v různých podobách (data, zprávy, studie, elektronické interaktivní zdroje) a doplněny o nástroje pro analýzu dostupných dat.

2. Kurz (MOOC) Rozhodování na základě dat v době pandemie (Data-informed Decision Making in a Pandemic)

Pracovní skupina COVID-19 vypracovala návrh interaktivního kurzu (zkr. MOOC) na téma „rozhodování na základě dat v době pandemie“. Kurz by měl pomoci nestatistikům (novinářům, politikům, …) k rozhodování na základě dat. Podkladem pro kurz budou materiály shromážděné pracovní skupinou a vycházející z rámce datové gramotnosti rozvíjeného think tankem Hochschulforum Digitalisierung. Kurz bude sponzorován a bude přístupný na stránkách německé edukační platformy Stifterverband/KI-Campus (AI Campus, viz https://ki-campus.org/), sponzorované německým ministerstvem školství a výzkumu.

3. Zásady datové gramotnosti

Asociace Stifterverband společně s dalšími partnery v lednu 2021 iniciovala sepsání Zásad datové gramotnosti (Data Literacy Charter). V tomto dokumentu se uvádí, co rozumíme datovou gramotností a zdůrazňuje se její význam pro vzdělávací procesy. Dokument je v souladu s datovou strategií německé spolkové vlády a s Berlínskou deklarací o digitální společnosti.

„Uprostřed krize obrovských rozměrů jsme velmi naléhavě potřebovali vysoce kvalitní statistiku, ale namísto toho nám hrozí, že se utopíme v oceánu dat. To je velmi pozoruhodné, protože žijeme v době, kdy jsou data považována za novou ropu, která by měla generovat pokrok a prosperitu. Co přesně chybí ve výzkumu a tvorbě statistik, abychom se dostali z této paradoxní pozice, to vyžaduje další šetření."

Profesor Bourgignon reagoval na iniciativu FENStatS následujícími slovy: „Přednesl jsem… vaši záležitost vědecké radě na našem nedávném říjnovém zasedání a došlo k jednomyslné shodě na důležitosti vysoce kvalitní statistiky a na naléhavé potřebě vyvinout novou metodiku, která by umožňovala interpretaci rostoucího množství dostupných dat." *Za účelem řešení potřeby lepšího financování statistického výzkumu* „Vědecká rada výjimečně přijala úpravu deskriptorů PE1 vytvořením nového deskriptoru (PE1_15) tak, aby konkrétně pokrýval 'Obecnou statistickou metodiku a modelování' (PE1_15 Generic statistical methodology and modelling). Předchozí statistický deskriptor byl přejmenován na 'Matematická statistika' (PE1_14 Mathematical statistics). Tyto změny vstoupí v platnost od prvních výzev v rámci programu Horizont Evropa (Horizon Europe), které budou zahájeny v roce 2021."

Díky těmto změnám se otevřely do budoucna dveře pro financování statistického výzkumu. Nyní je důležité, abychom této příležitosti aktivně využili. Jako prezident FENStatS proto vyzývám všechny kolegy ze všech členských společností, aby se seznámili se strukturami ERC a aktivně využívali tyto nové příležitosti, které se nám nabízejí.

Byl bych velmi rád, kdyby tato výzva padla na úrodnou půdu a silně vzrostl počet výzkumných návrhů a zastoupení statistiků ve výběrových komisích. Počítám také se zpětnou vazbou k národní statistické společnosti nebo přímo k FENStatS v případě potíží nebo překážek. V tomto smyslu bych chtěl požádat národní statistické společnosti organizované ve FENStatS o předání mé zprávy jejich členům.

S pozdravem

Walter J. Radermacher, Ph.D.

Je stále více zřejmé, že bez podpory Evropských grantů se do budoucna neobejdeme. Proto bychom se měli snažit tuto příležitost maximálně využít. Je totiž nebezpečí, že pokud nebude o tyto oblasti dostatečný zájem, ERC v budoucnu tyto deskriptory opět zruší.

Literatura

[1] ERC Evaluation Panels and Keywords, `https://erc.europa.eu/sites/default/files/document/file/ERC_Panel_structure_2020.pdf` *cit. 17*

který můžeme vyjádřit jako

$$\widehat{\boldsymbol{\mu}} = (1-\delta)\overline{\mathbf{X}} + \delta\bar{\mathbf{t}}, \quad \text{kde} \quad \delta = \frac{\gamma^{-2}}{n\sigma^{-2}+\gamma^{-2}} = \frac{\sigma^2}{n\gamma^2+\sigma^2}. \tag{10}$$

Ještě speciálněji předpokládejme, že každá z hodnot $\mathbf{t}_1, \ldots, \mathbf{t}_R$ získaná z předchozích měření pochází z celkového počtu n měření s rozptylem σ^2. Protože $\mathsf{var}\,\overline{\mathbf{X}} = \sigma^2\boldsymbol{\mathcal{I}}/n$, přirozeně volíme $\gamma^2 = \sigma^2/(nR)$. Získáváme pak

$$\delta = \frac{R}{R+1} \quad \text{a} \quad \widehat{\boldsymbol{\mu}} = \frac{1}{R+1}\overline{\mathbf{X}} + \frac{R}{R+1}\bar{\mathbf{t}}. \tag{11}$$

Tuto mnohorozměrnou obdobu (5) lze interpretovat jako odhad $\boldsymbol{\mu}$ smrštěný k hodnotě $\bar{\mathbf{t}}$. O smrštěných (*shrinkage*) odhadech pojednal Stein [19]; obvykle se smrštěný odhad $\boldsymbol{\mu}$ formuluje při vycentrování $\bar{\mathbf{t}} = 0$, kdy (11) odpovídá konvexní lineární kombinaci průměru pozorovaných hodnot a nuly [4].

3. Varianční matice mnohorozměrného normálního rozdělení

Je k dispozici p-rozměrný náhodný vektor $\mathbf{X}_1, \ldots, \mathbf{X}_n$ pocházející z normálního rozdělení $\mathsf{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ se známým $\boldsymbol{\mu}$ a neznámou maticí $\boldsymbol{\Sigma} \in \mathsf{PD}(p)$, kde $\mathsf{PD}(p)$ značí množinu všech pozitivně definitních matic velikosti $p \times p$. Je přirozené odhadovat $\boldsymbol{\Sigma}$ pomocí matice $\mathbf{U}/n$, kde

$$\mathbf{U} = \sum_{i=1}^{n} (\mathbf{X}_i - \boldsymbol{\mu})(\mathbf{X}_i - \boldsymbol{\mu})^T \tag{12}$$

s využitím známého $\boldsymbol{\mu}$. V dané situaci se matice $\mathbf{U}$ řídí Wishartovým rozdělením $W_p(\boldsymbol{\Sigma}, n)$ a tedy platí $E\,\mathbf{U}/n = \Sigma$. Je obvyklé modelovat nejistotu o $\boldsymbol{\Sigma}$ (vlastně $\boldsymbol{\Sigma}^{-1}$) pomocí Wishartova rozdělení $\boldsymbol{\Sigma}^{-1} \sim W_p(\nu + p - 1, \boldsymbol{\Omega}^{-1})$ pro určité $\nu > 0$ a určitou $\boldsymbol{\Omega} \in \mathsf{PD}(p)$ [5]. V takové situaci má $\boldsymbol{\Sigma}$ inverzní Wishartovo rozdělení $W_p^{-1}(\nu + p - 1, \boldsymbol{\Omega})$. Přitom platí

$$\mathsf{E}\boldsymbol{\Sigma}^{-1} = (\nu + p - 1)\boldsymbol{\Omega}^{-1} \quad \text{a ovšem} \quad \mathsf{E}\boldsymbol{\Sigma} = \boldsymbol{\Omega}/(\nu - 2). \tag{13}$$

Aposteriorní střední hodnota $\boldsymbol{\Sigma}$ je pak rovna

$$\widehat{\boldsymbol{\Sigma}} = \frac{\mathbf{U} + \boldsymbol{\Omega}}{n + \nu - 2}. \tag{14}$$

Bayesovské riziko empirických bayesovských odhadů $\boldsymbol{\Sigma}$, které odhadují hyperparametry z dat, v dané situaci studoval článek [7].

Zde ovšem uvažujme, že máme z předchozích měření spočítané odhady $\mathbf{\Sigma}$ ve tvaru $\mathbf{U}_1/n, \ldots, \mathbf{U}_R/n$. Jinými slovy představují matice $\mathbf{U}_1, \ldots, \mathbf{U}_R$ apriorní protějšky $\mathbf{U}$. Dejme tomu, že každý z předchozích experimentů proběhl přes n měření, a to se stejnou variabilitou jako současná měření. Je přirozené volit $\mathbf{\Omega}$ a ν tak, aby

$$\frac{1}{R}\sum_{r=1}^{R}\frac{U_r}{n} = \frac{\mathbf{\Omega}}{\nu - 2}, \tag{15}$$

tedy vzít $\mathbf{\Omega} = \sum_{r=1}^{R}\mathbf{U}_r$ a $\nu = Rn + 2$. Odhad (14) pak získáme jako intuitivní kombinaci výsledku měření s výsledky předchozích měření ve tvaru

$$\widehat{\mathbf{\Sigma}} = \frac{1}{n(R+1)}\left(\mathbf{U} + \sum_{r=1}^{R}\mathbf{U}_r\right). \tag{16}$$

4. Binomické rozdělení

Při výrobě nějakého výrobku v dané výrobní lince bylo zjištěno, že z celkového počtu n výrobků jich bylo m vadných. Nahlížíme na m jako na realizaci náhodné veličiny M, která se řídí binomickým rozdělením $Bi(n, p)$. Úkolem je odhadnout neznámou pravděpodobnost vadného výrobku p. Při bayesovském odhadu p je obvyklé modelovat nejistotu o p tak, že se za apriorní rozdělení pro p zvolí beta rozdělení $p \sim \mathsf{beta}(a, b)$ s konkrétními parametry $a > 0$ a $b > 0$; pak platí $\mathsf{E}\,p = a/(a+b)$. Tento model studoval Anděl ([1], str. 279) a odvodil střední hodnotu aposteriorního rozdělení jako

$$\mathsf{E}(p|m) = \frac{a+m}{a+b+n}. \tag{17}$$

Uvažujme nyní, že v dané situaci máme k dispozici R předchozích navzájem nezávislých kontrol kvality při výrobě daného výrobku. V r-tém z nich bylo zjištěno, že z celkového počtu n_r výrobků jich bylo m_r vadných.

Pak $\sum_{r=1}^{R} m_r$ představuje realizaci náhodné veličiny, která se řídí rozdělením $Bi(\sum_{r=1}^{R} n_r, p)$. Označíme $\overline{p} = \sum_{r=1}^{R} m_r/(\sum_{r=1}^{R} n_r)$. Pro apriorní beta rozdělení se nabízí jako přirozené zvolit parametry apriorního rozdělení tak, že

$$p \sim \mathsf{beta}\left(\sum_{r=1}^{R} m_r, \sum_{r=1}^{R}(n_r - m_r)\right), \tag{18}$$

protože s touto volbou je středního hodnota a rozptyl apriorního rozdělení

$$\mathsf{E}p = \overline{p} \quad \text{a} \quad \mathsf{var}\,p = \frac{(\sum_{r=1}^{R} m_r)(\sum_{r=1}^{R}(n_r - m_r))}{(\sum_{r=1}^{R} n_r)^2(1 + \sum_{r=1}^{R} n_r)}; \tag{19}$$

ZLEPŠENÍ MOŽNOSTÍ FINANCOVÁNÍ STATISTICKÉHO VÝZKUMU ZE STRANY EVROPSKÉ RADY PRO VÝZKUM

NEW EFFORT TO IMPROVE THE FUNDING CONDITIONS FOR STATISTICAL RESEARCH BY THE EUROPEAN COUNCIL

Gejza Dohnal

Adresa: FS ČVUT v Praze, Karlovo náměstí 13, 121 35, Praha 2

E-mail: gejza.dohnal@fs.cvut.cz

Hlavní úlohu ve financování výzkumných projektů v rámci Evropské unie hraje Evropská rada pro výzkum (ERC). Je to evropská grantová agentura, která pro plánování a hodnocení návrhů výzkumných grantů používá stejně jako například naše Grantová agentura tzv. panely. ERC používá systém 25 panelů, které by měly pokrývat všechny oblasti vědy, inženýrství a vzdělávání, rozdělené do třech skupin: sociální a humanitní vědy (6 panelů, SH1 – SH6), fyzikální vědy a inženýrství (10 panelů, PE1 – PE10) a přírodní vědy (9 panelů, LS1 – LS9). Každý z těchto panelů pokrývá několik oblastí vědy a výzkumu, označené tzv. deskriptory. Tento systém se neustále obnovuje a upravuje. Poslední verze systému deskriptorů ERC je z roku 2020 [1]. Oblast pravděpodobnosti a statistiky je v ní pokryta v panelu PE1 dvěma deskriptory: PE1_13 (Probability) a PE1_14 (Statistics).[1]

Federace Evropských národních statistických společností (FENStatS), jejímž členem je i naše společnost, vyvinula nové úsilí s cílem zlepšit podmínky financování statistického výzkumu ze strany Evropské rady. Korespondence mezi prezidentem FENStatS a předsedou ERC dává naději na zlepšení možností financování ze strany ERC. Prezident FENStatS Walter Radermacher rozeslal 19. února 2021 na adresy národních statistických společností následující dopis:

Vážení členové FENStatS, milí kolegové,

na našem Valném shromáždění loni v září jsem vás informoval, že jsem vynaložil nové úsilí na zlepšení podmínek financování statistického výzkumu ze strany Evropské rady pro výzkum. Ve svém dopise předsedovi ERC, profesorovi Jean-Pierre Bourgignonovi, ze dne 3. září 2020, jsem poukázal na to, že pandemická krize znovu zdůraznila význam statistiky a statistického výzkumu pro moderní společnosti:

[1] To je poněkud lepší situace než v GAČR, kde jsou oba tyto obory pokryty jediným širokým hodnotícím panelem P201 Matematika a informatika.

NOVINKY Z FEDERACE EVROPSKÝCH NÁRODNÍCH STATISTICKÝCH SPOLEČNOSTÍ FENSTATS
NEWS FROM THE FEDERATION OF EUROPEAN NATIONAL STATISTICAL SOCIETIES

Ondřej Vencálek

E-mail: ondrej.vencalek@upol.cz

Dne 13. září 2021 se prostřednictvím videokonference konala schůzka zástupců národních statistických společností sdružených ve *Federaci evropských národních statistických společností (FENStatS)*. Této schůzky jsem se jako zástupce České statistické společnosti zúčastnil a nyní bych rád pro členy ČStS shrnul hlavní body jednání týkající se činnosti FENStatS.

- Pracovní skupiny pro vzdělávání a pro pandemii COVID-19. Novinky ohledně činnosti pracovní skupiny jsou shrnuty v příspěvku její vedoucí Kathariny Schüller v tomto čísle Informačního bulletinu.
- Evropská statistická akreditace. O začínajícím projektu Evropské statistické akreditace jsme informovali v Informačním bulletinu 3/2020, viz též webové stránky https://www.fenstats.eu/accreditation. Předseda akreditační komise Magnus Pettersson informoval o současném stavu. Dosud se k akreditačnímu systému připojilo sedm členských asociací a bylo registrováno 42 žádostí o akreditaci, schváleny byly zatím tři akreditace.
- Evropská rada pro výzkum (ERC). Prezident FENStatS Walter Radermacher informoval o své komunikaci s (nově zvolenými) představiteli Evropské rady pro výzkum. Jeho úsilí, o němž informujeme v tomto Inf. bulletinu a informovali jsme na webu, viz http://www.statspol.cz/wp-content/uploads/2021/04/FENStatS-ERC-info.pdf, má za hlavní cíl zlepšit podmínky financování statistického výzkumu ze strany Evropské rady.
- Web a sociální sítě. Kromě webové stránky https://www.fenstats.eu/homepage se FENStatS nyní prezentuje také na sociálních sítích Twitter https://twitter.com/statfen?lang=cs a LinkedIn https://www.linkedin.com/company/fenstats/.
- Stíhání Andrease Georgiou, bývalého ředitele Řeckého statistického úřadu ELSTAT. Prezident FENStatS Walter Radermacher informoval o novém vývoji soudních procesů vedených proti Andreasi Georgiou (https://en.wikipedia.org/wiki/Andreas_Georgiou), jehož se představitelé FENStatS opakovaně veřejně zastávají.

hodnota $\mathsf{var}\, p$ je tedy blízká hodnotě $\mathsf{var}\, \overline{p} = p(1-p)/(\sum_{r=1}^{R} n_r)$. Nyní je aposteriorní střední hodnota p daná vzorcem (17) rovna

$$\widehat{p} = \frac{\sum_{r=1}^{R} m_r + m}{\sum_{r=1}^{R} n_r + n}, \tag{20}$$

tedy počtu vadných výrobků vydělenému celkovým počtem měření. Jinými slovy jde o zcela přirozený odhad pravděpodobnosti zdaru, kdy kombinujeme předchozí měření s nově naměřenou informací.

5. Lineární regrese

Uvažujeme klasický lineární regresní model

$$Y_i = \mathbf{X}_i^T \boldsymbol{\beta} + e_i, \quad i = 1, \ldots, n, \tag{21}$$

kde pozorujeme spojitou odezvu i p-rozměrné regresory (pro $p \geq 1$) pro celkový počet n pozorování. Maticově vyjádříme (21) jako $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$. Předpokládáme nezávislé normálně rozdělené chyby $e_i \sim \mathsf{N}(0, \sigma^2)$ pro $i = 1, \ldots, n$ s danou směrodatnou odchylkou $\sigma > 0$. Pak se i odhad parametru $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ metodou nejmenších čtverců řídí normálním rozdělením, konkrétně

$$\mathbf{b}_{LS} = (b_1^{LS}, \ldots, b_p^{LS})^T \sim \mathsf{N}_p(\boldsymbol{\beta}, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}). \tag{22}$$

Mějme k dispozici R předchozích sad měření, pro něž pomocí metody nejmenších čtverců získáme odhady $\boldsymbol{\beta}$ označené jako $\mathbf{b}_1, \ldots, \mathbf{b}_R$.

5.1. Hřebenová regrese

Uvažujme nejprve, že nejistotu ohledně $\boldsymbol{\beta}$ modelujeme pomocí apriorního rozdělení pro $\boldsymbol{\beta}$ za podmínky známé hodnoty σ^2 ve tvaru $\boldsymbol{\beta} \sim \mathsf{N}(\mathbf{0}, \sigma^2\boldsymbol{\Psi}^{-1})$ s danou maticí $\boldsymbol{\Psi} \in \mathsf{PD}(p)$. Nulovost střední hodnoty zjednoduší následující výsledky; můžeme bez újmy na obecnosti předpokládat, že data upravíme tak, aby průměr hodnot $\mathbf{b}_1, \ldots, \mathbf{b}_R$ byl nulový. Střední hodnota aposteriorního rozdělení pak dá odhad $\boldsymbol{\beta}$ ve tvaru (viz např. [9])

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X} + \boldsymbol{\Psi}\right)^{-1} \mathbf{X}^T\mathbf{Y}, \tag{23}$$

což je odhad známý jako hřebenová regrese (*ridge regression*), resp. hřebenovou regresí se ovšem často myslí speciální případ při $\boldsymbol{\Psi} = \lambda\boldsymbol{\mathcal{I}}$ s nějakým $\lambda > 0$.

Speciálně předpokládejme, že R předchozích měření proběhlo ve stejných podmínkách, v nichž byla získána nová data. To znamená předpokládat, že každé z $\mathbf{b}_r$ pro $r = 1, \ldots, R$ bylo získáno v lineárním modelu se stejnými regresory i se stejným počtem pozorování jako (21). Nyní je přirozené zvolit $\boldsymbol{\Psi}^{-1} = (\mathbf{X}^T\mathbf{X})^{-1}/R$ a tudíž $\widehat{\boldsymbol{\beta}}$ (23) má tvar

$$\widehat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{X} + R\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} = \frac{1}{R+1}\mathbf{b}_{LS}, \tag{24}$$

který odpovídá intuici, protože představuje jako v (11) konvexní lineární kombinaci klasického odhadu a nuly.

5.2. Lasso

Apriorní rozdělení pro β_j pro $j = 1, \ldots, p$ za podmínky známé hodnoty σ^2 nyní uvažujme jako Laplaceovo se střední hodnotou 0 a parametrem měřítka $\tau > 0$, tedy s rozptylem $\mathsf{var}\,\beta_j = 2/\tau^2$. Apriorní rozdělení pro celý vektor $\boldsymbol{\beta}$ přitom bereme jako součin jednotlivých apriorních rozdělení pro $\beta_1, \ldots, \beta_p$. Speciálně ještě předpokládejme $\mathbf{X}^T\mathbf{X} = \boldsymbol{\mathcal{I}}$. Směřujeme k tomu, abychom získali lasso odhad vektoru $\boldsymbol{\beta}$ jako bayesovský odhad. Není ovšem pravda (ani za daných specifických předpokladů), že by šlo získat lasso odhad jako střední hodnotu aposteriorního rozdělení. Je ovšem známo [22, 8], že modus aposteriorního rozdělení je dán jako řešení úlohy

$$\min_{\mathbf{b}\in\mathbb{R}^p}\left\{||\mathbf{Y} - \mathbf{X}\mathbf{b}||_2^2 + 2\tau\sigma^2\sum_{j=1}^{p}|b_i|\right\}; \tag{25}$$

jde o lasso odhad, v němž penalizační parametr lze označit jako $\lambda = 2\tau\sigma^2$.

Nechť je jako v kap. 5.1 k dispozici R předchozích měření se stejnými regresory a stejným počtem pozorování jako v (21). Máme tedy odhady $\boldsymbol{\beta}$ označené jako $\mathbf{b}_1, \ldots, \mathbf{b}_R$. Předpokládejme, že jejich průměr $\overline{\mathbf{b}}$ je roven nule. Samozřejmě platí $\mathsf{var}\,\overline{\mathbf{b}} = \sigma^2\boldsymbol{\mathcal{I}}/R$. Uvažujme nyní analogicky jako v předchozích kapitolách. Pokud položíme $\mathsf{var}\,\boldsymbol{\beta} = \mathsf{var}\,\overline{\mathbf{b}}$, tzn. pokud požadujeme $2/\tau^2 = \sigma^2/R$, pak dostáváme

$$\tau = \sqrt{2R}/\sigma \quad \text{a tedy} \quad \lambda = 2\sigma\sqrt{2R}. \tag{26}$$

[15] Kalina J., Soukup L. (2019): Průkopníci statistiky ve vědách o člověku v 19. století. *Informační bulletin České statistické společnosti* **30**(3), 1–15. *cit. 4*

[16] Neruda R., Vidnerová P. (2009): Learning errors by radial basis function neural networks and regularization networks. *International Journal of Grid and Distributed Computing* **1**, 49–57. *cit. 11*

[17] Pourahmadi M. (2013): *High-dimensional covariance estimation.* Wiley, Hoboken. *cit. 12*

[18] Robert C. P. (2001): *The Bayesian choice.* 2. vydání. Springer, New York. *cit. 4*

[19] Stein C. (1956): Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* **1**, University of California Press, Berkeley, 197–206. *cit. 7*

[20] Stigler S. M. (1973): Simon Newcomb, Percy Daniell, and the history of robust estimation 1885–1920. *Journal of the American Statistical Association* **68**, 872–879. *cit. 4*

[21] Stigler S. M. (1977): Do robust estimators work with *real* data? *Annals of Statistics* **5**, 1055–1098. *cit. 4*

[22] Tibshirani R. (1996): Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society B* **58**, 267–288. *cit. 10*

[23] van Wieringen W. N. (2020): Lecture notes on ridge regression. arXiv: 1509.09169. *cit. 11*

Literatura

[1] Anděl J. (1978): *Matematická statistika.* SNTL/ALFA, Praha. *cit. 3, 5 a 8*

[2] Calvetti D., Somersalo, E. (2018): Inverse problems: From regularization to Bayesian inference. *WIREs Computational Statistics* **10**, e1427. *cit. 12*

[3] Che M. H., Ibrahim J. G. (2003): Conjugate priors for generalized linear models. *Statistica Sinica* **13**, 461 – 476. *cit. 13*

[4] Efron B., Morris C. (1973): Stein's estimation rule and its competitors—An empirical Bayes approach. *Journal of the American Statistical Association* **68**, 117 – 130. *cit. 7*

[5] Evans I. G. (1965): Bayesian estimation of parameters of a multivariate normal distribution. *Journal of the Royal Statistical Society B* **27**, 279 – 283. *cit. 6, 7*

[6] Gauss C. F. (1809): Theoria motus corporum coelestium in sectionibus conicis solem ambientum. Perthes & Besser, Hamburg. *cit. 13*

[7] Haff L. R. (1980): Empirical Bayes estimation of the multivariate normal covariance matrix. *Annals of Statistics* **8**, 586 – 597. *cit. 7*

[8] Hans C. (2009: Bayesian lasso regression. *Biometrika* **96**, 835 – 845. *cit. 10*

[9] Hastie T., Tibshirani R., Friedman J. (2001): *The elements of statistical learning.* Springer, New York. *cit. 9, 11*

[10] Hastie T., Tibshirani R., Wainwright M. (2015): *Statistical learning with sparsity: The lasso and generalizations.* CRC Press, Boca Raton. *cit. 12*

[11] Haykin S. O. (2009): *Neural networks and learning machines: A comprehensive foundation.* 2. vydání. Prentice hall, Upper Saddle River. *cit. 11*

[12] Hušková M. (1985): *Bayesovské metody.* Univerzita Karlova, Praha. *cit. 3, 5*

[13] Hwang J. T. G., Liu P. (2007): Optimal tests shrinking both means and variances applicable to microarray data analysis. *Statistical Applications in Genetics and Molecular Biology* **9**, Article 36. *cit. 12*

[14] Jaynes E. T. (2003): *Probability theory. The logic of science.* Cambridge University Press, Cambridge. *cit. 4*

Lasso odhad daný (25), který označíme jako $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_p)^T$, lze vyjádřit explicitně (viz např. [23]). Za našich předpokladů vyjádříme složky (25) jako

$$\widehat{\beta}_j = \mathsf{sgn}(b_j^{LS}) \left(|b_j^{LS}| - \frac{\lambda}{2} \right)_+ = \mathsf{sgn}(b_j^{LS}) \max\left(0, |b_j^{LS}| - \frac{\lambda}{2}\right) =$$
$$= b_j^{LS} \max\left(0, 1 - \frac{\lambda}{2|b_j^{LS}|}\right), \quad j = 1, \ldots, p,$$

kde $(x)_+$ označuje kladnou část $x \in \mathbb{R}$. Získaný bayesovský odhad $\boldsymbol{\beta}$ zde závisí na hodnotě σ, což je rozdíl oproti hřebenové regresi; každopádně lze σ odhadnout z dat. Ani v našem speciálním případě ale (27) není lineární kombinací pozorovaných dat a apriorní informace, ale smrštění odhadu získaného z pozorování směrem k výsledkům předchozích měření. Přitom pokud pro některé $j = 1, \ldots, p$ platí

$$|b_j^{LS}| \leq \sigma\sqrt{2R}, \tag{27}$$

pak už nutně $\widehat{\beta}_j = 0$.

6. Regularizační sítě

Regularizační sítě (*regularization networks*) představují jednu z metod strojového učení, která je zajímavá tím, že ji lze odvodit v kontextu bayesovské statistiky [16]. Zdůrazněme ovšem, že nemáme na mysli regularizované sítě (*regularized networks*) [11], tedy regularizované verze obvyklých typů neuronových sítí.

Máme k dispozici n hodnot spojité odezvy $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$ a k tomu n hodnot p-rozměrných regresorů $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Cílem je odhadnout regresní funkci

$$f(\mathbf{x}) = \mathsf{E}(Y|\mathbf{X} = \mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^p, \tag{28}$$

na základě dat, přičemž tvar funkce f je neznámý a jen předpokládáme, že f existuje. Jednou možností je odhadnout f jako řešení optimalizační úlohy

$$\min_{f \in H_K} \left\{ \sum_{i=1}^{n} (Y_i - f(\mathbf{X}_i))^2 + \lambda ||f||_{H_K} \right\}, \tag{29}$$

s regularizačním parametrem $\lambda > 0$, kde f hledáme v prostoru s reprodukčním jádrem (RKHS, *reproducing kernel Hilbert spaces*) H_K, příslušném Hilbertově prostoru reálných funkcí na $\mathbb{R}^p$ [9]. V následujícím postupu se typicky

volí K jako Gaussovské jádro

$$K(\mathbf{x},\mathbf{y}) = \exp\left\{-\frac{||\mathbf{x}-\mathbf{y}||^2}{2\sigma^2}\right\}, \quad \mathbf{x},\mathbf{y}\in\mathbb{R}^p \tag{30}$$

pro pevné $\sigma > 0$, které lze odhadnout z dat. Uvažovanou úlohu (29) lze vyjádřit jako

$$\min_{\alpha}\left\{||\mathbf{Y}-\mathbf{K}\boldsymbol{\alpha}||^2 + \lambda\boldsymbol{\alpha}^T\mathbf{K}\boldsymbol{\alpha}\right\}, \tag{31}$$

kde $\boldsymbol{\alpha} = (\alpha_1,\ldots,\alpha_n)^T$ je vektor parametrů a $\mathbf{K}$ je symetrická matice s hodnotami $K_{ij} = K(\mathbf{X}_i,\mathbf{X}_j)$ pro $i,j = 1,\ldots,n$. Derivacemi zjistíme, že minimum nastává pro vektor

$$\begin{aligned}\widehat{\boldsymbol{\alpha}} &= \left(\mathbf{K}^T\mathbf{K}+\lambda\mathbf{K}\right)^{-1}\mathbf{K}^T\mathbf{Y}\\ &= \left[(\mathbf{K}+\lambda\boldsymbol{\mathcal{I}})\mathbf{K}\right]^{-1}\mathbf{K}^T\mathbf{Y}\\ &= \left(\mathbf{K}+\lambda\boldsymbol{\mathcal{I}}\right)^{-1}\mathbf{Y},\end{aligned} \tag{32}$$

který zřejmě odpovídá odhadu metodou hřebenové regrese v lineárním modelu $\mathbf{Y} = \mathbf{K}\boldsymbol{\alpha}+\mathbf{e}$; proto se v literatuře také nazývá *generalized ridge estimator*. Vyhlazená hodnota pro nové pozorování $\mathbf{Z}\in\mathbb{R}^p$ se pak získá jako

$$\widehat{f}(\mathbf{Z}) = \sum_{i=1}^{n}\widehat{\alpha}_i K(\mathbf{Z},\mathbf{X}_i). \tag{33}$$

Odhad regresních parametrů (32) lze získat bayesovským postupem podle kap. 5.1, pokud volíme apriorní rozdělení $\boldsymbol{\alpha}\sim\mathsf{N}\left(\mathbf{0},\sigma^2(\lambda\mathbf{K})^{-1}\right)$, které ovšem na rozdíl od volby v (24) není podle našeho názoru intuitivní.

7. Závěr

Vzhledem k tomu, jak narůstá objem dostupných dat v různých oborech lidské činnosti, narůstá i možnost (a někdy i potřeba) kombinovat naměřená data s výsledky předchozích měření. Čím dál častěji budeme mít k dispozici apriorní znalosti nebo představy o neznámých parametrech, které se mají odhadnout. Takto se v mnoha aplikačních oblastech rozšiřuje prostor pro možné uplatnění bayesovských metod.

Jako příklad lze uvést regularizované postupy pro vysoce dimenzionální data (tj. pro klasifikaci, regresi [10] či testování hypotéz [13]), které lze často odvodit v bayesovském kontextu [2, 17]. Ve strojovém učení má velký význam

bayesovské učení, které v posledních letech výrazně přispělo k pokrokům v robotice, a dále bayesovská optimalizace nebo bayesovské sítě; samotné bodové odhady pomocí neuronových sítí ale obvykle nelze uvažovat v bayesovském kontextu, protože většina odhadů v úlohách strojového učení ani není založena na věrohodnostní funkci.

Článek ilustruje na vybraných modelech, že bayesovské odhady umožňují přirozeným a intuitivním postupem provést syntézu naměřené a apriorní informace. Pokud je apriorní informace získaná z měření za stejných podmínek, za jakých jsou získána aktuální pozorovaná data, pak se konkrétní volba hyperparametrů nabízí přirozeným (odpovídajícím) způsobem. V některých z popsaných příkladů jsme také schopni uhodnout, jak by měl výsledný odhad vypadat. Všechny výsledky v tomto článku ve skutečnosti platí i pro $R = 0$.

Přestože je náš pohled díky specifickým předpokladům na apriorní měření značně zúžený, šlo by využít některé uvedené příklady např. při výuce statistiky, protože učební texty obvykle prezentují co nejobecnější verze daných postupů. Zároveň je třeba přiznat, že i v jiných poměrně základních úlohách může být bayesovské odhadování dost složité. To je třeba příklad logistické regrese, pro niž byl teprve v práci [3] navržen konjugovaný systém hustot[6].

Zatímco bayesovská statistika je na Wikipedii[7] podivně představena jako *„větev relativně moderní statistiky“*, můžeme závěrem říci, že bayesovské uvažování má nejen dlouhou a velmi zajímavou tradici[8], ale zároveň také slibnou budoucnost.

Poděkování

Článek vznikl s podporou grantů GA19-05704S a GA21-19311S Grantové agentury České republiky. Autoři děkují Petře Vidnerové (ÚI AV ČR) za diskusi ke kap. 6.

[6] Konjugovaným systémem se rozumí takový systém pravděpodobnostních rozdělení, kdy při volbě libovolné apriorní hustoty z daného systému do něj patří i aposteriorní hustota.

[7] `https://cs.wikipedia.org/wiki/Bayesovská_statistika`

[8] Zatímco je dobře známo, že Bayesovu větu publikovanou roku 1763 znovuobjevil v roce 1774 a následně zpopularizoval Pierre-Simon Laplace, zůstává obvykle opomíjeno, že bayesovsky uvažoval i Carl Friedrich Gauss. Právě bayesovskou úvahou ve svém významném díle [6] definoval metodu maximální věrohodnosti i získal normální rozdělení jako to, pro nějž maximálně věrohodný odhad střední hodnoty odpovídá aritmetickému průměru.