
SROVNÁNÍ KLASICKÉ A BAYESOVSKÉ PRAVDĚPODOBNOСТИ A STATISTIKY (2.)

Petr Hebák*

(Pokračování z čísla 1/2012 AOP.)

4. Bodové odhady

Klasická statistika často vychází z předpokladu, že dostupná data představují náhodný výběr z nezávislých stejně rozdělených náhodných veličin. *DeGroot* (1988) vysvětluje, proč tento předpoklad je logicky neudržitelný. *Samaniego* (2010, s. 15) se ptá: Je taková data možné získat? Mohou být podmínky při opakovaném házení mincí opravdu stejné? A sám si odpovídá NE, protože zcela obecně jsou identické pokusy fyzikálně nemožné. Nebylo by to tak, že kdyby byly podmínky pokusu opravdu stále zcela stejné, že by výsledky těchto pokusů musely být také stejné? Nicméně (*Samaniego* pokračuje) zkušenosti ukazují, že předpoklad nezávislých stejně rozdělených veličin (tzv. *i.i.d. variables*) je skvělá aproximace reality, takže v souvislosti se statistickými úvahami je učinění tohoto předpokladu relativně neškodné. Pro porovnání klasického a bayesovského přístupu k úsudkům nejde o ústřední spornou otázku a hlavní zdroj nedorozumění, protože obě porovnávané statistické školy tento předpoklad občas učiní nebo jej mlčky předpokládají, když popisují dostupná data.

Klasický přístup k bodovým odhadům

Teorie bodových odhadů v klasickém pojetí je velice propracovaná a není (ani v české literatuře) obtížné najít kvalitní zpracování této problematiky. Za všechny jmenuji jen ty publikace, které jsem četl a považuji je za typické pro tento příspěvek a uvádím je v pořadí podle data vydání. Za česky psané aspoň *Hátle – Likeš* (1972), *Anděl* (1978), *Rao* (1978 překlad se stručnými zmínkami o bayesovském výpočtu) a *Zvára – Štěpán* (1997). Za desítky starších anglicky psaných textů (znovu připomínám, že monografie před rokem 1990 bylo pro mne velice obtížné sehnat) tedy jen *Hogg – Craig* (1970) a *Dudewicz – Mishra* (1988) a z těch porovnávajících klasickou a bayesovskou statistiku aspoň už zmínění *Leamer* (1978), *Wang* (1993), *Bolstad* (2004), dále dva články *Bayarri – Berger* (2004, *Statistical Science*, s. 58 až 80) a nakonec tři z nejnovějších *Samaniego* (2010), *Jackman* (2010) či *Efron* (2010).

* Univerzita Hradec Králové, Fakulta informatiky a managementu; Vysoká škola ekonomická v Praze, Fakulta informatiky a statistiky (hebak@centrum.cz).

Stručný seznam těch nejlepších knih a článků zaměřených na odhady a testy hypotéz by vystačil na rozsáhlou publikaci, takže se raději pokusím pouze připomenout aspoň zcela nejdůležitější skutečnosti týkající se klasické teorie bodových odhadů. Zájemce o tuto problematiku by si podle mého názoru měl uvědomit, že bez jejich poznání a pochopení není možné se ani pokoušet o jakékoli kritické hodnocení či srovnání s bayesovským nebo jiným přístupem. Navíc by si to někteří vynikající statistici (v tomto směru především minulého století) nezasloužili, protože teorie bodového odhadu představuje ucelený přístup. Ve srovnání s teorií intervalových odhadů a testů hypotéz jsou studijní materiály týkající se bodových odhadů asi ty nejpropracovanější. Přesto některá východiska teorie bodových odhadů jsou pro bayesovskou školu nepřijatelná.

Uvažujme náhodný výběr Y_1, Y_2, \dots, Y_n z pravděpodobnostního rozdělení s neznámým parametrem θ a statistiku $T = T(Y_1, Y_2, \dots, Y_n)$, jejíž hodnoty $t = t(y_1, y_2, \dots, y_n)$ odpovídají všem možným náhodným výběrům rozsahu n . Statistika T je náhodnou veličinou a její rozdělení je označováno za *výběrové rozdělení*, což je pro běžné uživatele nebo méně zkušeného statistika velmi obtížně představitelný pojem. Anglická terminologie je v tomto případě o něco výhodnější, protože odlišuje pojem *sample distribution*, kterým se rozumí rozdělení hodnot y v jednom konkrétním (již pořízeném) výběru, od pojmu *sampling distribution*, což je *výběrové rozdělení* statistiky T . Statistika T je náhodná veličina, která *před pořízením výběru* je funkcí *nezávislých a stejně rozdělených náhodných veličin* Y_1, Y_2, \dots, Y_n , zatímco po pořízení výběru je její hodnota funkcí zjištěných **hodnot** veličiny Y v n náhodných pokusech. Znalost rozdělení T je pro klasicky prováděné úsudky z výběru na populaci naprosto rozhodujícím činitelem. Nezapomejme se zde tím, zda a jak lze *výběrové rozdělení* jednoznačně nebo asymptoticky určit, ale v každém případě bude záviset nejen na rozdělení Y , ale i na neznámém parametru θ nebo obecněji na všech neznámých parametrech rozdělení Y . Rozdělení T obsahuje hodnoty statistiky T , které odpovídají **všem možným výběrům** rozsahu n z populace rozsahu N anebo dokonce (při nekonečné nebo hypotetické populaci) zcela **neomezenému počtu výběrů**. Výběrové rozdělení je v ideálním případě v klasické teorii statistických úsudků rozdělením hodnot, které se mohly (nebo by se mohly) vyskytnout při **dané** hodnotě parametru θ . Potíž je tedy v tom, že v klasickém pojetí je parametr θ neznámá konstanta a nikoli (jako v bayesovském pojetí) náhodná veličina. Potom pravděpodobnostní funkce $P(t|\theta)$, hustota pravděpodobnosti $f(t|\theta)$ nebo distribuční funkce $F(t|\theta)$ popisují rozdělení (podle mého názoru všech možných nebo jen hypotetických) hodnot veličiny T , které *by mohly* nastat při dané hodnotě neznámého parametru, což je velice obtížná představa. Každopádně však tento přístup nemá nic společného s konkrétními daty získanými z právě pořízeného náhodného výběru. Klasická statistika však používá právě *výběrové rozdělení* k úsudkům o neznámých parametrech, přičemž navíc vychází z již zmíněného nerealistického předpokladu *nezávislých všech možných náhodných výběrů pořízených za zcela stejných podmínek*. Všechny uvažované pravděpodobnosti používané při induktivních úsudcích (se spornou výjimkou p -hodnot diskutovaných v posledním oddíle) proto nemají nic společného s daty z pořízeného náhodného výběru a jsou stanoveny předem, tedy dříve než výběr byl proveden. Tím však dochází k tomu, že kvalita použitých *statistik* je posuzována

na základě nerealistické abstrakce a teoretické (částečně hypotetické) úvahy. *Bolstad* (2004) dokonce přirovnává tento postup k situaci, kdy *vůz jede před koněm*. Zatímco bayesovská škola se snaží vyvarovat *průměrování* středních čtvercových chyb pro všechny možné hodnoty neznámých parametrů, tak četnostní škola tuto zvyklost obhajuje, s tím že představuje rozumnou aproximaci reality.

Klasický požadavek *nezkreslenosti* odhadu neznámých parametrů je jedna z mnoha intuitivně (*Samaniego* (2010) říká *ad hoc*) stanovených podmínek, které mohou být kladeny na *estimátory*. Jakmile je výběr *vhodných* statistik zúžen na třídu nezkreslených odhadů, může klasicky orientovaný statistik využít propracovanou teorii a s její pomocí najít odhad, který je za jistých podmínek a okolností vhodný nebo dokonce *nejlepší* ve třídě nezkreslených odhadů neznámých parametrů. Významnou roli v této teorii představují *úplné postačující* statistiky, ale i mnoho dalších užitečných nástrojů i kritérií kvality odhadu. Pro nedostatek prostoru se nebudu klasické teorií *bodových odhadů* už dále věnovat. Předpokládám, že čtenář zná nebo se v případě zájmu seznámí s obsahem pojmů, jako jsou *konzistence*; *asymptotická nezkreslenost*; *vydatnost*; *Fisherova míra informace*; *Raova-Cramerova nerovnost*; *metoda maximální věrohodnosti*; *střední čtvercová chyba* a různé podmínky a věty jako je *Raova-Blackwellova* či *Lehmannova-Scheffeova*, které s teorií bodových odhadů úzce souvisí. Vzhledem k zaměření tohoto článku ani nemám jinou možnost než případné zájemce o tuto problematiku odkázat na bohatou literaturu.

Bayesovský přístup k bodovým odhadům

Už bylo řečeno, že klasická statistická teorie považuje vektor $\mathbf{y} = [Y_1, Y_2, \dots, Y_n]^T$ za náhodný, zatímco parametry $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_s]^T$ rozdělí \mathbf{y} označuje za vektor konstant bez ohledu na skutečnost, že v případě úsudků z výběru na populaci jsou tyto parametry téměř výhradně neznámé. Bayesovská statistika rovněž považuje vektor pozorování \mathbf{y} za náhodný, ale na rozdíl od klasického přístupu za náhodný považuje i vektor neznámých parametrů $\boldsymbol{\theta}$. Tato skutečnost je jedním z kritických míst rozdílů mezi klasickými a bayesovskými statistickými úsudky. Bayesovce velmi zajímá *apriorní rozdělí* $\boldsymbol{\theta}$, které po získání a zahrnutí dat představuje užitečný krok k získání potřebného *posteriorního* rozdělí. Zde se budeme výhradně zabývat spojitými parametry, což je i nejčastější případ.

Simultánní hustotu pravděpodobnosti $\boldsymbol{\theta}$ a \mathbf{y} označme jako $h(\boldsymbol{\theta}, \mathbf{y})$, **apriorní hustotu pravděpodobnosti** $\boldsymbol{\theta}$ vyjadřující výchozí postoj k $\boldsymbol{\theta}$ jako $g(\boldsymbol{\theta})$ a *věrohodnostní funkci* vnímanou jako funkci $\boldsymbol{\theta}$ při daném \mathbf{y} jako $f(\mathbf{y}|\boldsymbol{\theta})$. Potom analogicky jako pro náhodné jevy platí, že

$$h(\boldsymbol{\theta}, \mathbf{y}) = f(\mathbf{y} | \boldsymbol{\theta}) g(\boldsymbol{\theta}) = g(\boldsymbol{\theta} | \mathbf{y}) f(\mathbf{y}). \quad (4)$$

Bayesův vzorec pro náhodné veličiny v tomto případě má tvar

$$g(\theta|y) = \frac{f(y|\theta)g(\theta)}{f(y)} = \frac{f(y|\theta)g(\theta)}{\int f(y|\theta)g(\theta)d(\theta)}, \quad (5)$$

ve kterém $g(\theta|y)$ je **posteriorní hustota pravděpodobnosti**, která vyjadřuje modifikovaný postoj k θ po zahrnutí výběrové informace. Vzhledem k θ lze hustotu pravděpodobnosti $f(y)$ považovat za *konstantu* a podmíněnou hustotu $f(y|\theta)$ zapsat (vzhledem k její vypovídací schopnosti) jako věrohodnostní funkci $\ell(\theta|y)$. Potom zjednodušeně můžeme Bayesův vzorec zapsat ve tvaru

$$g(\theta|y) \propto \ell(\theta|y) g(\theta), \quad (6)$$

ve kterém *konstanta* proporcionality je

$$\left[\int f(y|\theta) \cdot g(\theta) \cdot d(\theta) \right]^{-1}. \quad (7)$$

Uvedená úměra volně řečeno vyjadřuje, že posteriorní informace je úměrná (*proporcionální*) *součinu výběrové informace a apriorní informace*.

Uvažujme nejjednodušší případ jednoho neznámého parametru θ . Následující bayesovský přístup k bodovému odhadu vychází z předpokladu, že existuje vhodná ztrátová funkce $L(\theta; \hat{\theta})$, která vyjadřuje důsledky chybného použití $\hat{\theta}$ místo skutečné hodnoty θ . V souladu s *Judge et al.* (1982, s. 90) uvažujme kvadratickou ztrátovou funkci $L = c(\hat{\theta} - \theta)^2$, ve které c je konstanta. Minima tato ztrátová funkce dosahuje pro $\hat{\theta} = \theta$, ale to při neznámé hodnotě θ nemá žádný rozumný význam. Problém lze vyřešit tím, že místo minimalizace L se minimalizuje střední hodnota L přes všechny možné hodnoty θ při vahách $g(\theta|y)$. Potom hledaný odhad $\hat{\theta}$, který minimalizuje

$$E_{\theta|y} [L(\theta; \hat{\theta})] = \int c(\hat{\theta} - \theta)^2 g(\theta|y) d\theta, \quad (8)$$

dostáváme po úpravách, které uvádí *Spiegel* (1963) ve tvaru

$$\hat{\theta} = E(\theta|y) = \int \theta g(\theta|y) d\theta, \quad (9)$$

což není nic jiného než **podmíněná střední hodnota** posteriorního rozdělení θ . Analogicky při použití ztrátové funkce $L(\theta; \hat{\theta}) = c|\hat{\theta} - \theta|$ bychom minimalizací L získali odhad $\hat{\theta}$, kterým by byl 50% kvantil posteriorního rozdělení $g(\theta|y)$.

Aplikujeme-li pro srovnání použití kvadratické ztrátové funkce na klasický přístup při odhadu θ zjistíme, že *vhodným bodovým odhadem* θ by byla hodnota $\hat{\theta}$, která minimalizuje střední čtvercovou chybu $MSE(\hat{\theta}; \theta)$, ale tu nelze pro všechny hodnoty θ obecně určit. Běžně se stává, že určitá statistika má pro některé hodnoty θ menší střední čtvercovou chybu než konkurující statistiky, ale pro jiné hodnoty θ je tomu jinak.

Při klasickém pojetí má přednost nezkreslený odhad T nebo aspoň asymptoticky nezkreslený odhad θ , protože při *větším počtu výběrů* (což je jen hypotetická úvaha vzhledem k jedinému výběru, který už jsme provedli anebo máme v úmyslu provést) se střední hodnota T neliší od odhadované hodnoty θ (nebo se liší, ale s růstem n se rozdíl zmenšuje). Ve třídě nezkreslených odhadů je pak (jak už bylo uvedeno) pochopitelně preferován takový odhad, který má nejmenší rozptyl, protože malé kolísání kolem *správné* střední hodnoty působí velmi nadějným dojmem. Na stejných principech jsou pak vybudovány i metody intervalového odhadu, a s tím i úzce související testovací procedury. Teoreticky se v klasickém přístupu připouštějí i zkreslené odhady, ale jen tehdy, je-li střední čtvercová chyba

$$MSE(T; \theta) = E\{[T - \theta]^2\} = \int (\hat{\theta} - \theta)^2 f(\hat{\theta} | \theta) d\hat{\theta} = D(T) + [B(T; \theta)]^2 \quad (10)$$

menší než u jiných v úvahu přicházejících statistik. V uvedeném vyjádření je $f(\hat{\theta} | \theta)$ hustota pravděpodobnosti *výběrového rozdělení* $\hat{\theta}$ pro dané θ . Ve skutečnosti se však v *tomto pojetí* zkreslené odhady uvažují spíše jen teoreticky pro srovnání různých statistik, protože zkreslení odhadu $B(T, \theta) = E(T) - \theta$, a tedy rovněž střední čtvercová chyba $MSE(T; \theta)$, je funkcí nejen použité statistiky T , ale i hodnoty neznámého parametru θ . Bylo by pěkné minimalizovat $MSE(T; \theta)$ přes všechny hodnoty θ , ale to obecně není možné. Dokresleme si na následujícím příkladu možnost použití *zkresleného odhadu* místo *nejlepšího nezkresleného odhadu*.

Příklad 3: Bodový odhad neznámé střední hodnoty normálního rozdělení

Při tradičním výkladu teorie bodového odhadu je asi nejčastěji uváděným případem odhad střední hodnoty $E(X) = \mu$ veličiny X , která má normální rozdělení s parametrem μ a se známým nebo neznámým parametrem (rozptylem) $D(X) = \sigma^2$. Aritmetický průměr \bar{x} z n náhodně získaných pozorování je nejen nejlepším lineárním nezkresleným odhadem μ , ale má i jiné velmi dobré statistické vlastnosti. Nabízí se otázka, zda je \bar{x} dobrým nebo dokonce *nejlepším* odhadem i z hlediska střední čtvercové chyby $MSE(\bar{x}; \mu)$. Při hledání odpovědi na otázku, zda neexistuje jiný (připustíme, že zkreslený) odhad μ , jehož střední čtvercová chyba $MSE(\bar{x}; \mu)$ by byla menší než $D(\bar{x}) = \sigma^2/n$. Zvolme pro odhad μ statistiku $T = k\bar{x}$, kde k je konstanta. Střední hodnota $E(\bar{x}) = kE(\bar{x}) = k\mu$, takže T je zkresleným odhadem μ , přičemž velikost zkreslení $B(k\bar{x}; \mu) = k\mu - \mu = (k - 1)\mu$.

Rozptyl $D(T) = k^2 D(\bar{x}) = k^2 \sigma^2/n$, takže $MSE(k\bar{x}; \mu) = k^2 \sigma^2/n + (k - 1)^2 \mu^2$. Rozdíl čtvercových chyb

$$MSE(\bar{x}; \mu) - MSE(k\bar{x}; \mu) = \sigma^2/n - k^2 \sigma^2/n - (k - 1)^2 \mu^2 = (1 - k)(1 + k) \sigma^2/n - (1 - k)^2 \mu^2.$$

Pro konstantu $-1 < k < 1$ je první výraz kladný a k tomu, aby i rozdíl obou středních čtvercových chyb byl kladný, je třeba, aby $k > (c - 1)/(c + 1)$, kde $c = n\mu^2/\sigma^2$. Jinak řečeno, pro k větší než -1 i větší než $(c - 1)/(c + 1)$, ale zároveň menší než 1 , je $MSE(k\bar{x}; \mu)$ menší než $MSE(\bar{x}; \mu)$. Ukázali jsme si, že z hlediska střední čtvercové chyby existuje odhad, který pro *některá* n , μ a σ^2 je lepší než běžně používaný BLUE odhad \bar{x} . Pro $n = 10$, $\mu = 3$ a $\sigma = 10$ je $c = 0,9$. Pak

pro $-1/19 < k < 1$ a $n = \sigma = 10$ má odhad μ pomocí $T = k\bar{x}$ menší MSE než odhad μ pomocí \bar{x} . Z toho tedy vyplývá, že z hlediska MSE je odhad μ pro $|\mu| < \sqrt{190}$ a $n = \sigma = 10$ lepší pomocí $0,9\bar{x}$, než pomocí \bar{x} . Pro danou situaci tedy existuje z hlediska MSE lepší odhad pro neznámou střední hodnotu μ než je aritmetický průměr, ale jen když $|\mu| < \sqrt{190}$. Máme-li však navíc k dispozici informaci, že hodnota μ se s vysokou pravděpodobností nachází v tomto intervalu, pak bychom určitě (pro $n = \sigma = 10$) dali přednost odhadu $0,9\bar{x}$, před \bar{x} .

Hodnocení kvality bayesovského bodového odhadu

Kritériem kvality bayesovského bodového odhadu parametru je posteriorní střední hodnota čtvercových odchylek odhadu $\hat{\theta}$ od skutečné hodnoty θ která se značí $PMS(\hat{\theta})$, ve smyslu anglického názvu *posterior mean square of an estimate*. Poznamenejme, že anglická terminologie odlišuje pojem *estimator*, kterým se rozumí *statistika* (nebo též *estimátor*), tedy náhodná veličina určená k odhadu neznámého parametru, od pojmu *estimate*, což je označení konkrétní hodnoty této veličiny v daném výběru. Potom

$$PMS(\hat{\theta}) = \int (\theta - \hat{\theta})^2 g(\theta | y) d\theta, \quad (11)$$

kde $PMS(\hat{\theta})$ měří očekávaný posteriorní čtvercový rozdíl mezi θ a jeho odhadem $\hat{\theta}$. Po jednoduchých úpravách dostáváme

$$PMS(\hat{\theta}) = D(\theta | y) + [E(\theta | y) - \hat{\theta}]^2, \quad (12)$$

jelikož vzhledem k posteriornímu rozdělení oba výrazy v hranaté závorce jsou konstanty. Tedy ve shodě s již uvedeným výsledkem, který byl získán (již dříve) minimalizací střední hodnoty kvadratické ztrátové funkce, je vidět, že žádný jiný odhad $\hat{\theta}$ než posteriorní střední hodnota $E(\theta | y)$ nemůže mít menší hodnotu $PMS(\hat{\theta})$.

Klasické statistické úsudky i střední čtvercová chyba $MSE(\hat{\theta}; \theta)$ vycházejí z **výběrového rozdělení** $f(\hat{\theta} | \theta)$ v situaci **před** provedením výběru, kdy $\hat{\theta}$ je náhodná veličina a θ je neznámá konstanta. Všechny pravděpodobnosti a z nich vyplývající úsudky lze vztahovat jen k této situaci a nemají nic společného s konkrétními daty ze zamýšleného nebo už provedeného výběru. Jedinou velice spornou výjimkou jsou p -hodnoty při testování hypotéz, ale k těm se ještě vrátíme v oddíle 6. Podle *Bolstada* (2004) bayesovci proto všechny úsudky, které se opírají o výběrové rozdělení, označují za *PRE-DATA* úsudky. Ve srovnání s tím se bayesovským bodovým odhadem rozumí **konkrétní hodnota** charakteristiky posteriorního rozdělení, která je s pomocí **apriorního rozdělení** $g(\theta)$ vypočítaná na základě dat z provedeného **výběru**. Touto charakteristikou je nejčastěji posteriorní střední hodnota, protože minimalizuje posteriorní střední čtvercovou chybu, jak jsme dvakrát různým způsobem už ověřili. Proto se bayesovské úsudky na základě charakteristik posteriorního rozdělení označují za *POST-DATA* úsudky.

Bayesův vzorec umožňuje revidovat výchozí postoj k θ pomocí jeho pravděpodobnostního rozdělení, které je (vzhledem k tomu, že θ je považováno za spojitou náhodnou veličinu) popsáno apriorní hustotou pravděpodobnosti $g(\theta)$. Kritika

bayesovského přístupu se často dotýkala skutečnosti, že o parametru θ nemusí být k dispozici žádná výchozí znalost. V tomto případě se říká, že apriorní znalost o parametru θ je *neinformativní* (nebo též *vágní* či *difuzní*). Mohou nastat i případy, ve kterých jistá informace θ existuje, ale je obtížné funkci $g(\theta)$ najít nebo ji vhodným způsobem matematicky vyjádřit. Otázka hledání výchozího rozdělení θ není nijak nová a pokoušeli se na ni odpovědět už *Bayes* a *Laplace* a po nich i *K. Pearson*, *Student (Gosset)* a další. Tento problém je častým námětem téměř všech bayesovských zaměřených učebnic i širěji orientovaných monografií. Například *Bolstad* v *Introduction to Bayesian Statistics* (2004) věnuje této otázce velkou pozornost a v různých částech této kvalitní učebnice bayesovské statistiky se k ní vrací a nabízí některé postupy, které mohou uživateli pomoci při hledání vhodného apriorního rozdělení.

Stejně jako *Bayes* vychází i *Bolstad* (2004), *Press* (2003) a řada dalších statistiků z nejpropracovanějšího případu, kterým je apriorní i posteriorní rozdělení parametru π alternativního rozdělení. Jiným příkladem je *Zellner* (1971), který se nejdříve věnuje parametrům normálního rozdělení a teprve později dalším rozdělením a v souvislosti s volbou apriorního rozdělení začíná právě případem *vágní* informace. Pokračuje datově orientovanými informacemi z minulosti (DB) a teoretickými nebo jinými nedatově orientovanými informacemi (NDB). Dokonalým výkladem myšlenek bayesovské statistiky a významu správné volby apriorního rozdělení π binomického rozdělení pro konstrukci bodového i intervalového odhadu je článek *D. W. Lindleyho* a *L. D. Phillipse* (1976), který vznikl na základě třídílného televizního seriálu autorů před více než 35 roky.

Víme, že v bayesovské statistice (po získání dat \mathbf{y}) je posteriorní hustota pravděpodobnosti $g(\theta | \mathbf{y})$ normalizovaným součinem apriorní hustoty pravděpodobnosti $g(\theta)$ a věrohodnostní funkce $\ell(\theta | \mathbf{y})$. Patří-li rozdělení veličiny Y do *rodiny exponenciálního typu*, pak existuje *konjugované apriorní* rozdělení, které rovněž patří do této rodiny. Jinak řečeno, je-li posteriorní $g(\theta | \mathbf{y})$ členem stejné *rodiny rozdělení* jako apriorní $g(\theta)$, pak obě dvě tato rozdělení jsou *konjugovaná* (neboli obecně stejného typu).

Podobně jako v článku *Lindleyho* a *Phillipse* (1976) a v souladu s přístupem, který zvolil *Bolstad* (2004), jsem pro tento příspěvek zvolil (pro ilustraci bayesovského bodového odhadu) jako ukázkou parametr π alternativního rozdělení. *Press* (2003, s. 191 až 208) pro případné zájemce nabízí přehled apriorních rozdělení pro parametry téměř všech často se vyskytujících jednorozměrných rozdělení a uvádí zde i případ apriorního rozdělení pro parametry vícerozměrného normálního rozdělení. Pro tento krátký příspěvek je doufám dostačující následující ukáзка používaného postupu.

Apriorní rozdělení parametru π

Označme jako Y počet *úspěchů* v n nezávislých pokusech, když v každém pokusu je výsledkem *úspěch* nebo *neúspěch* a pravděpodobnost úspěchu je vždy π , $0 < \pi < 1$. Pak pravděpodobnostní funkce je

$$P(y|\pi) = \binom{n}{y} \pi^y (1-\pi)^{n-y} \text{ pro } y = 0, 1, \dots, n. \quad (13)$$

Neinformativním apriorním rozdělením je *rovnoměrné rozdělení* s hustotou pravděpodobnosti $g(\pi) = 1$ pro $0 < \pi < 1$, které modeluje situaci, ve které se o parametru π nic neví, a žádné hodnotě π se nedává přednost: Při použití apriorního rovnoměrného rozdělení tudíž všechny stejně velké intervaly hodnot π od nuly do jedné považujeme za stejně pravděpodobné. Ukažme si, že obecně vhodným modelem apriorního rozdělení je hustota pravděpodobnosti $g(\pi)$ beta *rozdělení* s nezápornými parametry a , b s hustotou pravděpodobnosti, které označíme jako $Be(a; b)$. Pak apriorní rozdělení parametru π lze zapsat ve tvaru

$$g(\pi; a, b) = k\pi^{a-1}(1-\pi)^{b-1}. \quad (14)$$

Pro známé kladné hodnoty a, b vypočítáme konstantu k jako

$$k = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \quad (15)$$

pomocí gama funkce $\Gamma(\cdot)$, přičemž je-li c přirozené číslo, pak platí, že $\Gamma(c) = (c-1)!$ Podstatné je, že výraz $\pi^{a-1}(1-\pi)^{b-1}$ určuje tvar rozdělení π při různých hodnotách a, b , zatímco konstanta k zabezpečuje, že nezáporná normalizovaná funkce $g(\pi)$ má vlastnosti hustoty pravděpodobnosti, takže

$$\int_0^1 g(\pi) d\pi = \int_0^1 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1} d\pi = 1. \quad (16)$$

Po provedení náhodného výběru rozsahu n z rozdělení veličiny Y s neznámým parametrem π dostáváme y jedniček a $n-y$ nul. Nestejná pravděpodobnostní úvaha o situaci před provedením a po provedení výběru vyžaduje odlišení zápisu formálně stejné *pravděpodobnostní funkce* $P(y|\pi)$, pro dané π a neznámé y , od významově obrácené *věrohodnostní funkce*

$$l(\pi|y) = \binom{n}{y} \pi^y (1-\pi)^{n-y}, \text{ pro dané } y \text{ a neznámé } \pi, 0 < \pi < 1. \quad (17)$$

Vzhledem k tomu, že

$$g(\pi|y) \propto g(\pi) \cdot l(\pi|y), \quad (18)$$

dostáváme (po vynechání konstant nezávisících na π apriorního beta rozdělení a věrohodnostní funkce) úměru

$$g(\pi|y) \propto \pi^{a+y-1}(1-\pi)^{b+n-y-1}. \quad (19)$$

Dělíme-li (podle Bayesova vzorce) pravou stranu výše uvedeného úměry *konstantou* $f(y)$ dostáváme hustotu pravděpodobnosti posteriorního rozdělení

$$\begin{aligned} g(\pi|y) &= \frac{g(\pi)P(y|\pi)}{\int_0^1 g(\pi)P(y|\pi)d\pi} = \\ &= \frac{\Gamma(a+b+n)}{\Gamma(a+y)\Gamma(b+n-y)} \pi^{a+y-1}(1-\pi)^{b+n-y-1}, 0 < \pi < 1. \end{aligned} \quad (20)$$

Dříve než přejdeme k bodovému odhadu π , uveďme čtyři důležité poznámky:

1. Pokud není zvolené apriorní rozdělení $g(\pi)$ konjugovaným rozdělením, může výpočet konstanty ve jmenovateli Bayesova vzorce vyžadovat i složité numerické řešení. Beta rozdělení však patří do rodiny konjugovaných rozdělení, takže se shoduje obecný funkční tvar apriorního a posteriorního rozdělení (až na konstanty nezávislé na π a hodnoty zvolených parametrů a, b). Výsledné posteriorní rozdělení je rovněž *beta* s parametry $a+y$ a $b+n-y$.
2. Pro určení sdružených, marginálních i podmíněných rozdělení není na překážku skutečnost, že náhodná veličina Y je diskrétní náhodná veličina, zatímco náhodná veličina π je (v bayesovském pojetí) spojitá náhodná veličina. Je-li X spojitá náhodná veličina s hodnotami x a Y je diskrétní náhodná veličina s hodnotami y , pak $f(x, y)$ je sdružená funkce těchto dvou náhodných veličin. Marginální pravděpodobnostní funkci $P(y)$ lze získat integrací sdružené hustoty $f(x, y)$ podle x a analogicky marginální hustotu pravděpodobnosti $f(x)$ lze získat součtem sdružené funkce $f(x, y)$ přes všechny hodnoty y . Podmíněná hustota pravděpodobnosti $f(x|y)$ je podílem $f(x, y)/P(y)$ a podmíněná pravděpodobnostní funkce $P(y|x)$ je analogicky podílem $f(x, y)/f(x)$, protože pro dané x je $f(x)$ konstanta určená jako součet $f(x, y)$ přes všechny hodnoty diskrétní veličiny Y .
3. Víme, že vhodným apriorním rozdělením pro případ, kdy nedisponujeme žádnou výchozí informací, je *rovnoměrné rozdělení* s hustotou pravděpodobnosti $g(\pi) = 1$, pro $0 < \pi < 1$. Apriorní hustota pravděpodobnosti $g(\pi) = g(\pi; a, b) = k\pi^{a-1}(1-\pi)^{b-1}$ se pro $a = b = 1$ redukuje na apriorní hustotu pravděpodobnosti $g(y) = 1$ neboli apriorní rovnoměrné rozdělení je speciální případ beta rozdělení $Be(1; 1)$. Apriorní beta rozdělení s $a = b$ je symetrické, když žádnému z výsledků pokusu nedáváme přednost, takže naše pravděpodobnost *úspěchu* se rovná pravděpodobnosti *neúspěchu*, neboli $\pi = 1 - \pi = 0,5$. Čím $a = b$ je větší, tím větší je i naše přesvědčení (pravděpodobnost), že oba možné výsledky jsou stejně pravděpodobné. Například při apriorní úvaze o pravděpodobnosti, že napínáček při jednom hození padne špičkou nahoru, by se mnohý z nás asi rozhodl pro neinformativní apriorní rovnoměrné rozdělení s $a = b = 1$, zatímco pro hod mincí by volba např. $a = b = 50$ vyjadřovala poměrně velkou jistotu, že obě strany mince jsou stejně pravděpodobné. Na bayesovských úsudcích je zvláště pěkné, že *postoj k napínáčkům je jiný než k mincím*. Liší se tedy i jejich apriorní a posteriorní pravděpodobnosti před a po získání výsledků z několika málo hodů.

4. Pro úsudky pomocí beta rozdělení lze využít centrální limitní větu. Aproximace normálním rozdělením při určování (např. kvantilů) jsou poměrně přesné už pro hodnoty a, b přibližně větší než 10. Posteriorní rozdělení s parametry $a + n$ a $b + n - y$ se tedy už pro poměrně malé rozsahy výběru rychle blíží svým tvarem k normálnímu rozdělení. Z toho vyplývá, že pro uvedený příklad s mincemi a s apriorním rozdělením $Be(50; 50)$ je jasné, že nejen náš výchozí, ale i posteriorní postoj k pravděpodobnostem výsledků hodů lze velmi dobře modelovat normálním rozdělením.

Střední hodnota a rozptyl apriorního beta rozdělení parametru π

Hustotu pravděpodobnosti apriorního beta rozdělení $Be(a; b)$ pro parametr $0 < \pi < 1$ a přirozená čísla a, b jsme už výše uvedli ve tvaru

$$g(\pi) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1}. \quad (21)$$

Apriorní střední hodnota a rozptyl tohoto rozdělení potom jsou

$$E(\pi) = \int_0^1 \pi \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1} d\pi = \frac{a}{a+b}, \quad (22)$$

$$D(\pi) = \int_0^1 \left(\pi - \frac{a}{a+b}\right)^2 \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \pi^{a-1} (1-\pi)^{b-1} d\pi = \frac{ab}{(a+b)^2(a+b+1)}. \quad (23)$$

V některých úlohách nemusí být pro uživatele snadné vyjádřit svůj apriorní postoj k hodnotám a, b . Je-li však pro něj jednodušší vyjádřit svůj postoj k apriorní střední hodnotě $E(\pi)$ a k apriornímu rozptylu $D(\pi)$ (nebo asi častěji k apriorní směrodatné odchylce), stačí do výše uvedených dvou rovnic dosadit apriorní střední hodnotu a apriorní směrodatnou odchylku a jejich řešením získat apriorní hodnoty a, b .

Charakteristiky posteriorního rozdělení parametru π

Z předchozích poznatků už vyplynulo, že po provedení výběru o rozsahu n a získání hodnoty y , kterou je počet výskytů jevu A v n *zaměnitelných* pokusech (v klasické terminologii hodnota výběrového úhrnu m statistiky M) při využití věrohodnostní funkce (17), dostáváme posteriorní beta rozdělení $Be(a + y; b + n - y)$ s hustotou pravděpodobnosti (20).

Velkou výhodou bayesovského přístupu je skutečnost, že k bodovému i intervalovému odhadu, jakož i pro testování statistických hypotéz, je znalost posteriorního rozdělení těchto parametrů zcela dostačující informací. Při této příležitosti poznamenejme, že zcela obecně požadavek *zaměnitelných pokusů* je mnohem mírnější a snadněji ověřitelný než požadavek nezávislých pokusů. Koncepce zaměnitelných pokusů je výsledek práce *de Finettiho* z roku 1937. Při provádění úsudků o parametru π stačí vědět, že všechna pořadí výsledků s y *úspěchy* a s $n - y$ *neúspěchy* mají stejnou pravděpodobnost $\pi^y(1 - \pi)^{n-y}$. Stručně uveďme jen zcela nejpoužívanější charakteristiky posteriorního rozdělení, kterými jsou

$$\text{posteriorní střední hodnota } E(\pi | y) = \frac{a + y}{a + b + n}; \quad (24)$$

posteriorní P100% kvantil π_p , pro $0 < P < 1$, určený řešením rovnice

$$P = \int_0^{\pi_p} g(\pi | y) d\pi; \quad (25)$$

$$\text{posteriorní rozptyl } D(\pi | y) = \frac{(a + y)(b + n - y)}{(a + b + n)^2(a + b + n + 1)}. \quad (26)$$

Bylo už dokázáno, že z hlediska posteriorní střední hodnoty čtvercových odchylek odhadu $\hat{\theta}$ od skutečné hodnoty θ je **nejlepším bayesovským odhadem střední hodnota posteriorního rozdělení**, neboli zde právě uvedená $E(\pi | y)$. Zmíňme, že při neurčitém (vágním) apriorním rozdělení s $g(\pi) = 1$ je (po dosazení $a = b = 1$) posteriorní střední hodnota $E(\pi | y) = (y + 1)/(n + 2)$.

Klasický přístup ve zcela stejné situaci *Bernoulliho* pokusů se dvěma možnými výsledky, kterými je *úspěch* (jev A) nebo *neúspěch* (jev opačný k jevu A), je sice dost známý a částečně jsme jej už popsali, ale kvůli srovnání s bayesovským bodovým odhadem si velice stručně shrňme jeho ústřední východiska.

Náhodná veličina Y , kterou je *počet úspěchů* (výskytů jevu A) při n nezávislých náhodných pokusech, má *binomické* rozdělení (12) s parametry n a π . Pravděpodobnost $P(A) = \pi$ je v každém pokusu stejná a podle charakteru úlohy (pravděpodobnostní nebo statistická) je π známá nebo neznámá konstanta. Konstanta π je interpretována jako *objektivní* pravděpodobnost jevu A v každém pokusu, bez ohledu na naši možnost ji stanovit. Podle běžných kritérií je *nejlepším* bodovým odhadem této pravděpodobnosti výběrová relativní četnost jevu A , neboli (po provedení výběru rozsahu n) hodnota

statistiky P , kterou je $p = m/n$, neboli podíl počtu *úspěšných* pokusů, ve kterých jev A nastal, k celkovému počtu provedených nezávislých pokusů. Ke statistice M (a tedy i ke statistice P) vede i *metoda maximální věrohodnosti*, jakož i další postupy či (převážně už zmíněná) kritéria kvality klasického bodového odhadu. Kromě jiných (příznivých) vlastností je tento odhad *nezkreslený* a jeho rozptyl $D(P) = \pi(1 - \pi)/n$ je nejmenší ze všech možných nezkreslených odhadů π . Při platnosti dříve uvedených podmínek je tedy v klasickém pojetí **nejlepším nezkresleným odhadem parametru π výběrová relativní četnost jevu A** .

Srovnání přesnosti klasického a bayesovského bodového odhadu parametru π

Několikrát jsem naznačil, že tento příspěvek se v mnohém opírá o rozsáhlou bayesovskou literaturu a týká se to i případu, kterým se teď zabýváme. Za nejvíce přesvědčivý považuji výklad *Bolstad* (2004) v části zaměřené na srovnání přístupu klasické a bayesovské školy k odhadu parametru π binomického rozdělení, a proto jsem i z některých jeho argumentů pro následující srovnání vycházel.

Víme, že bayesovským kritériem kvality odhadu je *posteriorní* střední čtvercová chyba, ale zvýhodněme klasický odhad a pro srovnání použijme střední čtvercovou chybu. Navíc připuťme (pro bayesovský odhad nejhorší) situaci, ve které není k dispozici žádná apriorní informace. Pro určení posteriorního rozdělení π vyjděme z apriorního rovnoměrného rozdělení, kterým je beta rozdělení s parametry $a = b = 1$.

Pomocí střední čtvercové chyby odhadu $MSE(\hat{\pi}; \pi)$ tedy porovnáváme klasický odhad $\hat{\pi}_K = y/n$ s bayesovským odhadem $\hat{\pi}_B = (a + y)/(a + b + n)$, který se pro $a = b = 1$ zjednoduší na $(y + 1)/(n + 2)$.

Klasický odhad je nezkreslený, takže střední čtvercová chyba odhadu se redukuje na rozptyl odhadu a dostáváme

$$MSE(\hat{\pi}_K; \pi) = D(\hat{\pi}_K) + 0^2 = \pi(1 - \pi)/n. \quad (27)$$

Při provedených zjednodušeních a při klasickém vnímání *výběrového rozdělení* statistiky $\hat{\pi}_B$ je rozptyl bayesovského odhadu $D(\hat{\pi}_B) = D[(y + 1)/(n + 2)] = n\pi(1 - \pi)/(n + 2)^2$. Bayesovské odhady jsou obecně zkreslené, ale v tomto případě je snadné ukázat, že velikost zkreslení je $(n\pi + 1)/(n + 2) - \pi$. Po několika úpravách (viz např. *Bolstad* 2004, s. 152) dostáváme střední čtvercovou chybu bayesovského odhadu, která je

$$MSE(\hat{\pi}_B; \pi) = D(\hat{\pi}_B) + [B(\hat{\pi}_B; \pi)]^2 = n\pi(1 - \pi)/(n + 2)^2 + [(1 - 2\pi)/(n + 2)]^2. \quad (28)$$

Tabulka 1**Srovnání středních čtvercových chyb odhadu pro různá n a π** Pro klasický odhad: $MSE(\hat{\pi}_K; \pi) = \pi(1 - \pi)/n$

n/π	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,95	Průměr	Sm. odch.
5	0,0095	0,0180	0,0320	0,0420	0,0480	0,0500	0,0480	0,0095	0,032125	0,016372
10	0,0048	0,0090	0,0160	0,0210	0,0240	0,0250	0,0240	0,0048	0,016063	0,008186
20	0,0024	0,0045	0,0080	0,0105	0,0120	0,0125	0,0120	0,0024	0,008031	0,004093
30	0,0016	0,0030	0,0053	0,0070	0,0080	0,0083	0,0080	0,0016	0,005354	0,002729
40	0,0012	0,0023	0,0040	0,0053	0,0060	0,0063	0,0060	0,0012	0,004016	0,002047
50	0,0010	0,0018	0,0032	0,0042	0,0048	0,0050	0,0048	0,0010	0,003213	0,001637
100	0,0005	0,0009	0,0016	0,0021	0,0024	0,0025	0,0024	0,0005	0,001606	0,000819

Pro bayesovský odhad: $MSE(\hat{\pi}_B; \pi) = n\pi(1 - \pi)/(n + 2)^2 + [(1 - 2\pi)/(n + 2)]^2$

n/π	0,05	0,1	0,2	0,3	0,4	0,5	0,6	0,95	Průměr	Sm. odch.
5	0,0214	0,0222	0,0237	0,0247	0,0253	0,0255	0,0253	0,0214	0,023686	0,001671
10	0,0089	0,0107	0,0136	0,0157	0,0169	0,0174	0,0169	0,0089	0,013637	0,003411
20	0,0036	0,0050	0,0074	0,0090	0,0100	0,0103	0,0100	0,0036	0,007376	0,002706
30	0,0022	0,0033	0,0050	0,0063	0,0071	0,0073	0,0071	0,0022	0,005055	0,002078
40	0,0016	0,0024	0,0038	0,0049	0,0055	0,0057	0,0055	0,0016	0,003852	0,001661
50	0,0012	0,0019	0,0031	0,0039	0,0045	0,0046	0,0045	0,0012	0,003102	0,001393
100	0,0005	0,0009	0,0016	0,0020	0,0023	0,0023	0,0023	0,0005	0,001578	0,000755

Výsledky výpočtů v uvedených tabulkách jednoznačně ukazují, že:

1. Pro uvažovaná n jsou $MSE(\hat{\pi}_K; \pi)$ vždy vyšší pro $\pi = 0,2; 0,3; 0,4; 0,5; 0,6; 0,7; 0,8$ než $MSE(\hat{\pi}_B; \pi)$ a nižší jen pro $\pi = 0,05; 0,10; 0,90; 0,95$.
2. Průměry středních čtvercových chyb klasického odhadu jsou pro všechna zvolená n a π vždy vyšší než bayesovského odhadu.
3. Směrodatné odchylky středních čtvercových chyb pro všechna zvolená n a π jsou vždy výrazně vyšší pro klasický odhad než pro bayesovský odhad.
4. Hodnoty středních čtvercových chyb obou odhadů jsou symetrické kolem $\pi = 0,5$ ve smyslu rovnosti $MSE(\hat{\pi}; \pi) = MSE(\hat{\pi}; 1 - \pi)$ a nejvyšší právě pro $\pi = 1 - \pi = 0,5$.
5. S růstem rozsahu výběru n klesají střední čtvercové chyby obou odhadů.
6. S růstem n se pro každé π zmenšují rozdíly mezi středními čtvercovými chybami klasického a bayesovského odhadu.

Provedené výpočty jen potvrdily teoreticky známé skutečnosti, jakož i závěry učiněné na základě výpočtů různých autorů. Vyplývá z nich (nejen pro π , ale i pro parametry jiných běžně používaných rozdělání), že bayesovské bodové odhady jsou *obecně přesnější* než klasické odhady. Tím se samozřejmě **nemyslí**, že střední čtvercové chyby při všech hodnotách neznámých parametrů **jsou vždy nižší** při bayesovském než při klasickém bodovém odhadu, což potvrzují i výpočty pro hodnoty π asi od 0 do 0,15 a od 0,85 do 1. Stejný závěr pro $n = 10$ je vidět i na grafu *Bolstad* (2004, s. 153) a zde je zřejmé, že to platí i pro jiná n . Připomeňme si, že v provedeném srovnání se v obou případech vycházelo z klasického kritéria střední čtvercové chyby a byl uvažován pro (bayesovce nejhorší) případ neexistující apriorní informace. Využila se jen možnost použití rovnoměrného apriorního rozdělání s $g(\pi) = 1$ pro všechna π . Nelze přehlédnout ani výrazně nižší směrodatné odchylky středních čtvercových chyb bayesovského odhadu, které jen dokumentují větší robustnost bayesovského odhadu $\hat{\pi}_B$ ve srovnání s klasickým odhadem $\hat{\pi}_K$.

Tabulka 2

Srovnání klasického odhadu a bayesovského odhadu pro různá n a y

n	y_1	y_2	y_3	y_1/n	y_2/n	y_3/n
5	2	3	4	0,400	0,600	0,800
10	1	4	8	0,100	0,400	0,800
20	4	8	18	0,200	0,400	0,900
30	10	17	26	0,330	0,570	0,870
40	18	27	35	0,450	0,675	0,875
50	20	35	45	0,400	0,700	0,900
100	39	40	80	0,390	0,400	0,800

n	y_1	y_2	y_3	$(y_1 + 1)/(n + 2)$	$(y_2 + 1)/(n + 2)$	$(y_3 + 1)/(n + 2)$
5	2	3	4	0,429	0,571	0,714
10	1	4	8	0,167	0,417	0,750
20	4	8	18	0,227	0,409	0,864
30	10	17	26	0,344	0,563	0,844
40	18	27	35	0,452	0,667	0,857
50	20	35	45	0,404	0,692	0,885
100	39	40	80	0,392	0,402	0,794

Vzhledem k neznámému π mnoho neříkají hodnoty $\hat{\pi}_K = y/n$ a $\hat{\pi}_B = (y + 1)/(n + 2)$ pro zvolené hodnoty $n = 5; 10; 20; 30; 40; 50; 100$ a tři různé skupiny hodnot úhrnů $y_1; y_2; y_3$. Nikoho nepřekvapí, že pro různé úhrny y a různé rozsahy výběru n může být klasický odhad vyšší i nižší než bayesovský odhad. Rovněž je samozřejmé, že s růstem rozsahu výběru n se zmenšuje rozdíl mezi hodnotami klasického a bayesovského

odhadu, bez jakékoli souvislosti s tím, zda původně byla vyšší hodnota klasického nebo bayesovského odhadu. Zajímavé je, že ve všech 21 případech je (z hlediska možné velikosti výběrové chyby) větší část intervalu možných hodnot π příznivá bayesovskému odhadu. Například pro první hodnotu $y_1 = 2$ při $n = 5$ je $\hat{\pi}_K = y/n = 2/5 = 0,4$, zatímco $\hat{\pi}_B = (y + 1)/(n + 2) = 3/7$, což je přibližně 0,429. Srovnání možných výběrových chyb pro $0 < \pi < 1$ jasně ukazuje, že hodnoty do **0,4 + (3/7 – 0,4)/2 = 29/70, tedy přibližně do 0,414**, by vedly k závěru, že klasický odhad je lepší, zatímco pro výrazně větší (!) interval hodnot **od 0,414** bychom preferovali bayesovský odhad. Vzhledem ke zvolenému apriornímu rovnoměrnému rozdělení, při kterém všechny stejně velké intervaly hodnot π jsou i stejně pravděpodobné, je možné říci (zatím jen v tomto případě), že bayesovský odhad π má větší pravděpodobnost menší výběrové chyby. Hodnoty y i rozsahy výběru jsem bez jakéhokoli záměru zvolil a přesto je z výpočtů vidět, že právě učiněný závěr se týká všech 21 uvedených dvojic odhadů.

Poznámka 1

Klasický odhad y/n se rovná střední hodnotě posteriorního rozdělení jen když apriorní rozdělení $g(\pi)$ je úměrné výrazu $\pi^{-1}(1 - \pi)^{-1}$. Takové apriorní rozdělení je však *nevhodné (improper)*, protože pro uvedené $g(\pi)$ nelze určit funkci, jejíž integrál přes všechny hodnoty π (od nuly do jedné) by se rovnal jedné.

Poznámka 2

Bodový odhad parametru je obecně *přijatelný (admissible)*, neexistuje-li žádný jiný odhad, který by měl menší střední čtvercovou chybu odhadu pro všechny možné hodnoty odhadovaného parametru. Střední hodnota posteriorního rozdělení při každém jiném (říkejme *vhodném*) apriorním rozdělení je vždy přijatelným odhadem a dokonce někdy i při nevhodném apriorním rozdělení může být střední hodnota posteriorního rozdělení přijatelným odhadem.

Literatura k 2. části

- ANDĚL, J. *Matematická statistika*. Praha : SNTL; Alfa, 1978.
- BAYARRI, M. J.; BERGER, J. O. The Interplay of Bayesian and Frequentist Analysis. *Statistical Science*. 2004, vol. 19, s. 58–80.
- BOLSTAD, W. M. *Introduction to Bayesian statistics*. Hoboken : John Wiley and Sons, 2004.
- EFRON, B. *Large Scale Inference – Empirical Bayes Methods for Estimation, Testing and Prediction*. Cambridge; New York : Cambridge University Press, 2010.
- DUDEWICZ, E. J.; MISHRA, S. N. *Modern Mathematical Statistics*. New York : John Wiley and Sons, 1988.
- HÁTLE, J.; LIKEŠ, J. *Základy počtu pravděpodobnosti a matematické statistiky*. Praha : SNTL; Alfa, 1972.
- HOGG, R. V.; CRAIG, A. T. *Introduction to Mathematical Statistics*. New York : Macmillan Publishing, 1970.
- JACKMAN, S. *Bayesian Analysis for the Social Sciences*. Chichester : John Wiley and Sons, 2010.
- JUDGE, G. G. et al. *Introduction to the Theory and Practice of Econometrics*. New York : John Wiley and Sons, 1982.
- KOLMOGOROV, A. N. *Information Theory and Statistics*. New York : Chelsea, 1956.

- LEAMER, E. E. *Specification Searches: Ad Hoc Inference with Nonexperimental Data*. New York : John Wiley and Sons, 1978.
- LINDLEY, D. V.; PHILLIPS, L. D. Inference for a Bernoulli Process (a Bayesian view). *The American Statistician*. 1976, vol. 30, s. 112–119, 1976.
- LINDLEY, D. V. *Bayesian Inference*. *Encyclopedia of Statistical Sciences*, 1. New York : John Wiley and Sons, 1982.
- PRESS, S. J. *Subjective and Objective Bayesian statistics*. New York : John Wiley and Sons, 2003.
- RAO, C. R. *Lineární metody statistické indukce a jejich aplikace*. [Překlad do češtiny]. Praha : Academia, 1978.
- SAMANIEGO, F. J. *A Comparison of the Bayesian and Frequentist Approaches to Estimation*. New York : Springer, 2010.
- SPEIGEL, M. R. *Theory and Problems of Advanced Calculus*. New York : Schaum Publishing, 1963.
- WANG, C. *Sense and Nonsense of Statistical Inference*. New York : Marcel Dekker, 1993.
- ZELLNER, A. *An Introduction to Bayesian Inference in Econometrics*. New York : John Wiley and Sons, 1971.
- ZVÁRA, K.; ŠTĚPÁN, J. *Pravděpodobnost a matematická statistika*. Praha : Matfyzpress, 1997.

A COMPARISON OF CLASSICAL AND BAYESIAN PROBABILITY AND STATISTICS (2)

Abstract: Statistics has been developing for almost 250 years – since the publication of an essay which included one theorem called Bayes' after the author. This whole period (since 1763 to this day) has been accompanied by a duel between the supporters of a subjective concept of probability and those who refuse everything but a purely objective concept of probability as well as statistics. While the 18th and 19th centuries accepted the importance of the subjective (let us say Bayesian) way of thinking for the development of probability and statistics without a problem, in the 20th century the classic (frequentist) way took over and has been dominant in teaching and textbooks to this day. Only in the second half of the 20th century did the situation begin to change slowly. Reasons for that are partly described in the present article, but arguments and simple examples supporting the Bayesian way in comparison with the classic one are clear and generally respected worldwide. Unsuspected new computing possibilities have caused an explosive development of Bayesian statistics, which has infiltrated almost all the areas of statistics and a number of other scientific fields. It is not possible to expect a retreat of the different philosophical or pedagogical positions of the fighting schools of thought (even though it is really needed), but the use of advantages of both the approaches is methodologically not only possible, but even expected. Part of the teaching of statistics must be prepared for these changes, but it has not been the case in the Czech Republic at all so far.

Keywords: subjective probability; frequentist statistics; classical and Bayesian approach and thinking; Bayes' theorem; point estimation; prior and posterior distribution; Bayesian Credible interval; hypothesis testing

JEL Classification: E21, C82

(Dokončení stati v čísle 3/2012 AOP.)