SROVNÁNÍ KLASICKÉ A BAYESOVSKÉ PRAVDĚPODOBNOSTI A STATISTIKY (1.)

Petr Hebák*

Úvod

Hlavním cílem tohoto příspěvku je upozornit výzkumníky různých vědních oborů, že bayesovská pravděpodobnost a statistika je mnohem více než všeobecně známý Bayesův vzorec a jeho občasné použití v ukázkových či ilustrativně orientovaných příkladech při výkladu operací s pravděpodobnostmi náhodných jevů. Zobecňujícími úsudky se ve statistice běžně rozumí bodové odhady a intervalové odhady (jednorozměrných či vícerozměrných) populačních *parametrů* či *charakteristik* na základě náhodných výběrů a testy hypotéz nejrůznějšího typu a zaměření. Mnohdy mezi autory článků v nestatisticky orientovaných časopisech existuje až fetišizovaná představa, že odborný příspěvek neobstojí bez statistického *testu významnosti*. Úvodem je proto třeba varovat nebo aspoň připomenout odborné veřejnosti skutečnost, že nejen *mechanické používání* nástrojů a terminologie z oblasti statistických úsudků, ale **samotné** tyto **statistické nástroje** jsou oprávněně odbornou veřejností z různých hledisek už *dlouhodobě* kritizovány a zpochybňovány.

Za zmínku stojí výborná kniha *Chamonta Wanga*, která pod názvem *Sense and Nonsence of Statistical Inference* byla napsána před 18 lety. Je soustředěna na rostoucí omyly, chybná použití nebo nedorozumění především v oblastech *testů významnosti*, *statistického zobecňování*, *statistické kauzality* a *subjektivních úsudků*. Wang v úvodu říká, že *mnoho let se cítil frustrovaný a podvedený při aplikacích tzv. statistických modelů a způsobem používání statistických testů v oblasti neexperimentálních dat*. Kniha Wanga ani zdaleka nechce být jen kritikou statistiky. Částečně je naopak projevem vděčnosti za ty přístupy k používání statistických úsudků (zaznamenal jsem, že mezi ně počítá i *bayesovskou* analýzu), které podle jeho slov dokreslují, že *statistika je nekonečně bohatá ve svých jemnostech a skrytých krásách*. Přes tato vznešená slova se text zmíněné knihy převážně soustřeďuje na chybná použití nebo dokonce zneužití statistických úsudků, protože ty převládají a mají za následek, že statistika se už dlouhou dobu dostává do slepé uličky.

Jiným dokladem skutečnosti, že nedostatky klasického přístupu v oblasti modelování a statistických úsudků jsou dlouhodobé a zásadní, je dodnes poměrně

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

často citovaná kniha *Edwarda E. Leamera*: *Ad Hoc Inference with Nonexperimental Data*, kterou vydalo prestižní nakladatelství *Wiley* už v roce 1978, tedy o 15 let dříve než vyšla kniha Wanga. Přečtení knihy Leamera (na začátku osmdesátých let v *ruském* překladu, když anglický originál byl nedostupný) pro mne znamenalo zásadní změnu postoje. Změna se týkala nejen tradičních statistických úsudků a často diskutovaných definicí pravděpodobnosti. Důvodem byl hlavně způsob *modelování ekonomických jevů a procesů pomocí stochastických modelů* regresního a příbuzného ekonometrického typu, jež jsou převážně používány bez dostatečné ekonomické teorie a na základě nepříliš kvalitních ekonomických dat, která většinově k náhodným výběrům mají hodně daleko.

Leamer (1978) se v této (spíše metastatisticky orientované) knize od začátku netají svým (na studiích získaným) bayesovským přesvědčením. Leamer nejméně před 33 roky poznal přednosti bayesovské (*subjektivní*) interpretace pravděpodobnosti, jakož i nezanedbatelné výhody bayesovského způsobu přístupu ke statistickým úsudkům a k modelování. Později však, když začal pracovat s ekonomickými daty ve výzkumu na Michiganské univerzitě, sice neztratil důvěru k bayesovskému způsobu myšlení, ale trochu modifikoval svoje názory. Došel k závěru, že bayesovská statistika je sice lepší než klasická, ale také nevyléčí všechny neduhy analýzy statistických dat a statistického modelování vztahů mezi ekonomickými veličinami. Podle Leamera mají ekonomové zcela obecně lhostejný až *cynický* vztah k práci s reálnými daty. Podle něj o důkladnou analýzu dat vlastně ani nemají velký zájem. Tyto úvahy se sice týkají už poměrně vzdálené doby, ale podle mého názoru i současných zkušeností s ním lze souhlasit, jakož i s jeho zajímavou *metodou Sherlocka Holmese*, jejíž podstatu v knize popsal a vysvětlil.

Tradiční (říkejme *klasický* nebo *četnostní*) přístup dominoval téměř celé minulé století, takže pořád v aplikacích a ve výuce pravděpodobnosti a statistiky výrazně převládá. Přednosti bayesovského přístupu jsou zřetelné a uznávané a současná konjunktura bayesovské statistiky byla některými předními světovými statistiky předpovídána a očekávána. Pokusil jsem se v tomto krátkém příspěvku na velmi jednoduchých situacích a příkladech ukázat hlavní rozdíly mezi tradičním a bayesovským **způsobem uvažování** a stručně popsat metodologická východiska a odlišnosti obou těchto přístupů. Nelze očekávat, že se mi podaří více než jen naznačit některé myšlenky a argumenty bayesovců. Chtěl bych hlavně varovat před současnými negativními zvyklostmi některých *výzkumníků*, zvláště ve společenských vědách, ale zároveň upozornit na potřebu zásadních změn ve způsobu výuky pravděpodobnosti a statistiky obecně a statistických úsudků speciálně. Nejde o jakousi *zjednodušenou* náhradu klasické metodologie jinou metodologií, ale o postupný krok k racionálnímu spojení předností klasické a bayesovské školy s tím, že v budoucnosti (v době stále více a více převažujících velice rozsáhlých datových souborů) ani toto metodologické propojení nemusí stačit.

První oddíl je nezbytným pohledem do historie a druhý oddíl porovnává *objektivní* vnímání pravděpodobnosti v klasickém pojetí se *subjektivním* chápáním pravděpodobnosti v bayesovském pojetí. Třetí oddíl se týká Bayesova vzorce; jeho úkolem je na jednoduchých příkladech ukázat a vysvětlit, že kromě jednoduchého a všeobecně známého zdůvodnění formální správnosti tohoto vzorce existují i důvody, kvůli kterým se později stal základem bayesovské statistiky. Ve čtvrtém a v pátém oddílu jsou popsány přednosti bayesovských bodových a intervalových odhadů ve srovnání s klasickou teorií odhadu při stále ještě převládajících tradičních zvyklostech. Poslední šestý oddíl se podobným způsobem zabývá testováním statistických hypotéz. Patrně by bylo dobré zmínit, že některé části jsou obsahem pokračování článku.

Závěrečné shrnutí naznačuje i některé směry současného vývoje, společně s několika poznámkami o zaměření současné klasické a bayesovské statistiky, a je zde i velice stručná zmínka o bayesovských výpočtech. Vzhledem k již zmíněným rozsáhlým datovým souborům je totiž zřejmé, že pomyslným přijetím jakési *čisté* bayesovské statistiky vývoj statistické metodologie zdaleka neskončí. Přednosti různých přístupů budou nejen dále zkoumány a kombinovány, ale nelze vyloučit ani zcela jiný budoucí charakter statistických úsudků než lze v současnosti předvídat.

Je škoda, že se pro nedostatek místa v tomto příspěvku důkladněji nezabývám teorií souvisící s bayesovskými výpočty, jejíž pokrok zásadně přispěl ke zvýšenému zájmu o bayesovskou statistiku. Vyhýbám se i rozhodovací teorii, ve které bayesovský způsob myšlení dlouhodobě dominuje, a nezmiňuji ani žádné ukázky z rozsáhlých bayesovských aplikací v některých oblastech statistiky, které zasahují do řady vědních oborů. Snažím se aspoň naznačit význam a závažnost současných změn ve statistické metodologii, ale nedostanu se příliš přes začátek, protože knihy stejného nebo podobného zaměření jako má tento článek, nejsou vzácností, a proto jich i několik uvádím v seznamu použité literatury. Studenti i absolventi společenskovědních vysokých škol by měli aspoň dostat příležitost se hlouběji seznámit s přednostmi bayesovského způsobu myšlení, a proto je nezbytné, aby budoucí a snad i současní vyučující (nejen statistiky) samostatně nebo v rámci povinných či volitelných předmětů studovali bohatou literaturu. Na rozdíl od mnohých vyspělých zahraničních univerzit přetrvává v České republice tradiční výuka statických úsudků bez jediné změny či aspoň zmínky o existenci bayesovské statistiky. Přidáme-li k tomu celkově převažující nízkou úroveň kvantitativního myšlení mnohých absolventů středních a vysokých škol, nelze se ani divit současnému stavu. Zdánlivě paradoxně s tím souvisí relativně často v odborných časopisech nebo knihách publikované nepříliš kvalitní modely různého typu s pochybnými interpretacemi. Často se objevují teorií nedostatečně zdůvodněné a analýzou dat nepodložené lineární regresní modely s několika desítkami vysvětlujících proměnných všeho typu, které rovněž potvrzují pravdivost slov Leamera (1978) o *cynickém přístupu ekonomů k analýze dat* i názory Wanga (1993), který v této souvislosti říká, že *...the practice of statistical inference has drifted into never-never land for too long"*.

1. Krátký pohled do historie

Odmysleme si aspoň pro začátek úvah dlouhodobý *souboj* mezi *subjektivisty* a *objektivisty*, který trval více než dvě století, a respektujme současný stav. Myslím, že kritika bayesovského způsobu myšlení ve *světě středoškolských nebo vysokoškolských pedagogů* je dnes spíše už jen *umíněností* některých vyučujících i výzkumníků, vyplývající z částečné nebo úplné *neznalostí argumentů* bayesovců, a někdy dokonce jen *jakoby nadřazenou neochotou* některých *klasiků* respektovat vývoj, ke kterému nesporně došlo. Bayesovské nástroje definitivně přestaly být v odborné literatuře kritizovány, ale přesto bayesovské *pojetí pravděpodobnosti* (např. skutečnost, že neznámé parametry pravděpodobnostních rozdělení jsou bayesovci vnímány jako náhodné veličiny s pravděpodobnostním rozdělením), jsou stále ještě *trnem v oku* statistiků vychovaných výhradně v zajetí *objektivity* a *dokonalých* klasických *postupů*. Ještě dnes se však v soukromých debatách či polemikách setkávám s výroky typu: *Je to možná zajímavé, ale já se zabývám něčím jiným*; *příliš tomu nevěř*ím; *je to složité a výpočetně náročné*; *celé je to metastatistika*; nebo s podobnými dalšími spíše jen *jako* kritickými poznámkami přímo namířenými proti bayesovskému přístupu. Poznamenejme, že poslední výrok přirovnávající bayesovskou statistiku k metastatistice je v zásadě oprávněný (viz například Leamer, 1978), ale svědčí o naprostém nepochopení, měl-li by být používán jako negativní argument nebo hanlivé označení odlišného přístupu.

Bradley Efron (2010) ve stručném úvodu (volně přeloženo P. Hebákem – dále PH) říká, že ... *za cenu drastických zjednodušení by rozdělil historii statistiky jako uznávané disciplíny na tři období. Prvním je doba Queteleta a jeho následovníků, kdy rozsáhlá data ze sčítání lidu bylo třeba využít k zodpovězení jednoduchých, ale důležitých otázek typu*: *Rodí se více chlapců než dívek? Druhým je klasické období Pearsona, Fishera, Neymana, Hotellinga a jejich následovníků, intelektuálních géniů, kteří rozvinuli teorii optimálních úsudků do podoby, která dovoluje vyždímat každý kousek informace… Otázky stále směřovaly k jednoduchosti typu, zda postup A je lepší než postup B. Vhodnost nových metod však byla posuzována na základě malých datových souborů, které jednotliví vědci mohli shromáždit. Třetí je doba hromadné produkce, kdy nové technologie dovolují jednomu týmu vědců vytvořit soubory takového rozsahu, že by Quetelet mohl jen závidět. Tuto záplavu dat však doprovází povodeň otázek; možná tisíce odhadů či testů hypotéz, takže statistik má nálož k současnému zodpovězení mnohem větší než si klasičtí mistři mohli představit*.

Vrátíme-li se pro lepší představu o budoucnosti statistiky zpátky o 250 let na začátek debat mezi *bayesovci* a *klasiky*, lze zjednodušeně konstatovat, že druhá polovina 18. století a téměř celé 19. století bylo z hlediska statistické metodologie spíše *bayesovsky* (nebo možná lépe řečeno) převážně *subjektivně* orientované. Každopádně na začátku tohoto období byli *Thomas Bayes* (1701 až 1761) a *Pierre Simon Laplace* (1749 až 1827), ale ti měli mnoho význačných pokračovatelů. Konec 19. století a téměř celé 20. století byla nesporně doba velkého rozvoje klasické statistické školy, nebo (jinak vyjádřeno) doba *objektivistického* vymezení pravděpodobnostních a statistických úsudků. Je zcela nesporné, že tento směr z hlediska statistické metodologie jednoznačně a téměř výlučně dominoval. Za všechny, kteří se o to zasloužili, stačí jmenovat takové osobnosti, jako byli *Karl Pearson* a jeho syn *Egon Pearson*, *Jerzy Neyman* nebo *William Sealy Gosset* (Student) a především (podle některých geniální) *Sir Ronald Alymer Fisher*. Přínos R. Fishera k rozvoji statistiky a ke statistickému myšlení je všeobecně uznávaný a nesporný, ale také nelze z řady materiálů přehlédnout, že R. Fisher byl odpůrcem tzv. *neobayesovců* a odmítal jiný (než jeho) objektivistický způsob uvažování.

Asi v polovině dvacátého století se přesto začínají pomalu probouzet původní *bayesovské* myšlenky a hlavně se k nim přidávají mnohé nové. Následuje nejméně dvacet let bouřlivých (někdy až příliš polemicky agresívních) diskusí o (ne)oprávněnosti tzv. neobayesiánství. Jména jako *Bruno de Finetti*, *Leonard Jimmy Savage*, *Arnold Zellner*, *Harold Jeffreys*, *Jose Miguel Bernardo*, *Dennis Lindley*, *Morris H. DeGroot* a další připomínají období, kdy bayesovská škola postupně sílí. Především tito statistici, společně s dalšími nejmenovanými, se zasloužili o současnou pozici bayesovské školy. Za symbolicky zlomový bych si dovolil označit rok 1994, ve kterém známý statistik Sir *Maurice Kendall*, vydavatel všeobecně uznávané statistické řady publikací se zastřešujícím označením *The Advance Theory of Statistics,* s poměrně velkým zpožděním (ale přece jen) zařadil do své série monografií díl 2b pod názvem *Bayesian Inference*, jehož autorem je *Anthony O'Hagan*. Dnes nejspíše neexistuje žádný významný statistický časopis, který by pravidelně nezařazoval nejrůzněji zaměřené články k bayesovské problematice a výjimkou nejsou případy, kdy celá čísla prestižních časopisů jsou věnována bayesovské statistice.

Devadesátá léta (asi logicky) zároveň představují přibližný začátek rozsáhlých polemik o přednostech různých **způsobů výuky statistiky** (a to nejen pro nestatistiky) zaměřených na možnost vyučovat pravděpodobnost a statistiku výhradně z bayesovského hlediska. Postupně začaly vycházet vysokoškolské i středoškolské učebnice s východisky založenými na bayesovském způsobu uvažování. Takové publikace sice existovaly už v šedesátých letech (např. učebnice *D. Blackwella*: *Basic Statistics* z roku 1969), ale byly to tehdy spíše jen ojedinělé pokusy. Ústředním cílem příznivců bayesovského způsobu myšlení v debatách vhodném způsobu výuky pravděpodobnosti a statistiky bylo (a stále ještě částečně je) vysvětlit důvody subjektivní interpretace pravděpodobnosti a přesvědčit uživatele, že klasická statistika se při odhadech parametrů a testech hypotéz, na rozdíl od bayesovského přístupu, neopírá o pravděpodobnosti **vycházející z daného konkrétního výběru**, ale o hůře představitelnou situaci **všech možných výběrů z dané populace**. Bayesovci tvrdí, že pojmy *opakované výběry, výběrové rozdělení* a na nich založené statistické úsudky (především bodové odhady, intervaly spolehlivosti a testy hypotéz) jsou často nejen pro začátečníka mnohem méně pochopitelné a představitelné než aparát a pojmy vycházející z bayesovského pojetí pravděpodobnosti a statistických úsudků. Pro rozhodnutí, **jak učit teorii statistických úsudků**, by podle bayesovců měla být rozhodující skutečnost, že bayesovské postupy založené na charakteristikách posteriorního rozdělení jsou většinou přesnější, jednodušší, a tedy i pro začátečníky pochopitelnější než klasické postupy. *Asymptotické* vlastnosti výběrových charakteristik sice při velkých výběrech z výpočetního hlediska eliminují zásadní rozdíly mezi oběma přístupy, ale interpretační potíže a metodologické nejasnosti s klasickými odhady a testy přetrvávají, což si částečně ukážeme i v následujících oddílech.

Většina statistiků preferujících tradiční přístup považuje za hlavní rozdíl *objektivisticky* četnostní (klasickou, tradiční) interpretaci proti *subjektivistické* bayesovské interpretaci pravděpodobnosti. Tento rozdíl skutečně existuje a je zásadní, i když z hlediska důsledků pro způsob uvažování a formu provádění statistických úsudků není zdaleka jediný. Přesto právě tato otázka je častým námětem filozoficky orientovaných debat. Inspirujícím důkazem významu subjektivity vědců ve výzkumu je kniha, kterou společně napsali *S. James Press* a *Judith M. Tanur* (2001). Dvanáct v této knize uvažovaných význačných osobností (Aristotelés, Galileo Galilei, Viliam Harvey, Isaac Newton, Antoine Lavoisier, Alexander von Humboldt, Michael Faraday, Charles Darwin, Louis Pasteur, Sigmund Freud, Marie Curie a Albert Einstein), všeobecně považovaných za *ikony vědy*, přesvědčivě svým přístupem demonstruje potřebu i schopnost vědeckého využití osobního přesvědčení, intuice a předchozích i jiných znalostí. Tito význační vědci a odborníci ve své profesi jednoznačně prokázali potřebu a význam subjektivity pro získání nových poznatků, jakož i možnost její kombinace s empirickými výsledky bádání. Svým chováním představují skvělý příklad bayesovského myšlení, i když sami o Bayesově vzorci asi nikdy nic neslyšeli.

Závažná námitka proti používání bayesovské statistiky, týkající se velké výpočetní náročnosti při práci se složitějšími (hlavně víceparametrickými) modely, padla v posledním desetiletí minulého století. Bayesovci za to vděčí úspěšné možnosti využívání simulací založených na metodě Monte Carlo a na Markovských řetězcích. Možnost provádět výběry přímo z posteriorního rozdělení vedla k pochopitelné explozi zájmu o bayesovské statistické úsudky, které tím v různých podobách a variantách postupně pronikly a pronikají do téměř všech vědních oborů a téměř do všech oblastí statistiky.

Obrovský zájem o bayesovskou statistiku a její přístup v teorii a i v aplikacích potvrzují i některé následující údaje. *James O. Berger* (JASA, 2000) v článku *Bayesian Analysis* a *Look at Today and Thoughts of Tomorrow* uvádí, že zatímco za 200 let od roku 1769 do roku 1969 celkově vyšlo jen 15 bayesovsky zaměřených knih, za období 1970 až 1989 jich vyšlo už dalších 30 a v dalším desetiletí od roku 1990 do roku 1999 se počet vytvořených bayesovsky orientovaných knih zvýšil přibližně o dalších 60. Podobný vývoj (při přibližném dvojnásobném růstu každých deset let) uvádí Berger i u počtu publikovaných článků. Zmiňuje desítky organizací či konferencí, příklady z mnoha internetových stránek a odkazů na aplikace v nejméně patnácti vědních oborech téměř ze všech oblastí statistiky. Podobný vývoj stále pokračuje, což jen dokresluje zásadní změnu postoje k bayesovskému přístupu i způsobu myšlení. Jen pro zajímavost vyhledavač Google k dnešnímu dni (14. 9. 2011) uvádí přibližně 10,8 mil. odkazů na jméno *T. Bayes* a dále 1,65 mil. odkazů na bayesovské *konference*, 2,73 mil. odkazů na bayesovský *software* a 2,88 mil. odkazů na bayesovsky orientované *články*.

2. Rozdíl mezi četnostní a bayesovskou pravděpodobností

Formálně matematicky je pravděpodobnost jednoznačně určena axiómy, které *Kolmogorov* (1933) vytvořil s tím, že budou sloužit jako konvenční a moderní matematická definice pravděpodobnosti. Při vynechání podrobností týkajících se teorie míry a σ-algebry, které zájemce může najít u více matematicky orientovaných textů, např. u *Anděla* (1998, s. 23) nebo u *Zváry, Štěpána* (1997, s. 21 až 25) či u *Jackmana* (2010, s. 497) lze definici Kolmogorova v souladu s *Jackmanem* (2010, s. 3 až 4) zjednodušeně reprodukovat takto: Je-li Ω neprázdná množina jevů a $P(A)$ je funkce, která přiřazuje reálná čísla jevům A, pak $P(A)$ je mírou pravděpodobnosti, jestliže

1. $P(A) \geq 0,\ \forall A \subset \Omega$ (pravděpodobnosti jsou nezáporné),
2. $P(\Omega) = 1$ (pravděpodobnost jistého jevu je jedna),
3. Pro neslučitelné jevy A a B platí $P(A \cup B) = P(A)+P(B)$.

Axiómem Kolmogorova se řídí každá statistická škola, a tedy nejen klasická, ale i bayesovská. Avšak uvědomíme-li si, že Kolmogorov (1956, s. 1) uvedl (překlad PH), že

...teorie pravděpodobnosti jako matematická disciplína může a měla by být rozvíjena z axiómů přesně stejným způsobem jako geometrie a algebra. To znamená, že jakmile máme definovány prvky, které jsou předmětem studia a jejich hlavní vztahy, a máme stanoveny axióm, kterými jsou tyto vztahy ovládány, pak každý další výklad musí být založen výhradně na těchto axiómech, nezávisle na běžném konkrétním významu těchto prvků a jejich vztahů.

Z těchto slov (stále podle *Jackmana* 2010, s. 5) jednoznačně vyplývá, že otázka interpretace pravděpodobnosti Kolmogorova vůbec nezajímala, takže všechny statistické školy vycházející z této definice mají naprosto otevřený prostor k vlastní interpretaci. Kdokoli se však snaží v jakékoli situaci pravděpodobnosti interpretovat a využívat může zjistit, že Komogorovova definice pravděpodobnosti je pro něj naprosto nedostačující. Nejenže podle ní nejde žádnou pravděpodobnost určit ani vypočítat, ale navíc může dojít až k situacím, ve kterých všechny podmínky definice jsou splněny, ale nikdo by v nich slovo *pravděpodobnost* v obvyklém významu nepoužil. Všeobecně známý a často v této souvislosti citovaný je příklad *Leamera* (1978, s. 24), ve kterém uvádí, že lidská ruka představuje 10 % váhy těla, ale nikoho by přesto nenapadlo tvrdit, že ruka má pravděpodobnost 0,1.

Četnostní definice pravděpodobnosti jevu *A* (používá se označení *statistická* definice pravděpodobnosti) vychází z předpokladu *nezávislých náhodných pokusů opakovaných za stejných podmínek* s tím, že při velkém počtu pokusů (teoreticky neomezeném) lze *relativní četnost jevu A* považovat za odhad pravděpodobnosti jevu *A*. Vychází se z toho, že dochází-li s růstem počtu pokusů ke *stabilizaci* relativní četnosti jevu *A* kolem nějakého čísla, pak lze toto číslo považovat za odhad pravděpodobnosti jevu *A*. Tento odhad je z hlediska přesnosti tím lepší, čím větší je počet provedených pokusů. Oporou pro tento přístup je i dílčí věta zákona velkých čísel *Jacoba Bernoulliho* (1654 až 1705), ze které vyplývá, že při dlouhodobě nezměněných podmínkách nezávislých pokusů konverguje relativní četnost jevu A podle pravděpodobnosti k pravděpodobnosti tohoto jevu. Tímto způsobem (necháme-li stranou případ *konečného počtu stejně pravděpodobných elementárních jevů*) klasická (*četnostní*) statistika definuje pravděpodobnost všech výsledků *za stejných podmínek opakovatelných* náhodných pokusů. Třeba výsledky herních situací, četnost výskytu vlastnosti lidí nebo zhotovených výrobků či vypěstovaných rostlin anebo četnosti demografických jevů, to vše a mnohé další podléhá stejnému principu výpočtu (= odhadu) pravděpodobnosti založenému na stálosti podmínek a stabilizaci dlouhodobé relativní četnosti.

Zdůrazňuje se, že je-li realita, ve které žijeme, nezávislá na našem vědomí, je popsaný postup i v tomto smyslu *objektivní*. Například *schopnost mince* padnout při jednom hodu *hlavou vzhůru* či *schopnost napínáčku* padnout při jednom hodu *špičkou dolů* jsou neznámá čísla, která je třeba (*objektivně*) odhadnout provedením dostatečného počtu nezávislých hodů opakovaných za stále stejných podmínek. Míry této schopnosti (pravděpodobnosti) jsou tedy neznámé konstanty, které neumíme změřit či jinak zjistit, takže v obou případech je to stejné a je jen třeba pro ně získat nějaký vhodný odhad. Klasická statistika tedy vnímá pravděpodobnost jevu *A* jako *objektivně danou vlastnost tohoto jevu*, bez ohledu na to, zda ji umíme či neumíme určit. V našem příkladu tedy bez ohledu na to, zda něco o mincích nebo napínáčkách víme či nevíme a zda k výsledkům provedených hodů máme stejný či zcela odlišný postoj. Sporná místa této definice komentujme otázkami a označme je otazníky: Při klasickém pojetí neexistuje žádná jiná rozumná možnost, jak pravděpodobnost určit, než provést za stejných podmínek (*je to fyzikálně možné*?) dostatečně velký počet (*nejlépe nekonečný*?) nezávislých (*jak to poznáme, když pojem nezávislosti vyžaduje znalost pravděpodobnosti*?) pokusů a ztotožnit výslednou relativní četnost (*pro jaký počet pokusů je to výsledná*?) s neznámou pravděpodobností (*je opravdu zcela neznámá a nic o ní nevíme*?). Například v případě napínáčku či mince opakovat hody tímto ideálním nebo běžným (*jak to poznáme*?) napínáčkem či mincí za víceméně stejných podmínek (*co se tím myslí*?). Příliš nám nepomůže případné zjištění, že ani při velkém počtu pokusů se relativní četnost nestabilizovala z důvodu nesplnění předpokladu nezávislých pokusů prováděných za stejných podmínek. Je jasné, že jedinečné pokusy a neopakovatelné jevy už jsou předem vyloučeny, takže v tomto pojetí u nich o žádné pravděpodobnosti ani nelze uvažovat. Otazníky jsem jen chtěl upozornit na slabiny četnostní definice pravděpodobnosti a na případnou obtížnost dodržet stanovená pravidla. Kromě toho, označíme-li náhodné pokusy za nezávislé v případě, že jejich výsledky neovlivňují pravděpodobnosti výsledků ostatních pokusů, pak posouzení nezávislosti předpokládá znalost pravděpodobností, ale i naopak určení pravděpodobností výsledků vyžaduje nezávislé pokusy. Rovněž představa, že se daný objekt v průběhu velkého počtu pokusů nemění, je sice možná, ale přinejmenším optimistická.

Jeffreys (1957) formuloval základní pravidla induktivních úsudků, která jsou uvedena v knize *Zellnera* (1971, s. 7) a v jednom z těchto pravidel říká, že *jakákoli definice pravděpodobnosti založená na nekonečně velkém počtu pozorování nebo pokusů je bezcenná, protože se opírá o nemožný experiment, když nekonečně velký počet pokusů nelze provést.*

Četnostní definice pravděpodobnosti předpokládá libovolně velký počet *opakovatelných* nezávislých náhodných pokusů, a tedy se týká výhradně *hromadných* náhodných jevů. Tím jsou přinejmenším značně sníženy možnosti využití, protože jedinečné (neopakovatelné) jevy či pokusy jsou touto definicí pravděpodobnosti zcela vyloučeny.

Zamysleme se nad následujícími otázkami: *Jaká je pravděpodobnost atomového útoku teroristů na Evropu ještě v tomto desetiletí*? *Jaká je pravděpodobnost, že do tří let zcela skončí ekonomická krize v zemích Evropské unie*? *Jaká je pravděpodobnost, že moje znalosti získané studiem na školách, které jsem absolvoval, byly dostatečné pro můj další odborný růst*? Ve všech případech jediná (!) korektní odpověď zastánce četnostní definice pravděpodobnosti je 0 nebo 1 bez ohledu na to, zda správnou odpověď zná či nezná. Takové otázky jsou totiž pro četnostního statistika zcela nepřijatelné. V horším případě je jednoduše odmítne se zdůvodněním, že nepatří do počtu pravděpodobnosti, a v lepším případě odpoví podobnými slovy jako *Blais Pascal* fiktivního dialogu knihy *Alfreda Renyiho*: *Dialogy o matematice*, která vyšla v českém překladu v roce 1967. Na námitku oponenta, že v jedinečných případech nelze stanovit relativní četnost, Pascal údajně reagoval: *Mohu a nemohu. Mohu provést jedno pozorování. Právě jen jedno. Hodnota relativní četnosti bude 1 nebo 0. Nic jiného…* Komentář asi není zapotřebí, protože tato odpověď diskutovanou situaci neřeší.

Uvedené a mnohé jiné otázky jsou závažné nebo aspoň zajímavé, ale každopádně vyžadují lepší odpověď. Ve zcela stejné situaci jsme u mnohem jednodušší otázky: *Jaká je pravděpodobnost, že* 13. července 2011 právě *u naší chaty na Vysočině byla v noci bouřka s nebezpečnými blesky, když v Praze bouřka s blesky řádila téměř dvě hodiny*? Otázka se týká minulosti, takže v četnostním pojetí je zřejmá správnost opět jenom jedna ze dvou variant 0 nebo 1 s tím, že my to buď víme, nebo nevíme. Osobně si jsem téměř jistý, že i někde na Vysočině byla aspoň jedna nebezpečná bouřka a telefonickým dotazem u naší pani sousedky by se dalo zjistit, zda byla i u naší chaty. Opět tedy můžeme otázku zamítnout jako nevhodnou pro určení pravděpodobnosti, ale příliš to nepomůže, i když to pro mne je či mohla být velice důležitá otázka. Určitě si dovedeme představit, že by nás velikost pravděpodobnosti mohla zajímat, a to nejen v situacích, ve kterých by šlo o určení výše sázky. Tím jsme se nepřímo dostali k bayesovské pravděpodobnosti, protože v daných situacích žádnou jinou možnost nemáme, a jak uvidíme i pro opakovatelné nebo dokonce libovolněkrát opakovatelné pokusy může být lepší možností, než je použití četnostní definice.

Bayesovská definice pravděpodobnosti je skutečně **subjektivní** a představuje **stupeň přesvědčení** jednotlivce (běžně označovaný za *degree of belief*), že daný jev nebo daná situace nastane. Jiná osoba může mít jiný názor, a tedy i jinou pravděpodobnost. Tuto myšlenku formuloval už zmíněný *Jacob Bernoulli* ve své knize *Ars Conjectandi* (Umění předvídat), která vyšla po jeho smrti v roce 1713. Například mohu říci při hodu mincí, že pro mne je vše (mince i způsob házení) zcela v souladu s pravidly hry, takže moje pravděpodobnost je 0,5. Někomu jinému se mince jeví jako lehce asymetrická a může se rozhodnout, že jeho pravděpodobnost je 0,45. Moje pravděpodobnost jevu *A* v bayesovském pojetí *není vlastností tohoto jevu*, ale pouze *můj* postoj k této vlastnosti. Například *moje* pravděpodobnost schopnosti tohoto napínáčku padnout špičkou dolů je můj postoj k této schopnosti.

Subjektivní vnímání pravděpodobnosti ve smyslu *degrese of belief* v žádném případě neznamená odmítnutí pravděpodobnostních **axiómů** nebo nerespektování vlastností pravděpodobnosti či nedodržování pravidel pro operace s pravděpodobnostmi, ze kterých vyplývá i Bayesův vzorec. Parametry pravděpodobnostních rozdělení jsou v klasické statistice označovány za známé či neznámé konstanty, zatímco v bayesovské statistice jsou parametry náhodné veličiny, jejichž pravděpodobnostní rozdělení je užitečným nástrojem. Nelze vyloučit, že přes všechny uvedené argumenty čtenáři připadá subjektivní definice pravděpodobnosti přece jen málo vědecká, takže je třeba dále argumentovat.

V souladu s *Hebák* (1999, s. 45) se nejdříve zamysleme, zda je **stupeň přesvědčení** (*degree of belief*) jednotlivce vhodný pro vědu nebo patří jinam (třeba dokonce do náboženství). Vědci zcela běžně přijímají nebo diskutují odlišné teorie o některé oblasti světa, aniž by věděli, která teorie je správná. Někteří vědci mají vysoký stupeň důvěry v určité teorie. Dochází i k tomu, že existují dvě vylučující se teorie, a ačkoli jen jedna může být pravdivá, tak dva vědci věří právě jedné z nich. V knize *A Brief History of Time* (Bantam Books, New York 1988) její autor fyzik *S. W. Hawking* jednoznačně přiznává, že jeho chápání vědy je **subjektivní**. Používá výrazy jako *bylo všeobecně přijímáno; my nyní věříme; oni věřili; věříte-li* atd. Stupeň *víry* (*důvěry*, *přesvědčení*) v určitý výsledek záleží na konkrétní osobě a na jevu, kterého se tato víra týká.

Subjektivní interpretace pravděpodobnosti má několik výhod. První výhodou je, že je možné ji použít vždy, když osoba, které se problém týká, má **názor** (*přesvědčení*, *znalost*). Předpokládat absolutní nevědomost o posuzované otázce je velmi slabý předpoklad, i když samozřejmě tato situace může nastat a nesmíme ji vyloučit. Další výhodou je možnost **změny názoru** po získání nových informací. Právě tato možnost je velmi užitečná a pro bayesovskou statistiku zcela **typická a rozhodující**.

Už bylo řečeno, že i bayesovská statistická škola vyšla z axiómů Kolmogorovova, ale částečně je modifikovala. První formální zmínky o subjektivní pravděpodobnosti učinil *Ramsey* (1926), zásadním způsobem přispěli *Savage* (1954) a *de Finetti* (1974). Všichni tři přešli na **konečný** aditivní systém na rozdíl od **spočetného** aditivního systému Kolmogorova a uvažovali výlučně **subjektivní** interpretaci pravděpodobnosti. *Samaniego* (2010), *Jackman* (2009) a další uvádějí celou řadu variant axiómů subjektivní pravděpodobnosti. Shrnutí různých pohledů s následující diskuzí nabízí *Fishburn* (1986, Statistical Science, s. 335 až 358). Není sice známé (kromě výše uvedeného jeho citátu) jaký přesně byl postoj Kolmogorova k interpretaci pravděpodobnosti, ale vyjdeme-li z jeho konstatování, že *interpretace funkce P*(.) *nezávisí na jednotlivci*, pak dojdeme k názoru, že objektivní vnímání pravděpodobnosti mu bylo bližší. Výhody konečného aditivního systému vysvětluje právě Fishburn a oponenti citovaného výroku.

Pravděpodobnost založená na *stupni přesvědčení* je subjektivní ohodnocení nadějí, že daný jev nastane nebo nastal. Měření stupně důvěry jednotlivce v nastoupení nebo nenastoupení určitého jevu může být sporné, ale je třeba mít na mysli, že je to pouze první a výchozí krok. Při dotazech týkajících se pravděpodobnosti je též třeba mít nějakou škálu měření, stejně jako má každý jiný experiment. Pro *stupeň důvěry* může být touto škálou *kalibrační experiment*, což je přístup, při kterém se využívají *symetrické* náhodné pokusy, při kterých se předpokládá, že **konkrétní osoby** budou považovat všechny elementární jevy za *stejně* pravděpodobné (tak jak je to běžné v *klasické definici pravděpodobnosti*). Osoba, které se otázka týká, musí být tedy schopna si představit tento symetrický experiment se stejně pravděpodobnými výsledky. Po lepší představu vyjděme z jednoduchého a zcela srozumitelného příkladu, jehož myšlenku jsem převzal a upravil podle *D. A. Berryho* (1996, s. 122).

Příklad 1: Váha dospělých tučňáků

Rád bych znal Vaši pravděpodobnost jevu *A*, který spočívá v tom, že průměrný dospělý tučňák (samec) váží více než 25 kg. Nezdá se, že byste mi právě Vy mohl rovnou odpovědět, takže budu hádat. Pro některé čtenáře bude možná moje hádání nesprávné, jelikož o váze tučňáků vědí více než jiní a možná dokonce, že někdo z Vás je v této oblasti přímo odborník. Sledujte následující úvahu a pokusíme se společně postupně modifikovat získané výsledky a určit Vaši pravděpodobnost jevu *A*. Představme si misku, která obsahuje jednu zelenou a jednu červenou kuličku. Nabízím Vám možnost vyhrát tisíc korun. Můžete si vsadit, že náhodně vybraná kulička z misky bude zelená, nebo na jev *A*, který spočívá v tom, že průměrný dospělý tučňák (samec) váží více než 25 kg. Podle Vaší odpovědi bude měřena Vaše pravděpodobnost jevu *A*. Vsadíte-li na jev *A,* znamená to, že Vaše pravděpodobnost jevu *A* je nejméně 1/2, protože předpokládám, že pravděpodobnost zelené kuličky jste ocenili právě na 1/2. Pokud však vsadíte na zelenou kuličku, ohodnotím Vaši pravděpodobnost jevu *A* jako nejvýše 1/2. Připusťme však, že jste vsadili na jev *A*. V takovém případě Vám nabídnu novou možnost vybrat si mezi sázkou na jev *A* a sázkou na náhodné vybrání zelené kuličky z misky se třemi zelenými kuličkami a s jednou červenou. Dali-li jste teď přednost sázce na zelenou kuličku, pak Vaše pravděpodobnost jevu *A* je nejvýše 3/4. Pro upřesnění Vaší pravděpodobnosti jevu *A* Vám nabídnu sázku na tento jev nebo na zelenou kuličku náhodně vybranou ze třetí misky, ve které je pět zelených kuliček a tři červené kuličky. Rozhodnete-li se teď pro jev *A*, plyne z toho, že Vaše subjektivní pravděpodobnost tohoto jevu je aspoň 5/8 a nejvýše 3/4, tedy od 0,625 do 0,75. Takto bychom mohli pokračovat až do dosažení potřebné přesnosti týkající se (výhradně Vaší) pravděpodobnosti jevu *A*.

Uvedený postup je jednoduchý a velice snadno použitelný, ale přesto určité problémy při tomto způsobu *kalibrace* pravděpodobnosti nastat mohou. Především dotazovaný nemusí rozumět skutečnosti, že velikost nabídnuté částky nehraje při kalibraci žádnou roli, protože cílem postupu není nic jiného než jen zjištění jeho osobní pravděpodobnosti. Konkrétně v tomto umělém příkladu s tučňákem se můžeme podívat do encyklopedie a zjistit, že průměrný dospělý tučňák (samec) opravdu váží aspoň 25 kg a nejsou výjimečné případy, kdy váží až 40 kg, ale u jiných výroků či vážnějších úloh z běžného života možnost ověřit určité hypotézy či pravdivost výroků většinou nemáme. Lze namítnout, že při hypotetickém dotazování podobného typu záleží i na serióznosti odpovídajícího, což je jistě pravda, a je třeba to individuálně posuzovat. Skutečným problémem však je, že pro odpovídajícího mnohdy není snadné si vybrat jednu z možností a neví si rady, jakmile úloze úplně nerozumí anebo ji považují za příliš složitou.

Všechny používané subjektivní pravděpodobnosti musí být vnitřně *konzistentní* (říká se **koherentní**), čímž se myslí, že **neodporují jmenovaným axiómům, a tedy ani z nich vyplývajícím pravidlům pro počítání s pravděpodobnostmi**. Například říká-li někdo, že *jeho pravděpodobnosti* nezávislých jevů *A* a *B* jsou 0,8 a 0,5 a přitom tvrdí, že *jeho pravděpodobnost* současného nastoupení těchto dvou jevů je 0,2, pak tato osoba uvažuje *nekoherentně*, jelikož z věty o pravděpodobnosti průniku dvou nezávislých jevů vyplývá, že pravděpodobnost průniku těchto jevů *A* a *B* je 0,4. Chce-li někdo učinit sázku, která vychází z *nekoherentních* pravděpodobností, může se mu snadno stát, že mylně akceptuje takové uspořádání sázek, při kterém utrpí ztrátu vždy, bez ohledu na variantu, pro kterou se rozhodne. Takové sázce se říká *Dutch Book* a podle *Bernardo, Smith* (1994, s. 86, 88 a 95) *de Finetti* předložil (1937/1964) elegantní demonstraci koherentního stupně důvěry. Podle de Finettiho lze zjednodušeně říci, že neexistence nevýhodného uspořádání sázek (ve smyslu *Dutch Book*) je základním kritériem *koherence* subjektivních pravděpodobností. Myšlenky de Finettiho o koherenci významným způsobem zobecnili *Savage* (1971) a *Lindley* (1982). Požadavek a princip *koherence* úzce souvisí s racionálním rozhodováním, a tedy i s racionálními akcemi, následně tedy i s užitkem či ztrátami, které v důsledku těchto akcí nastávají. Vstup do této problematiky, která velice úzce souvisí s bayesovským rozhodováním, je sice potřebný, ale překročil by možnosti tohoto příspěvku. Zájemci o důkladnější studium bayesovské teorie lze doporučit už zmíněnou významnou monografii, kterou veřejnosti předložili pod názvem *Bayesian Theory José M. Bernardo* a *Adrian F. M. Smith* (1994).

3. Bayesův vzorec pro náhodné jevy a hypotézy

Bayesův vzorec je asi všeobecně známý, ale přesto si dovolím jej připomenout. Důvodem je nejen skutečnost, že při běžném výkladu počtu pravděpodobnosti se více zdůrazňuje technická nebo formální správnost vzorce a menší pozornost je věnována jeho použitelnosti a interpretaci, ale i potřeba na jednoduchých příkladech ukázat možnosti různých interpretací a předností bayesovského pravděpodobnostního myšlení. Klasicky orientovaný výklad většinově zůstává u jednoho typu příkladu týkajícího se náhodných jevů, vynechává rozšíření použitelnosti vzorce pro hodnocení vyslovených hypotéz o náhodných veličinách, nezmiňuje předpokládaný výlučně subjektivní charakter některých nebo všech uvažovaných pravděpodobností, a proto se i vyhýbá vysvětlení potřeby odlišení apriorních a posteriorních pravděpodobností. Potencionálnímu zájemci o důkladnější přístup tak téměř uniká význam Bayesova vzorce, protože o bayesovské statistice se nedozví nic nebo téměř nic, natož konkrétně o porovnání rozdílů a zdůraznění předností bayesovského přístupu ve srovnání s klasickým. Týká se to však nejen statistických úsudků, ale i teorie rozhodování, rozhodovacích stromů, problémů i teorií souvisících s rizikem, vícerozměrných metod a dalších oblastí.

Pro vysvětlení vzorce úplné pravděpodobnosti jevu A vyjděme z náhodného pokusu, při kterém může nastat tento jev jen ve spojení s právě jedním z úplné skupiny neslučitelných jevů $B_1, B_2, \dots , B_J$. Má-li tedy jev A nastat, musí být výsledkem pokusu právě jeden z jevů

$$A \cap B_1, A \cap B_2, \dots , A \cap B_J.$$

Jev A lze tedy zapsat jako sjednocení J průniků $A \cap B_j$, protože

$$A = (A \cap B_1) \cup (A \cap B_2) \cup \dots \cup (A \cap B_J).$$

Podle axiómu o pravděpodobnosti sjednocení neslučitelných jevů pravděpodobnost jevu A

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_J),$$

takže v kombinaci s větou o násobení pravděpodobností dostáváme následující *vzorec* (někdy též *zákon*) *úplné pravděpodobnosti*

$$P(A) = P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + ... + P(B_J)P(A \mid B_J) = \\ = \sum_{j=1}^{J} P(B_j)P(A \mid B_j). \quad (1)$$

An Essay Toward Solving a Problem in the Doctrine of Chances

Pod tímto názvem vyšla esej s *Bayesovým vzorcem* dva roky po jeho smrti v roce 1763. Samotný vzorec se subjektivní pravděpodobností sice souvisí, což je zvláště zřetelné z úvah o apriorních a posteriorních pravděpodobnostech, ale lze jej snadno odvodit i bez zmínky o nich, jen na základě vět o násobení a sčítání pravděpodobností. Bayesovu vzorci se také říká *zákon o inverzní pravděpodobnosti*, částečně asi i z toho důvodu, že ukazuje, jak lze po získání *nové informace* o nastoupení jevu A, transformovat podmíněné pravděpodobnosti jevu A vzhledem k jevům B_1, B_2, ... , B_J na podmíněné pravděpodobnosti jevů B_1, B_2, ... , B_J vzhledem k jevu A. Pojem inverzní pravděpodobnosti se především používá v souvislosti s úsudky o *neznámém* modelu na základě *známých dat*, na rozdíl od přímých pravděpodobností, které se naopak používají pro úsudky o *neznámých* skutečnostech na základě *známého* modelu. Toto odlišení přímých a inverzních pravděpodobností sice připomíná rozdíly mezi deduktivními a induktivními úsudky, ale na mne použití tohoto označení pro Bayesův vzorec působí dojmem snahy o zdůraznění typu úloh a situací, pro které je tento vzorec převážně určen.

Obecněji z věty o pravděpodobnosti průniku náhodných jevů ve tvaru

$$P(A \cap B_j) = P(B_j)P(A \mid B_j) = P(A)P(B_j \mid A)$$

je možné po drobné úpravě dojít k Bayesovu vzorci ve tvaru

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{P(A)} = \frac{P(B_i)P(A \mid B_i)}{\sum_{j=1}^{J} P(B_j)\, P(A \mid B_j)},$$

pro každé $i, j = 1, 2, ... , J$. (2)

Při úvahách o volbě vhodného příkladu pro ilustraci Bayesova vzorce, a s tím souvisícímu vysvětlení rozdílu mezi *apriorními* a *posteriorními* pravděpodobnostmi, jsem nakonec zvolil velice jednoduchou a srozumitelnou situaci.

Příklad 2: Pohádka o kuličkách s bayesovským prodavačem

Můj čtyřletý vnuk Lukáš viděl, že v jednom obchodě prodávají pytlíčky s deseti kuličkami, a přemluvil mne, abych mu jeden koupil. Vešli jsme do obchodu a požádali o jeden pytlíček s deseti kuličkami. Prodavač nám však řekl, že už má jen dva poslední, přičemž v jednom z nich je deset bílých kuliček, zatímco ve druhém je pět bílých a pět černých kuliček. Lukáš tvrdohlavě trval na tom, že musí mít ten s pěti bílými a s pěti černými, ale prodavač řekl, že neví, který pytlíček má jen samé bílé a který polovinu bílých a polovinu černých. Prodavač mi dále prozradil, že jeho zálibou je bayesovská statistika, a proto pro mne stanovil následující podmínky pro usnadnění volby, který pytlíček koupíme. Náhodně nám vybere jeden z nich a umožní nám z něj bez nahlédnutí dovnitř vybrat až tři kuličky. Čím méně jich vybereme, tím bude cena pytlíčku nižší, takže nejlevnější možností by bylo koupit jeden z pytlíčků a žádnou kuličku nevybírat. S pravděpodobností 0,5 bychom získali nejlevnější pytlíček, ale představa Lukášova křiku až by zjistil, že nemá ten pravý, mi nedovolila tuto možnost využít. Přijal jsem prodavačem náhodně zvolený pytlíček s kuličkami a vzal si papír a začal uvažovat:

Skutečnost, že mi prodavač vybral pytlíček se samými bílými kuličkami, jsem označil jako B_1, a skutečnost, že mi prodavač vybral pytlíček s pěti bílými a pěti černými jako B_2. Skutečnost, že první vybraná kulička bude bílá, jsem označil jako A_1 s tím, že jsem tajně doufal, že hned první kulička bude černá. Získal bych slevu a měl bych jistotu, že máme pytlíček s 5 bílými a 5 černými kuličkami.

Byl jsem si poměrně jistý, že prodavač dal oběma pytlíčkům stejnou pravděpodobnost, neboli $P(B_1) = P(B_2) = 1/2$. Víme, že $P(A_1 \mid B_1) = 1$ a $P(A_1 \mid B_2) = 5/10 = 1/2$, takže podle vzorce úplné pravděpodobnosti určím pravděpodobnost bílé kuličky v prvním tahu jako $P(A_1) = 1/2 \,.\, 1 + 1/2 \,.\, 1/2 = 3/4$.

Zatřepal jsem pytlíčkem a vybral jednu (*bílou*) kuličku (říkejme, že jev A_1 nastal). Podle Bayesova vzorce jsem vypočítal $P(B_1 \mid A_1) = (1/2)/(3/4) = 2/3$ a $P(B_2 \mid A_1) = (1/4)/(3/4) = 1/3$. Kdybych v tomto okamžiku řekl prodavači, že koupím ten druhý pytlíček, měl bych pravděpodobnost správného pytlíčku 2/3 = 0,67. Stálo by mne to sice méně, než když budu pokračovat, ale výraz Lukáše v obličeji, když viděl bílou kuličku, mi řekl, že pokračovat musím.

Stejným způsobem jsem se rozhodl vypočítat pravděpodobnost, že druhá kulička vybraná ze zbývajících devíti bude též bílá. Je jasné, že

$$\mathrm{P}(A_2 \mid A_1 \cap B_1) = 1 \text{ a } \mathrm{P}(A_2 \mid (A_1 \cap B_2) = 4/9.$$

V tomto okamžiku dochází k typické bayesovské situaci, že původně použité apriorní pravděpodobnosti $P(B_1) = P(B_2) = 1/2$ odporují mému současnému přesvědčení. Vypočítané posteriorní pravděpodobnosti $P(B_1 \mid A_1) = 2/3$ a $P(B_2 \mid A_1) = 1/3$ proto použiji jako **nové** apriorní pravděpodobnosti pro výpočet pravděpodobnosti druhé bílé kuličky. Znovu potom po dosazení do vzorce úplné pravděpodobnosti dostávám pravděpodobnost $P(A_2 \mid A_1) = 2/3 \,.\, 1 + 1/3 \,.\, 4/9 = 22/27 = 0{,}815$. Zatřepal jsem znovu pytlíčkem a vybral druhou (opět *bílou*) kuličku a stejným způsobem vypočítal posteriorní pravděpodobnosti $P(B_1 \mid (A_1 \cap A_2) = 9/11$ a $P(B_2 \mid (A_1 \cap A_2) = 2/11$. Kdybych v tomto okamžiku řekl prodavači, že koupím druhý pytlíček ještě bych trochu ušetřil a zabezpečil Lukášovi i sobě (vyšší) pravděpodobnost 9/11 = 0,818, že bude mít pytlíček s pěti bílými a pěti černými kuličkami. Naděje, že aspoň ta třetí vybraná kulička bude černá, už mi však nedovolila nedokončit nabídnutou možnost vybrat až tři kuličky. Pravděpodobnosti se opět změnily, takže zcela stejně

$$\mathrm{P}(A_3 \mid (A_1 \cap \mathrm{A}_2 \cap B_1) = 1 \quad \text{a} \quad \mathrm{P}(A_3 \mid (A_1 \cap A_2 \cap B_2) = 3/8.$$

Analogicky opět nové apriorní pravděpodobnosti jsou 9/11 a 2/11, takže pravděpodobnost bílé kuličky ve třetím tahu vychází 39/44 = 0,886. Naposledy jsem zatřepal pytlíčkem se zbývajícími osmi kuličkami a opět vytáhl *bílou* kuličku. Výpočty dokončuji a dostávám poslední posteriorní pravděpodobnosti

$$\mathrm{P}(B_1 \mid (A_1 \cap A_2 \cap A_3) = 12/13 = 0{,}923,$$

takže pro úplnost

$$P(B_2 \mid (A_1 \cap A_2 \cap A_3) = 1/13 = 0{,}077.$$

Ještě než jsem vrátil prodavači pytlíček se všemi deseti kuličkami, zaplatil a vzal si ten druhý pytlíček, ve kterém mělo být s pravděpodobností 0,923 pět bílých a pět černých kuliček, mne napadlo, zda by totéž vyšlo, kdybych byl rovnou náhodně bez vracení vybral tři kuličky a zjistil, že jsou všechny bílé. Označil jsem jako jev $A = A_1 \cap A_2 \cap A_3$ a s pomocí hypergeometrického rozdělení vypočítal, že pravděpodobnosti $P(A \mid B_1) = 1$ a $P(A_1 \mid B_2) = 10/120 = 1/12$, přičemž apriorní pravděpodobnosti obou pytlíčků zůstávají 1/2. Potom $P(A) = (1/2) \cdot 1 + (1/2) \cdot (1/12) = 13/24$, z čehož posteriorní pravděpodobnosti $P(B_1 \mid A) = 12/13 = 0{,}923$ a $P(B_2 \mid A) = 1/13 = 0{,}077$. S potěšením jsem konstatoval, že oba výsledky jsou shodné. Pohádka má šťastný konec, protože prodavač nám dal ten druhý pytlíček, který Lukáš hned otevřel a spokojeně zjistil, že v něm je pět bílých a pět černých kuliček. Jisté to nebylo, ale pravděpodobnost 0,923 zase nebyla tak malá.

Poznamenejme, že **postupné změny** apriorních a posteriorních pravděpodobností jevů B_1 a B_2 po každém novém zjištění nepřicházejí při klasickém pojetí v úvahu, protože by nebylo možné je rozumně interpretovat**.** Před provedením výběru pytlíčku by (za předpokladu nastoupení jevu A) klasický statistik (pomocí Bayesova vzorce) výsledek 0,923 samozřejmě také získal, ale po provedení výběru by měl mít potíže s jeho interpretací.

Bayesův vzorec je vhodný i pro případ, ve kterém $B_1, B_2, \dots, B_J$ jsou v postavení hypotéz a jev A v úloze dat či jiné podstatné informace. Pro odlišení od předchozího textu stačí změnit symboliku a jev B_j *vnímat* jako hypotézu H_j, j = 1, 2, …, J a analogicky jev A *vnímat* jako data nebo jako jinou podstatnou informaci (měnící stupeň důvěry ve vyslovené hypotézy) a značit D. Stále uvažujeme Bayesův vzorec, ale (kvůli interpretaci) s jiným značením ve tvaru

$$P(H_j \mid D) = P(D \mid H_j)P(H_j)/P(D), \; j = 1, 2, \dots, J,$$

kde

$$P(D) = \sum_{j=1}^{J} P(H_j)P\left(D \mid H_j\right), \tag{3}$$

který vyjadřuje způsob přeměny původního (*apriorního*) *stupně důvěry* k posuzovaným hypotézám $P(H_j)$ k modifikovanému (*posteriornímu*) stupni důvěře $P(H_j \mid D)$ po získání dat D.

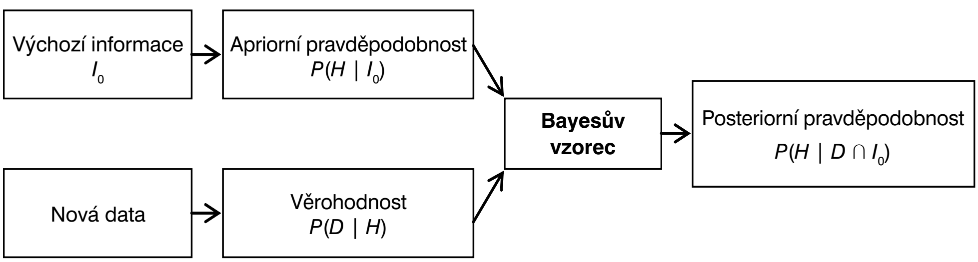
Podle *Bernardo, Smith* (1994, s. 43) se v různých podobách v bayesovské statistice vyskytují čtyři významově odlišné prvky $P(H_j)$, $P(D \mid H_j)$, $P(H_j \mid D)$ a $P(D)$, takže je užitečné mít pro ně vhodné pojmenování. Pro tento účel autoři definují *apriorní, posteriorní* a *prediktivní pravděpodobnosti* následujícím způsobem.

Představují-li H_j, $j = 1, 2, ..., J$, úplnou skupinu neslučitelných jevů (*hypotéz*), pak:

– $P(H_j)$, jsou **apriorní pravděpodobnosti** hypotéz H_j,
– $P(D \mid H_j)$, jsou **věrohodnosti** hypotéz H_j pro daná data D,
– $P(H_j \mid D)$, jsou **posteriorní pravděpodobnost** hypotéz H_j,
– $P(D)$ je **prediktivní pravděpodobnost** D vyplývající z věrohodností a apriorních pravděpodobností.

Kvůli rozdílu mezi *apriorními* a *posteriorními* pravděpodobnostmi zdůrazňují autoři *relativní* přínos informace z dat a domnívají se, že vhodnější by bylo vyjádřit stupeň důvěry v platnost hypotéz H_j *místo* $P(H_j)$ jako $P(H_j \mid I_0)$, kde I_0 označuje stav před získáním nové informace (před pořízením dat D), a tedy i analogicky místo $P(H_j \mid D)$ psát $P(H_j \mid I_0 \cap D)$ pro označení stupně důvěry po získání dat při respektování původního stavu I_0 a stupně důvěry $P(H_j)$. Tento přístup odpovídá i smyslu obrázku 1 z knihy *Zellnera* (1971, s. 10), který vyjadřuje proces postupného získávání nových informací a následné modifikace stupně důvěry v platnost posuzovaných hypotéz.

Obrázek 1
Proces revize pravděpodobností po získání nových dat



Příklad 3: Příprava studentů ke zkoušce ve vztahu k jejich výsledku u zkoušky

Přednáším studentům různých oborů základní kurz z počtu pravděpodobnosti. Na první přednášce studentům kromě příslušné literatury a potřeby návštěv přednášek i cvičení doporučím, aby během nastávajících třinácti týdnů samostatně spočítali přibližně sto příkladů, jejichž zadání (a někdy i řešení) jsou ve jmenovaných sbírkách příkladů nebo byly zadány přímo na cvičení či probrány na přednášce. Přitom vzpomenu, že moji spolužáci i já jsme kdysi před lety tímto způsobem získali sice minimální, ale docela solidní základy pravděpodobnostního myšlení.

Studenti mne většinou neposlechnou a podle pozdějších průzkumů jich nejméně 95 % potřebný počet příkladů nespočítá. Na konci semestru je zkouškový test s příklady, při kterém mohou používat připravené vzorce. Pro úspěšné absolvování předmětu stačí získat polovinu z maximálního počtu dosažitelných bodů. Podle mých záznamů mají studenti velmi malou úspěšnost a při svém prvním pokusu (mají k dispozici až tři pokusy) odhaduji, že úspěšných z těch, kteří doporučený počet příkladů nespočítali, je asi 20 %, zatímco u těch, kteří příklady spočítali je asi 99 %.

Označme jako H_1 hypotézu, že náhodně vybraný student příklady spočítal a jako H_2 hypotézu opačnou, že tento student doporučené příklady nespočítal. Dále jako D označme úspěšný test tohoto studenta. Podle uvedených údajů jsou moje stupně důvěry v jeho úspěšnost či neúspěšnost

$$P(H_1) = 0{,}05 \text{ a } P(H_2) = 0{,}95.$$

Dále

$$P(D \mid H_1) = 0{,}99 \text{ a } P(D \mid H_2) = 0{,}2,$$

z čehož pravděpodobnost

$$P(D) = 0{,}05 \;.\; 0{,}99 + 0{,}95 \;.\; 0{,}2 = 0{,}0495 + 0{,}19 = 0{,}2395.$$

Tedy asi 24 % studentů dlouhodobě test zvládá. Tento výpočet odpovídá i mým dlouhodobým zkušenostem, podle kterých test při prvním pokusu zvládne necelá čtvrtina studentů. Nás však zajímá konkrétní student, jehož výsledek testu (jak se dodatečně ukázalo) byl úspěšný. Pomocí *Bayesova vzorce* modifikujme po zjištění jeho úspěšnosti v testu apriorní pravděpodobnost 0,05 na posteriorní a dostáváme

$$P(H_1 \mid D) = 0{,}0495 \,/\, 0{,}2395 = 0{,}207,$$

takže

$$P(H_2 \mid D) = 0{,}793.$$

Závěr z těchto pro mne (velice málo povzbudivých) výsledků si jistě čtenář udělal sám, ale přidávám ujištění, že se snažím, aby testy studentů byly zvládnutelné a nebyly pro studenty překvapením (přibližně stovka testů je na internetu k dispozici všem zájemcům). Důvody neúspěchů je asi třeba hledat jinde.

Literatura k 1. části

ANDĚL, J. *Statistické metody*. Praha : Matfyzpress, 1998.

BAYES, T. An Essay Towards Solving a Problem in the Doctrine of Chances. *Philos. Trans. Royal Soc.* London, 1763, vol. 53, s. 370–418,.

BERGER, J. O. Bayesian Analysis a Look at Today and Thought of Tomorrrow. *J. American Statistical Association.* 2000, vol. 95, s. 1269–1276.

BERNARDO, J. M.; SMITH, A. F. M. *Bayesian Theory*. Chichester : John Wiley and Sons, 1994.

BERNOULLI, J. *Ars Conjectandi*. Book 4. Basel : Impensis Thurnisiorum, 1713.

BERRY, D. A. *Statistics – A Bayesian perspective*. San Francisco : Duxbury Press, 1996.

BLACKWELL, D. *Basic Statistics*. Los Angeles : McGraw Hill, 1969.

EFRON, B. *Large Scale Inference – Empirical Bayes Methods for Estimation, Testing and Prediction*. Cambridge; New York : Cambridge University Press, 2010.

de FINETTI, B. *La Prevision*: *Ses Lois Logique, ses Sources Subjectives*. Ann de l'institut Henri Poincare. 1937, vol. 7, s. 1–68. Přeloženo do angličtiny: *Studies in Subjective Probability*. New York : John Wiley and Sons, 1964.

de FINETTI, B. *Theory of Probability*. New York : John Wiley and Sons, 1974.

FISHBURN, P. C. The Axioms of Subjective Probability (with Discussion). *Statistical Science.* 1986, vol. 1, s. 335–358.

HAWKING, S. W. *A Brief History of Time*. New York : Bantam Books, 1988.

HEBÁK, P. *Texty k bayesovské statistice*. Praha : VŠE, 1999.

JACKMAN, S. *Bayesian Analysis for the Social Sciences*. Chichester : John Wiley and Sons, 2010.

JEFFREYS, H. *Scientific Inference*. 2. vyd. Cambridge : Cambridge University Press, 1957.

KOLMOGOROV, A. N. *Foundations of the Theory of Probability.* /Překlad z němčiny/. Berlin : Springer, 1933; 1950.

LEAMER, E. E. *Specification Searches*: *Ad Hoc Inference with Nonexperimental Data*. New York : John Wiley and Sons, 1978.

LINDLEY, D. V. Bayesian Inference. *Encyclopedia of Statistical Sciences, 1.* New York : John Wiley and Sons, 1982.

O'HAGAN, A. *Kendall´s Advanced theory of Statistics, Vol. 2B*, *Bayesian Inference*. London : Edward Arnold, 1994.

PRESS, S. J.; TANUR, J. M. *The Subjectivity of Scientists and Bayesian Approach* New York : John Wiley and Sons, 2001.

RAMSEY, F. P. *Truth and Probability*. Ed. BRAITHWAITE, R. B. London : Humanities Press, 1926.

RENYI, A. *Dialogy o matematice.* Praha : Mladá fronta, 1967, s. 124–182.

SAMANIEGO, F. J. *A Comparison of the Bayesian and Frequentist Approaches to Estimation*. New York : Springer, 2010.

SAVAGE, L. J. *The Foundations of Statistics*. New York : John Wiley and Sons, 1954.

SAVAGE, L. J. Elicitation of Personal Probabilities and expectations. *Journal of the American Statistical Association.* 1971, vol. 66, s. 781–801.

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.

ZILIAK, S. T.; McCloskey, D. N. *The Cult of Statistical Significance: How the Standard Error Costs Us Jobs, Justice, and Lives*. Michigan : University Press, 2007.

ZVÁRA, K.; ŠTĚPÁN, J. *Pravděpodobnost a matematická statistika*. Praha : Matfyzpress, 1997.

A COMPARISON OF CLASSICAL AND BAYESIAN PROBABILITY AND STATITICS (1)

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 world-wide. 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

(Pokračování stati v čísle 2/2012 AOP.)