

Učitel matematiky

Petr Emanovský

Statistická inference podle Thomase Bayese

Učitel matematiky, Vol. 32 (2024), No. 1, 1–14

Persistent URL: <http://dml.cz/dmlcz/152493>

Terms of use:

© Jednota českých matematiků a fyziků, 2024

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ*:
The Czech Digital Mathematics Library <http://dml.cz>

STATISTICKÁ INFERENCE

PODLE THOMASE BAYESE

PETR EMANOVSKÝ¹

Úvod

Současná matematická statistika využívá ve značné míře tradičních metod, které byly vyvinuty pro zkoumání hromadných jevů. Jsou založeny na srovnání vlastností vzorku „náhodně“ vybraného z populace s teoretickým pravděpodobnostním modelem. Na základě tohoto srovnání usuzujeme o vlastnostech celé populace, přičemž dokážeme kvantifikovat riziko omylu. Při tomto přístupu interpretujeme pravděpodobnost jako relativní četnost výskytu náhodného jevu. Jedním z významných nástrojů kvantitativního výzkumu, který odpovídá tomuto frekventistickému přístupu, je statistické testování hypotéz. Právě statistické testování hypotéz neboli určování statistické významnosti představuje jednu z typických forem statistického usuzování, při němž se uplatňuje princip statistické inference. Tato metoda je podrobně popsána v každé učebnici základů matematické statistiky, viz např. (Hendl, 2004; Chráska, 2007; Klementa et al., 1984; Komenda, 1994). Myšlenky, na nichž je tato metoda založena, se objevují již počátkem 18. století v práci skotského matematika Johna Arbuthnota (Arbuthnot, 1710; Emanovský, 2021). Do dnešní podoby byla tato metoda rozvinuta až ve 20. a 30. letech 20. století, zejména matematiky Ronaldem Fisherem (1890–1962), Jerzym Neymanem (1894–1981) a Egonem Pearsonem (1895–1980) (Emanovský, 2022). Bohužel,

¹Článek vznikl za podpory projektu Univerzity Palackého Olomouc „Algebraické a geometrické struktury“ IGA_PrF_2022.017.

tento nástroj je často používán nesprávně a statistická významnost jevů bývá špatně interpretována a má také některé „nedostatky“, viz např. články (Soukup, 2010; Soukup & Rabušic, 2007). Alternativní bayesovský přístup ke statistické inferenci je založen na Bayesově větě, která byla publikována již v 18. století.

Bayesovský přístup ke statistické inferenci

Thomas Bayes (1702–1761) byl anglický duchovní, jehož jméno je dnes spojováno hlavně s jeho větou o podmíněné pravděpodobnosti. Hlavní výsledky tohoto geniálního myslitele byly publikovány až dva roky po jeho smrti, zásluhou jeho přítele Richarda Price (1732–1791) (Bayes & Price, 1763). Bayes ukázal využití podmíněné pravděpodobnosti k výpočtu pravděpodobnosti předpovídaných událostí na základě výskytu událostí z nich vyplývajících (Bellhouse, 2004; Berkson, 1930). Tuto myšlenku tzv. inverzní pravděpodobnosti dále rozvíjel Pierre Simon Laplace (1749–1827) a další vědci 19. století, potom však upadla na delší dobu v zapomnění (Fienberg, 1992). Nový zájem o tuto metodu se začíná objevovat od poloviny 20. století v pracích mnoha autorů. Patří k nim např. Bruno De Finetti, Richard Cox, Richard Jeffrey a další. Tito statistici přispěli především k rozvoji metody statistické inference založené na Bayesově větě. Teprve zakladatel frekventistické statistiky R. A. Fisher však poprvé použil termín „bayesovská statistika“ (Fienberg, 2006). Hlavní myšlenky bayesovského přístupu ke statistice jsou následující (Lindley, 1972; Bolstad, 2007):

- Protože neznáme přesnou hodnotu populačních parametrů, považujeme je za náhodné veličiny.
- Pravidla pro počítání s pravděpodobnostmi jsou použita přímo k inferenci populačních parametrů.
- Pravděpodobnostní tvrzení o parametrech musí být interpretováno jako „stupeň přesvědčení“ (prior). Pravděpodobnostní rozdělení priorů musí být subjektivní. Každá osoba může mít svůj vlastní prior obsahující relativní váhy, které tato osoba dává všem možným hodnotám parametru. Tím je vyjádřena subjektivní přijatelnost jednotlivých hodnot parametru před tím, než osoba pozoruje data.

- Revidujeme naši předběžnou představu o parametrech na základě zjištěných dat užitím Bayesovy věty. Tím dostaneme aposteriorní rozdělení, které poskytuje relativní váhy hodnot všech parametrů po analýze dat. Aposteriorní rozdělení pochází ze dvou zdrojů: z apriorního rozdělení a z pozorovaných dat.

Podmíněná pravděpodobnost a Bayesův vzorec

Z teorie pravděpodobnosti známe pojem podmíněné pravděpodobnosti náhodného jevu A za podmínky, že nastal jev B (předpokládáme, že $P(B) \neq 0$). Tato pravděpodobnost je dána vztahem

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Běžně se používá rovněž Bayesův vzorec ve tvaru

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

který umožňuje výpočet podmíněných pravděpodobností $P(B_i|A)$ neslučitelných jevů $B_1, B_2, \dots, B_n$ za podmínky, že nastal jev A . Předpokládá se přitom, že $P(B_1 \cup B_2 \cup \dots \cup B_n) = 1$.

Bayesovská inference pro diskrétní náhodnou veličinu

Při klasické statistické analýze považujeme zkoumaný parametr za pevně danou neznámou konstantu, kterou odhadujeme ze získaných dat. Pracujeme tedy s podmíněnou pravděpodobností $P(\text{data}|\text{parametr})$. Bayesovský přístup předpokládá, že parametr je náhodná veličina, jejíž pravděpodobnostní rozdělení určujeme při daných datech. Zajímá nás tedy podmíněná pravděpodobnost $P(\text{parametr}|\text{data})$. Výpočet takových podmíněných pravděpodobností nám právě umožňuje Bayesův vzorec pro diskrétní náhodnou veličinu. Předpokládejme, že parametr považujeme za diskrétní náhodnou veličinu X s možnými hodnotami $x_1, x_2, \dots, x_m$

(neznámá data) a pozorovaná data popisuje náhodná veličina Y s možnými hodnotami $y_1, y_2, \dots, y_n$. Podmíněné pravděpodobnosti jednotlivých hodnot parametru $x_1, x_2, \dots, x_m$ za podmínky, že byla pozorována hodnota y_j , jsou dány Bayesovým vzorcem pro diskrétní náhodnou veličinu (f značí pravděpodobnostní funkce pozorovatelné náhodné veličiny, g značí pravděpodobnostní funkce nepozorovatelné náhodné veličiny):

$$g(x_i|y_j) = \frac{g(x_i)f(y_j|x_i)}{\sum_{j=1}^n g(x_i)f(y_j|x_i)},$$

kde $g(x_i)$ je tzv. apriorní funkce (prior), $f(y_j|x_i)$ je věrohodnostní funkce (likelihood), $g(x_i|y_j)$ je aposteriorní funkce (posterior) a $\sum_{j=1}^n g(x_i)f(y_j|x_i)$ je tzv. evidence (marginal likelihood). Apriorní funkce zachycuje naše znalosti o pravděpodobnostním rozdělení parametru (náhodné veličiny X) ještě před pozorováním dat. Věrohodnostní funkce představuje podmíněnou pravděpodobnost, že byla pozorována hodnota y_j za podmínky, že parametr má hodnotu x_i . Aposteriorní funkce vyjadřuje podmíněnou pravděpodobnost, že parametr má hodnotu x_i , za podmínky, že byla uvažována pozorovaná data. Evidence představuje celkovou pravděpodobnost naměřených dat, kterou dostaneme sečtením společných pravděpodobností $f(x_i, y_j)$ přes všechny naměřené hodnoty.

Ukažme si podstatu bayesovského testování hypotéz na následujícím jednoduchém příkladu:

Příklad 1. Uvažujme diskrétní náhodnou veličinu r , která může nabývat hodnot 1, 2, 3 a 4, přičemž pravděpodobnostní rozdělení hodnot této veličiny závisí na parametru θ , který nabývá hodnot 0, 1 a 2 (tab. 1).

Tab. 1: Pravděpodobnostní rozdělení náhodné veličiny r
(Christensen, 2005)

r	1	2	3	4
$f(r 0)$	0,980	0,005	0,005	0,010
$f(r 1)$	0,100	0,200	0,200	0,500
$f(r 2)$	0,098	0,001	0,001	0,900

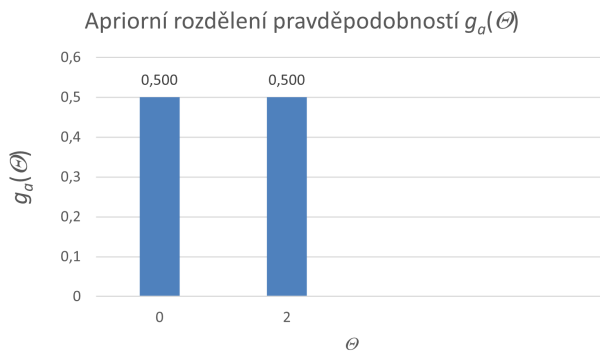
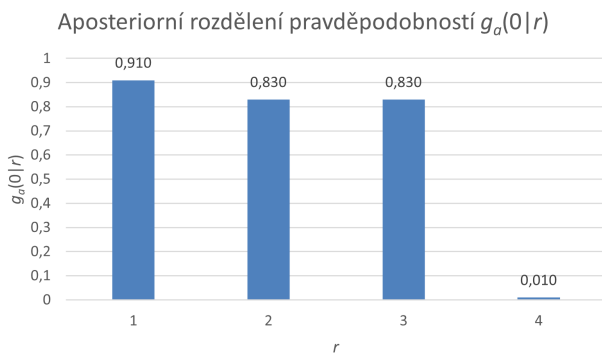
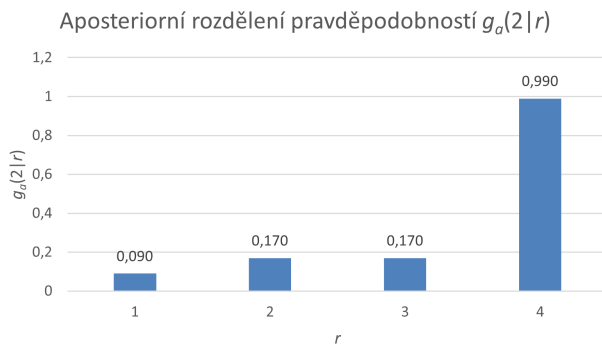
Při bayesovské analýze nepracujeme s nulovou a alternativní hypotézou, nýbrž porovnáváme dvě rovnocenné hypotézy a snažíme se určit, která z hypotéz je pravděpodobnější. Uvažujeme hypotézu $H_A: \theta = 0$ a hypotézu $H_B: \theta = 2$. Bayesovská analýza vyžaduje znalost apriorních pravděpodobností výskytu hodnot parametru θ . Pro náš příklad můžeme Bayesův vzorec pro diskrétní náhodnou veličinu přepsat do tvaru:

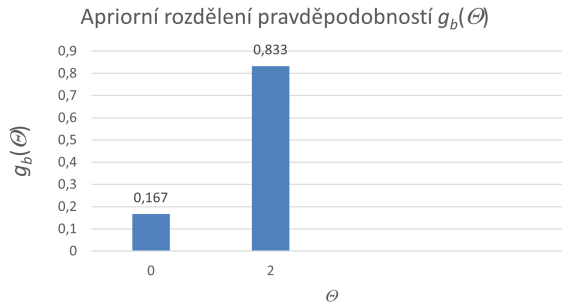
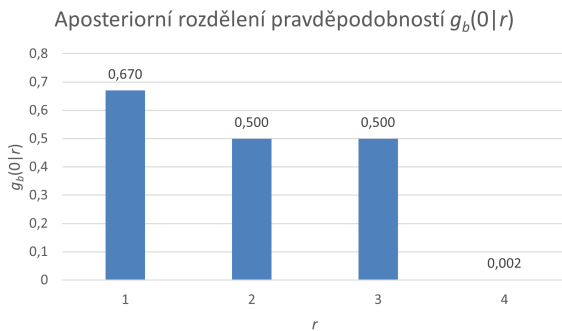
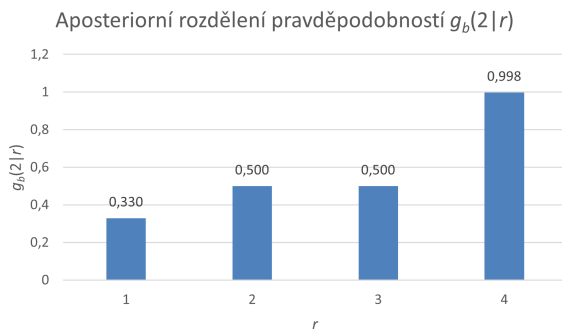
$$g(\theta|r) = \frac{f(r|\theta)g(\theta)}{f(r|0)g(0) + f(r|2)g(2)}, \quad \theta \in \{0; 2\}.$$

Tedy $g(\theta)$ označuje apriorní funkci, $f(r|\theta)$ je věrohodnostní funkce, $g(\theta|r)$ je aposteriorní funkce a $f(r|0)g(0) + f(r|2)g(2)$ odpovídá evidenci. Rozhodovací proces je založen na porovnávání hodnot aposteriorních pravděpodobností (tab. 2). Uvažujeme zde dvě apriorní rozdělení a a b . V prvním rozdělení a mají obě hodnoty parametru θ stejnou pravděpodobnost 0,5, v rámci druhého rozdělení b je hodnota $\theta = 2$ pětkrát pravděpodobnější než hodnota $\theta = 0$ (obr. 1–6).

Tab. 2: Aposteriorní podmíněné pravděpodobnosti hodnot $\theta \in \{0; 2\}$ pro dvě apriorní rozdělení a, b . (Christensen, 2005)

Prior	r	1	2	3	4
	$f(r 0)$	0,980	0,005	0,005	0,010
	$f(r 2)$	0,098	0,001	0,001	0,900
$g_a(0) = \frac{1}{2}$	$g_a(0 r)$	0,910	0,830	0,830	0,010
$g_a(2) = \frac{1}{2}$	$g_a(2 r)$	0,090	0,170	0,170	0,990
$g_b(0) = \frac{1}{6}$	$g_b(0 r)$	0,670	0,500	0,500	0,002
$g_b(2) = \frac{5}{6}$	$g_b(2 r)$	0,330	0,500	0,500	0,998

Obr. 1: Apriorní rozdělení $g_a(\theta)$ Obr. 2: Aposteriorní rozdělení $g_a(0|r)$ Obr. 3: Aposteriorní rozdělení $g_a(2|r)$

Obr. 4: Apriorní rozdělení $g_b(\theta)$ Obr. 5: Aposteriorní rozdělení $g_b(0|r)$ Obr. 6: Aposteriorní rozdělení $g_b(2|r)$

Bayesova eseje

Bayesovo dílo *An essay towards solving a problem in the doctrine of chances* z roku 1763 patří k jedné z nejobtížnější čitelných prací v historii statistiky. Jedná se o práci mimořádně duchaplného a extrémně schopného matematika a filosofa, která je navíc napsaná těžkopádným jazykem Newtonovy doby (Stigler, 1982). Není tedy divu, že čtenáři tohoto díla byli často odrazeni náročností četby ani že autorovy myšlenky nebyly vždy správně pochopeny. To se týká nejen Bayesových současníků, ale i čtenářů novějších vydání eseje, které jsou dnes běžně k dispozici (Barnard, 1958). Zkusme si v následujícím odstavci vysvětlit některé hlavní Bayesovy myšlenky pomocí současných matematických pojmů, symboliky a terminologie. Je třeba si uvědomit, že Bayes byl téměř současníkem Isaaca Newtona (1643–1727). Žil tedy v době, kdy matematická disciplína, kterou dnes nazýváme infinitezimální počet, byla opravdu v plenkách.

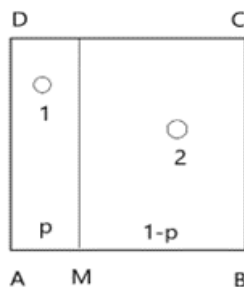
V úvodní části eseje autor nadšeně prezentuje základní pojmy teorie pravděpodobnosti, z nichž většinu bychom dnes pokládali za elementární. Bayes definoval pravděpodobnost jako zdánlivě subjektivně určené očekávání budoucího výsledku náhodného pokusu. Tím se jeho přístup lišil od frekventistů, kteří pravděpodobnost chápali jako relativní četnost opakujícího se náhodného pokusu. Bayes zavedl pojem podmíněné pravděpodobnosti a dokázal, že pro dvě události A a B takové, že A předchází v čase B , platí

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Bayes geniálně popsal analogii mezi matematickým modelem binomického rozdělení pravděpodobnosti a fyzikálním modelem „biliárového stolu“ (Stigler, 1982; Gillies, 1987; Mačák, 1997; Bellhouse, 2004). Bayesova analogie je založena na myšlenkovém modelu dokonale rovného a vyváženého stolu čtvercového půdorysu o jednotkové straně (obr. 7, 8). Předpokládáme, že vhodíme-li na tento stůl biliárovou kouli, může se zastavit v kterémkoliv bodě pomyslného jednotkového čtverce se stejnou pravděpodobností. Zvolme jednu stranu čtverce, např. AB , za část x -ové osy, na kterou budeme na-

nášet hodnoty vodorovné souřadnice konečné polohy dané koule. Předpokládejme, že se první vhozená koule (červená) zastavila v pozici s horizontální souřadnicí dané bodem M (obr. 7). Poté byla stejným způsobem vhozena na stůl bílá koule, a to opakovaně n -krát. Za úspěšný pokus je považována situace, kdy se bílá koule zastaví nalevo od konečné polohy červené koule, tj. horizontální souřadnice bílé koule je menší než u červené koule (např. pozice 1 na obrázku 7). Zastaví-li se bílá koule napravo od červené (např. v poloze 2), je to považováno za neúspěch. Řešme následující dva problémy:

1) Na stůl byla vhozena první (červená) koule, která zaujala konečnou polohu danou bodem M . Jaká je pravděpodobnost toho, že při následném n -krát opakovaném hodu bílou koulí budeme pozorovat právě k úspěchů?



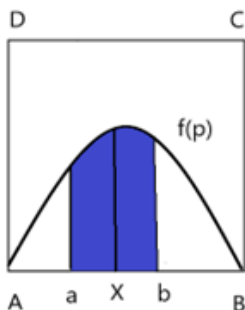
Obr. 7: Bayesův model biliárového stolu po vhození první koule

Tato situace je popsána diskrétní náhodnou veličinou X s binomickým rozdělením pravděpodobností odpovídajícím n nezávislým pokusům s pravděpodobností p úspěchu v každém jednotlivém pokusu. Je známo, že náhodná veličina X nabývá hodnot $k = 0, 1, 2, \dots, n$ (počet úspěchů) s pravděpodobnostmi

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}.$$

Hodnota parametru p v tomto vztahu je určena polohou červené koule. Platí totiž $p = \frac{d(AM)}{d(AB)} = d(AM)$. Délka úsečky AM je tedy rovna pravděpodobnosti úspěchu při každém jednotlivém hodu bílou koulí.

2) Uvažujme situaci před vhozením červené koule a opět se ptejme, jaká je pravděpodobnost, že budeme pozorovat právě k úspěchů při n hodech bílou koulí. Pro každou konkrétní polohu X červené koule bude pravděpodobnost úspěchu p dána délkou úsečky AX a pravděpodobnost neúspěchu $1 - p$ dána délkou úsečky XB (obr. 8). Budeme-li tedy hodnoty x -ových souřadnic konečné polohy červené koule považovat za hodnoty parametru p , dosáhneme touto myšlenkovou konstrukcí toho, že parametr p bude mít uniformní rozdělení na jednotkovém intervalu. Vzhledem k tomu, že uspořádání množiny bodů úsečky AB je spojitě, dostáváme tak spojitou náhodnou veličinu definovanou na intervalu $\langle 0, 1 \rangle$ s hustotou pravděpodobnosti $f(p) = \binom{n}{k} p^k (1 - p)^{n-k}$. Graf funkce $f(p)$ je znázorněn na obrázku 8.



Obr. 8: Bayesův model biliárového stolu před vhozením první koule

Bayes ve své eseji řešil problém, jak určit podmíněnou pravděpodobnost $P(a < p < b \mid Y = k)$, kde Y je náhodná veličina s definičním oborem všech možných poloh všech koulí a oborem hodnot odpovídajícím množině všech možných výsledků, tj. počtů úspěchů $k \in \{0, 1, \dots, n\}$.

Bayes ukázal, že

$$P(a < p < b \mid Y = k) = \int_a^b \binom{n}{k} p^k (1 - p)^{n-k} dp.$$

Této pravděpodobnosti geometricky odpovídá obsah vybarvené plochy pod křivkou (obr. 8). Rozšířením intervalu (a, b) na celý

jednotkový interval $\langle 0, 1 \rangle$ dostaneme jako důsledek

$$P(Y = k) = \int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp.$$

Tato pravděpodobnost je geometricky určena obsahem celkové plochy pod křivkou (obr. 8). Odtud

$$P(a < p < b \mid Y = k) = \frac{\int_a^b \binom{n}{k} p^k (1-p)^{n-k} dp}{\int_0^1 \binom{n}{k} p^k (1-p)^{n-k} dp}.$$

Bayes dále ukázal, že když binomický parametr p má uniformní rozdělení pravděpodobností, pak pravděpodobnost

$$P(Y = k) = \frac{1}{n+1} \text{ pro jakékoliv } k \in \{0, 1, \dots, n\}.$$

To znamená, že náhodná veličina Y má diskrétní uniformní rozdělení pravděpodobností. Při n -krát opakovaném hodu bílou koulí tedy není důvod domnívat se, že by byl některý počet úspěchů pravděpodobnější než jiný.

Závěr

Bayesovské metody jsou zpravidla početně náročnější, mají však oproti klasickým metodám některé výhody. Mimo jiné jsou použitelné i v případě jevů, které nejsou hromadné povahy, dají se tedy aplikovat i v kvalitativním výzkumu (sociální výzkum, pedagogický výzkum). Výhody využití bayesovské statistické analýzy v rámci sociálních a biologických věd jsou popsány v (Hendl, 2011):

1. Jsou dostupné určité apriorní znalosti zkoumaného fenoménu, které lze efektivně využít v rámci bayesovského přístupu.
2. Kromě systémové variability popsané nějakou dostupnou teorií, existuje v sociálních i biologických vědách značná individuální variabilita. Bayesovské metody mohou v některých případech poskytnout kvalitnější rámec pro naše usuzování.

Bayesovské metody dnes mají široké uplatnění všude tam, kde se pracuje s nejistými informacemi, tedy např. ve finanční matematice, managementu, psychologii, sociologii, pedagogickém výzkumu, medicínském výzkumu, soudnictví, kriminologii, ekologii, při ekonomických studiích apod. Bayesovský přístup hraje rovněž důležitou roli v matematické logice.

Žádný výzkumník, který se seriózně zabývá statistickou analýzou dat, by se neměl omezovat pouze na klasické frekventistické metody, ale měl by se seznámit i s některými alternativními přístupy, mezi které patří i bayesovská analýza.

Literatura

- [1] Arbuthnot, J. (1710). An argument for divine providence, taken from the constant regularity observed in the births of both sexes. *Philosophical Transactions of the Royal Society*, 27, 186–190. <http://www.jstor.org/stable/103111>
- [2] Barnard, G. A. (1958). Studies in the history of probability and statistics: IX. Thomas Bayes's essay towards solving a problem in the doctrine of chances. *Biometrika*, 45(3/4), 293–315.
- [3] Bayes, T., & Price, R. (1763). An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53, 370–418.
- [4] Berkson, J. (1930). Bayes' theorem. *The Annals of Mathematical Statistics*, 1(1), 42–56.
- [5] Bellhouse, D. R. (2004). The reverend Thomas Bayes, FRS: A biography to the celebrate the tercentenary of his birth. *Statistical Science*, 19(1), 3–43. <https://doi.org/10.1214/088342304000000189>
- [6] Bolstad, W. M. (2007). *Introduction to Bayesian statistics*. John Wiley & Sons.
- [7] Christensen, R. (2005). Testing Fisher, Neyman, Pearson, and Bayes. *The American Statistician*, 59(2), 121–126. <https://doi.org/10.1198/000313005X20871>

- [8] Emanovský, P. (2021). První statistické testování hypotézy podle Johna Arbuthnota. *Učitel matematiky*, 29(1), 26–36.
- [9] Emanovský, P. (2022). Statistické testování hypotéz třikrát jinak. *Učitel matematiky*, 30(1), 3–14.
- [10] Fienberg, S. E. (1992). A brief history of statistics in three and one-half chapters: a review essay. *Statistical Science*, 7(2), 208–225.
- [11] Fienberg, S. E. (2006). When did Bayesian inference become "Bayesian"? *Bayesian Analysis*, 1(1), 1–40.
- [12] Gillies, A. D. (1987). Was Bayes a Bayesian? *Historia Mathematica*, 14, 325–346.
- [13] Hendl, J. (2004). *Přehled statistických metod – zpracování dat*. Portál.
- [14] Hendl, J. (2011). Bayesovská statistická analýza s podporou výpočetně intenzivních procedur. In *Medsoft 2011, sborník příspěvků* (s. 22–32). Creative Connections. <https://www.creativeconnections.cz/medsoft/2011.html>
- [15] Chráška, M. (2007). *Metody pedagogického výzkumu*. Portál.
- [16] Klementa, J, Komenda, S., & Kunert, E. (1984). *Statistické metody v pedagogickém výzkumu*. VUP.
- [17] Komenda, S. (1994). *Biometrie*. VUP.
- [18] Lindley, D. V. (1972). *Bayesian statistics a review*. Society for Industrial and Applied Mathematics.
- [19] Mačák, K. (1997). Poznámky k formování teorie pravděpodobnosti v 17. a 18. století. In J. Bečvář & E. Fuchs (Eds.), *Historie matematiky II* (s. 29–67). Prometheus. <https://www.dml.cz/handle/10338.dmlcz/401036>
- [20] Soukup, P. (2010). Nesprávná užívání statistické významnosti a jejich možná řešení. *Data a výzkum – SDA Info*, 4(2), 77–104.
- [21] Soukup, P., & Rabušic, L. (2007). Několik poznámek k jedné obsesi českých sociálních věd, statistické významnosti. *Sociologický časopis/Czech Sociological Review*, 43(2), 379–395.

- [22] Stigler, S. M. (1982). Bayes's Bayesian inference. *Journal of the Royal Statistical Society Series A (General)*, 145(2), 250–258.

Abstract

Traditional (frequentist) statistical methods are based on probabilistic models applicable only to mass phenomena, the occurrence of which can be monitored repeatedly in many situations. In this case, the probability is understood as the relative frequency of the occurrence of the observed phenomenon. In contrast, the Bayesian approach defines probability as a measure of the belief in the truth of a given statement. The article focuses on clarifying the basic ideas of Bayesian statistic analysis and notes the history of its origin and development.

Petr Emanovský

Přírodovědecká fakulta Univerzity Palackého v Olomouci

17. listopadu 1192/12

771 46 Olomouc

e-mail: petr.emanovsky@upol.cz