Jerzy Neyman
Vznik matematické statistiky

Pokroky matematiky, fyziky a astronomie, Vol. 27 (1982), No. 3, 136--147

Persistent URL: http://dml.cz/dmlcz/139699

Vznik matematické statistiky*)

Jerzy Neyman, Berkeley, USA

Jerzy Neyman, původem Polák, se narodil 16. dubna 1894 v Benderách v Rusku. V roce 1916 ukončil studium matematiky na charkovské univerzitě a v letech 1917—1921 přednášel na charkovské technice. Od roku 1921 do roku 1923 pracoval jako statistik v zemědělském výzkumném ústavu v Bydgoszci. Poté, co dosáhl doktorátu matematiky na varšavské univerzitě, tam přednášel v letech 1923—1934 a zároveň byl vedoucím biometrické laboratoře v Nenckého ústavu ve Varšavě. V letech 1926—27 pobýval v Londýně a v Paříži jako stipendista Rockefellerovy nadace. Od roku 1934 do roku 1938 působil v University College v Londýně nejprve jako asistent, později jako docent matematiky. Od roku 1938 je profesorem a ředitelem statistické laboratoře na kalifornské univerzitě v Berkeley.

Neymanovy vědecké publikace zasahují do těchto oblastí: teorie množin, pravděpodobnosti, statistiky, astronomie, biologie, meteorologie a filozofie vědy. Získal pět čestných doktorátů (na universitách v Chicagu, Kalifornii, Varšavě, Stockhomu a v Indickém statistickém ústavu). Je čestným členem (i medailistou) Královské statistické společnosti v Londýně a londýnské matematické společnosti. Je členem Mezinárodního astronomického svazu a čestným předsedou Mezinárodního statistického institutu. Byl redaktorem a spoluredaktorem sborníků Proceedings of Berkeley Symposia on mathematical statistics z let 1945, 1946, 1950, 1955, 1960, 1965, 1970—71. Redigoval Koperníkův svazek Akademie věd USA v roce 1974. V roce 1968 obdržel vyznamenání U.S. National Medal of Science.

Vědecké problémy jako podnět pro rozvoj nových matematických odvětví

Je obecným pravidlem, že nové matematické disciplíny mají své kořeny v problémech vědy, to znamená v úsilí člověka porozumět zákonitostem vesmíru. Aritmetika má svůj původ v tom, že si naši dávní předkové uvědomili rozdíl mezi pojmy „málo“ a „mnoho“. Geometrie vznikla z představ, co je blízko a daleko, v pokročilejším stupni pak z potřeby měřit půdní plochy. Diferenciální počet vycházel z empirických představ rychlosti a zrychlení. Tyto příklady se týkají přímého spojení mezi matematickými disciplínami a empirickým bádáním. Souvislost mezi vědou a některými jinými odvětvími matema-

---

*) Překlad 1. části článku Jerzyho Neymana: *The Emergence of Mathematical Statistics* (*General Background*), uveřejněného ve Sborníku z konference "On the History of Statistics" (ed. Donald B. Owen), Marcel Dekker, Inc., NY 1977.

Reprinted from Proceedings of the Conference „On the History of Statistics“, pp. 149—166, by curtesy of Marcel Dekker, Inc.

tiky, jako je na příklad teorie množin, také existuje, není ale úplně bezprostřední. Tato odvětví vznikla, když se v již existujících disciplínách, souvisejících s reálným světem úžeji, objevily obtíže logické povahy. Poté, co začne zkoumat soubor přirozených jevů, se matematická disciplína obvykle vzdálí od své „mateřské domény" vědy a začne žít svým životem jako abstraktní teorie a jenom příležitostně se objeví zpětná vazba. Moderní geometrie má tedy jen málo co dělat s měřením ploch polí; tím se zabývá zeměměřičství. Zpětná vazba moderní geometrie a empirické vědy se dá ukázat na příkladu neeuklidovské geometrie, jež byla nejprve pokládána za logickou strukturu bez praktického užitku, nyní se však mnoho astronomů domnívá, že reálně charakterizuje prostor, v němž žijeme.

Počátky teorie pravděpodobnosti a matematické statistiky jsou také empirické. Nyní obě dosáhly stavu zralosti a žijí svým vlastním životem. Vzhledem k určitým okolnostem, které budou probrány dále, jsme však svědky zdánlivě nepřetržité posloupnosti zpětných vazeb k empirickým vědám a současně podobného sledu podnětů plynoucích z empirických věd do matematické teorie pravděpodobnosti i statistiky. Také ostatní matematické disciplíny zásobují pravděpodobnost a statistiku novými analytickými nástroji.

Tak, jak bylo právě popsáno, by se celý proces jevil jako velmi harmonický. Dojem harmonie však mizí, jakmile obrátíme svou pozornost k praktikům místo k různým disciplínám. Proud nových problémů, nezvyklých myšlenek a nových technik ohromuje lidi starší generace, kteří by rádi odpočívali na vavřínech. Avšak energická a matematicky lépe vybavená „mládež" je nenechá na pokoji. Tak nevyhnutelně pokračuje konflikt mezi otci a syny, často přinášející něco hořkosti. Situaci ilustruje citát z Williama Fellera [1], poměrně nedávno zesnulého vynikajícího učence:

> Bylo konstatováno, že moderní teorie pravděpodobnosti je příliš abstraktní a příliš obecná na to, aby byla užitečná... Proti tomu by se dalo poukázat na nové neočekávané aplikace, které umožnila abstraktní teorie stochastických procesů, nebo na nové pohledy, jež poskytuje moderní teorie fluktuací a které opět popírají intuici... Diskuse je však zbytečná, odsoudit je příliš snadné. Věci praktické dnes byly ještě včera pomlouvány jako neužitečné a teorie, které budou užitečné zítra, vždy budou praktici dneška označovat za bezcenné hry.

Empirické pozadí teorie pravděpodobnosti

Jsou dvě třídy jevů, které si žádají rozvoj matematické teorie, jež by se dala nazvat teorií pravděpodobnosti. Jednou z nich je třída zřejmě stabilních relativních četností, která dala vzniknout tak zvané četnostní teorii. Druhá je třída psychologických jevů spojených s pocity důvěry a nedůvěry. Odpovídající teorie pravděpodobnosti se nazývá subjektivní. Tento článek pojednává pouze o četnostní teorii.

Základní představou četnostní teorie pravděpodobnosti je náhodový mechanismus. Mechanismus nazveme náhodovým, jestliže (1) jeho činnost má za následek pouze jeden z několika možných výsledků $A, B, \ldots, C$, a je prakticky nemožné předvídat, který z nich nastane v daném případě, a jestliže (2) četnosti těchto výsledků při mnohokrát opakované činnosti mechanismu se zdají být předpověditelné. Házení mincí, vrhání kostkou, vytahování kuliček z osudí jsou dobře známé příklady jednoduchého náhodového mechanismu. Původ představy náhodového mechanismu je obtížné vysledovat. Dost možná, že pocta autorství náleží prvnímu podvodníkovi, který zatížil svou kostku. To se muselo stát v šerém dávnověku. Slyšel jsem, že v pohřebních komorách jistých egyptských faraónů bylo nalezeno několik souprav kostek, některé správné a jiné falešné. Před zatížením kostky si musel být podvodník vědom důležité skutečnosti, že četnost šestky v dlouhé řadě vrhů kostkou K_1 není nutně rovna odpovídající četnosti u kostky K_2. Tedy četnosti dopadů kostky jedním ze šesti možných způsobů se jevily jako její měřitelné vlastnosti, jako je objem a hmotnost. Když si toho falešný hráč povšiml, musel přemýšlet, jak vyrobit kostku s více či méně předem určenými frekvencemi dopadů kostky tím či oním způsobem.

Vše, co bylo dosud řečeno, poukazuje na to, že jsme si vědomi, že náhodový mechanismus je empirickou skutečností. Počátky četnostní teorie pravděpodobnosti sahají až k problému, zda se dá vypočítat četnost jevu E v dlouhé řadě pokusů ze známých četností jevů $A, B, \ldots, C$ s jevem E souvisejících. Při nutné míře zjednodušení se dá říci, že teorie pravděpodobnosti vznikla v roce 1713, kdy byla publikována kniha *Ars Conjectandi* od Jacoba Bernoulliho [2].

Při tomto empirickém pozadí by se mohla četnostní teorie pravděpodobnosti spíše nazývat počet četnosti nebo tak nějak. Přiřazení slova pravděpodobnost asi povstalo z intuitivní souvislosti mezi pojmy četný a pravděpodobný.

Bylo podniknuto mnoho pokusů, jak formulovat základy abstraktní, čistě matematické teorie pravděpodobnosti jako pojmového modelu empirického světa relativních četností. Nejúspěšnější práce v tomto směru náleží A. N. Kolmogorovi [3]. Práce s titulem *Grundbegriffe der Wahrscheinlichkeitsrechnung* byla publikována v roce 1933 u J. Springera v Berlíně. Toto dílo je čistě matematické. Autor však poukazuje na to, že souvislost jeho teorie s empirickým světem je četnostní, stejně jako je tomu u teorie současně hlásané Richardem von Misesem [4]. Základní otázka zní: „Jak často?“

„Individualistické“ a „pluralistické“ předměty výzkumu ve vědě

Když Newton a později Laplace pracovali na nebeské mechanice, jejich studie byly „individualistické“. Typickým problémem bylo vypočítat dráhu dané planety, řekněme Marsu, za předpokladu newtonské přitažlivosti Slunce. V několika studiích tohoto druhu Laplace shledal, že orbitální roviny všech studovaných planet jsou si navzájem dosti blízké. Čtenář si všimne, že poslední tvrzení je kvalitativně odlišné od rozličných závěrů týkajících se určité planety, řekněme Marsu. Výrok „Orbitální roviny všech planet si jsou navzájem blízké“ se netýká individuální planety, nýbrž celé kategorie nebeských těles nazývaných planety. Tady se setkáváme s novým předmětem studia: kategorií objektů, které všechny vyhovují jisté definici, avšak liší se ve svých individuálních charakteristikách. Takový předmět studia se nazývá „pluralistický“.

Když se nyní opět vrátíme do 18. století, můžeme se s Laplaceem [5] ptát, zda komety náleží jako planety ke sluneční soustavě, a otázka je opět pluralistická. V té době byla pozorování několika komet postačující pro výpočet jejich drah a Laplace si všiml, že roviny těchto drah si zdaleka nejsou tak blízké, jako je tomu u planet. Byla to tato okolnost, která přivedla Laplacea k další pluralistické otázce, zda se kategorie předmětů, zvaných komety, liší od jiné kategorie, tak zvaných planet.

Předměty studia v moderním vědeckém výzkumu jsou především pluralistické. Na příklad v astronomii se ptáme, zda eliptické galaxie (kategorie galaxií) jsou hmotnější než spirály (jiná kategorie galaxií). V medicíně jsou důležité otázky týkající se vnitřních chorob pacientů, jevících určitý soubor pozorovatelných příznaků (tj. určité kategorie pacientů). V dopravním inženýrství je jednou ze závažných otázek, zda při automobilové dopravě na dvouproudé dálnici (kategorie dopravy) dochází k více nehodám než na tříproudé dálnici atd.

V moderní terminologii se kategorii předmětů, objektů pluralistického studia, říká populace. Obvykle tedy mluvíme o populacích lidí vyhovujících určité definici, o populacích eliptických nebo spirálových galaxií, o populaci molekul plynu v uzavřené nádobě apod. Zde je důležité zaznamenat souvislost s teorií pravděpodobnosti. Tato souvislost se projevuje prostřednictvím všudypřítomné otázky „Jak často?“, která se objevuje ve všech pluralistických studiích. Laplaceovo podezření, že komety nejsou řádnými členy sluneční soustavy, bylo založeno na rozdělení četností úhlů jejich orbitálních rovin s ekliptikou. Otázka hmotností eliptických a spirálových galaxií je ve skutečnosti otázkou rozdělení četností hmotností v těchto dvou populacích. Výše zmíněný lékařský problém je otázkou stupně souvislosti mezi vnějšími příznaky a fyzickými chorobami pacientů. Jak často na příklad pacienti, kteří mají tyto určité příznaky, trpí rakovinou?*)

Matematická statistika jako odvětví matematiky vytvořené pro pluralistické studie přírody

Zatímco se pluralistické studie od dob Laplaceových stále znovu objevovaly, trvalo několik desetiletí, než si výzkumní pracovníci uvědomili, že představují novou kategorii a že tato nová kategorie vědeckých problémů se rozpadá na několik podkategorií, z nichž každá vyžaduje novou matematickou disciplínu. Soubor matematických disciplín, které byly rozvíjeny, aby splnily tyto potřeby, tvoří to, co nazýváme matematická statistika (nebo příležitostně „statistická matematika“).

Popisná statistika. Prvním odvětvím matematické statistiky (nikoli však historicky) je to, co nyní nazýváme popisná statistika. Původní termín pro označení tohoto odvětví vytvořil v 19. století G. T. Fechner [6]. Zněl *Kollektivmasslehre.* Je poněkud dlouhý, ale velmi názorný. Je-li předmětem našeho studia populace jedinců, z nichž každý je charakterizován vlastností X, měnící se od individua k individuu, vyžaduje toto studium

*) Význam této otázky by neměl být zaměňován s významem otázky následující: Jak často se shledá, že oběti rakoviny byly kuřáky?

matematickou disciplínu zvláště uzpůsobenou k popisu toho, co nyní nazýváme *rozdělením* vlastnosti X v populaci. Jak je dobře známo, realizace potřeby popsat empirická rozdělení vytvořila několik systémů teoretických četnostních křivek a ploch, pocházejících od Brunse, Grama, Charliera a nejúspěšnějšího z nich, Karla Pearsona [7]. Další zobecnění náleží Romanovskému a mnoha dalším autorům.

Problém „dobré shody“. Jednou z bezprostředně následujících partií matematické statistiky je ta, která se zabývá „dobrou shodou“. Souvisí to se skutečností, že navzdory tomu, že principiální zájem bývá soustředěn na soubor nějakých objektů, řekněme eliptických galaxií, nelze tento soubor podrobit vyčerpávajícímu studiu. Místo celé populace, která nás zajímá, jsme nuceni se zabývat pouze souborem z ní vybraným. Je-li dána funkce $F(x)$, o níž předpokládáme, že představuje rozdělení veličiny X na populaci, vzniká otázka, zda odchylky od F, zjištěné na výběrovém souboru, mohou být přisouzeny výběrové chybě. První známá metoda, jak řešit tento problém, je χ^2-test, který se přičítá Karlu Pearsonovi. Byl publikován v roce 1900 [8]. O několik desetiletí později bylo navrženo mnoho alternativních metod, z nichž nejznámější pocházejí od H. Craméra ze Švédska [9], Richarda von Miessa z Německa [10] a A. N. Kolmogorovova a N. V. Smirnova ze Sovětského svazu [12]: jaké to mezinárodní úsilí!

Problém dobré shody souvisí s dalšími statistickými úlohami, jako je odhad a testování statistických hypotéz, o nichž se bude hovořit zanedlouho [13].

Konstrukce modelu: problém náhodového mechanismu generujícího dané rozdělení

V závislosti na hlavním zájmu badatele se dají níže uvedené studie rozdělit do dvou odlišných kategorií: (a) reálné a (b) teoretické.

(a) Studie reálné kategorie se zabývají otázkami typu: co se asi děje uvnitř určité populace (řekněme galaxií nebo buněk tvořících lidské tělo nebo řidičů autobusů na londýnských ulicích), že pozorujeme právě určité rozdělení veličiny X? Nezřídka je tato vědecká otázka spojená s utilitářskou: Dají se nějak upravit procesy v populaci, aby se odstranil určitý neuspokojující rys rozdělení X?

Příkladem studia takových reálných situací může být práce G. Udny Yulea, kterou prováděl v 20. letech společně s M. Greenwoodem a E. M. Newboldovou a jež se zabývala nehodami způsobenými řidiči autobusů v Londýně [14]. Základem modelu byla představa, že počet nehod způsobených určitým řidičem za jednotku času je náhodná veličina s Poissonovým rozdělením se střední hodnotou λ, nazvanou „řidičův sklon k nehodám“. Hodnota se mění od řidiče k řidiči podle gamma rozdělení. Podle tohoto modelu je tedy pozorované rozdělení počtu nehod na řidiče a rok gamma směsí Poissonových rozdělení, což je negativně binomické rozdělení. „Utilitářská“ otázka zněla, zda může být pozměněn soubor řidičů (vhodným výběrem) tak, aby se zmenšil počet nehod. Pod ní se skrývá vědecká otázka, zda se hypotetické λ každého daného řidiče v čase znatelně mění či nikoli.

V pozdější době vzniklo a bylo řešeno obrovské množství podobných úloh týkajících se tvorby modelu. Uvádím několik příkladů, úmyslně různorodých:

1. Je mechanismus vývoje rakoviny jednostupňový nebo vícestupňový? [15]
2. Spočívá mechanismus pozorovaného uzavřeného sdružení galaxií v gravitačním „chycení“ „bludné“ galaxie (Lamaitre) nebo v kataklysmických explozích uvnitř jader velmi velikých galaxií (např. Ambartsumian) [16]?
3. Závisí nápadná rozdílnost v množství dešťových srážek na daném místě (vzpomeňte na biblických „sedm úrodných let“ a po nich následujících „sedm hladových let“) na variabilitě počtu částic, které jsou zárodky ledových krystalků v mracích? (Jestliže ano, dala by se zmírnit krutá sucha vypouštěním částic do mraků? [17]).

(b) Alternativní přístup ke konstrukci modelu je charakteristický pro vědce, kteří mají především matematické zájmy a schopnosti, jsou však zároveň schopni zajímat se o reálné otázky toho či onoho druhu. Jejich myšlení postupuje zhruba takto: Existuje jev, řekněme nákaza, jenž dělá starosti lidem, kteří se zabývají reálnými situacemi. Také však je znám náhodový mechanismus M, který nabízí jisté možnosti modifikace. Otázkou je, zda nějaká modifikace M může produkovat rozdělení veličiny X srovnatelné s tím, jež charakterizuje skutečný jev, jako jsou nehody u řidičů autobusů, případy chřipky atd. Obecněji, mohou modifikace M vytvářet rozdělení charakterizovaná všemi Pearsonovými křivkami, které vznikly, aby vyhovovaly překvapivé rozmanitosti empirických rozdělení? Právě tím se zabývá vynikající v podstatě pravděpodobnostní studie G. Pólyy z r. 1930 [18]. Pólya tehdy působil v Curychu, odtud později přešel do Stanfordu.

Zajímavé je, že jeden z Pólyových modelů „nákazy“, předpokládající, že nehody způsobené v minulosti zvyšují pravděpodobnosti více nehod v budoucnosti, poskytoval negativně binomické rozdělení počtu nehod na řidiče za jednotku času, shodné s rozdělením v modelu Yulea, Greenwooda a Newboldové. Posledně jmenovaný model se dá nazvat „smíšený beznákazový mechanismus“. Tento speciální Pólyův výsledek demonstroval nepředvídaný jev — dva úplně odlišné stochastické mechanismy mohou produkovat totožná rozdělení téže veličiny X. Studium takového rozdělení tedy nemůže zodpovědět otázku, který ze dvou mechanismů se uplatňuje ve skutečnosti. Nyní se tento jev označuje jako *neidentifikovatelnost.* V mnoha případech, zejména v ekonometrii a v psychologické i jiné faktorové analýze, se neidentifikovatelnost projevila jako nejvíce zavádějící a podnítila mnoho matematických studií. Naštěstí se shledalo, že systém náhodových mechanismů, které jsou neidentifikovatelné pomocí rozdělení jedné veličiny X, mohou se stát identifikovatelnými pomocí sdruženého rozdělení veličiny X a ještě některých jiných veličin $Y, \ldots, Z$. V ekonometrii dosáhl několika velice zajímavých výsledků týkajících se identifikovatelnosti O. Reiersøl z Oslo [19].

Po druhé světové válce výzkum stochastických modelů značně pokročil jak z hlediska studia reálných situací, tak z hlediska matematického. Ilustrují to termíny běžné v matematické literatuře, obvykle však vypůjčené z biologie nebo z fyziky jako procesy rození a zániku, větvící se procesy, procesy obnovy atd. Výborným zdrojem informací o těchto otázkách je kniha Williama Fellera [1]. Jiná pěkná kniha od Theodora E. Harrise [20] poskytuje bohatou informaci jak o skutečných zdrojích rozmanitosti matematických pojmů, tak o matematických výsledcích celé mezinárodní plejády vynikajících autorů.

Když se vytváří stochastický model přírodního jevu, vznikají hned dva různé problémy, oba spojené s otázkou, zda je model realistický či nikoli.

Jedním z problémů je, jak nejlépe přizpůsobit model pozorováním, která jsou k dispozici. Model obvykle poskytuje pouze systém rozdělení, která veličina X musí mít, ale ne hodnoty parametrů (tzv. rušivých), které identifikují členy tohoto systému. Tak na příklad ačkoli modely sklonu k nehodám, vypracované Yulem a kol. a také Pólyou, vedly k závěru, že počet nehod X se musí řídit negativně binomickým rozdělením, dva parametry a, b tohoto rozdělení nejsou určeny žádným z modelů. Vzniká problém odhadu těchto parametrů z dat: kterých funkcí $\hat{a} = \hat{a}(X_1, \ldots, X_n)$, $\hat{b} = \hat{b}(X_1, \ldots, X_n)$ při daných pozorováních $X_1, \ldots, X_n$ by se mělo užít, aby se dostaly odhady $\hat{a}$, $\hat{b}$ v nějakém smyslu nejlepší?

Druhý problém, týkající se testování statistické hypotézy, ihned následuje. I když je hypotetický model shodný s modelem, který skutečně „funguje" v pozadí studovaného jevu, a i když odhady všech rušivých parametrů jsou prosty chyb (což je více, než můžeme doufat), náhodná proměnlivost obsažená ve shromažďování dat dá vzniknout neshodám mezi empirickým a teoretickým rozdělením. Naopak ovšem může být model chybně zkonstruován a odhalení této skutečnosti má zásadní význam.

Máme tedy dva problémy matematické teorie statistiky: problém bodového odhadu a problém testu statistické hypotézy (tj. testu hypotézy o rozdělení veličiny X).

První správná formulace úlohy bodového odhadu náleží Laplaceovi. Poněkud více uspokojující formulaci podal později Gauss [21]. Oba případy uvažují pevnou neznámou hodnotu Θ parametru a jistý počet n jejích měření X_i, jež všechna podléhají náhodné chybě. V obou případech je také zahrnuta formulace toho, co nyní nazýváme ztrátovou funkcí $L(\hat{\Theta}, \Theta)$. Předpokládá se o ní, že představuje pokutu, kterou musí zaplatit statistik, jenž přijme $\hat{\Theta}$ za odhad neznámého parametru, když jeho skutečná hodnota je Θ. Stručně řečeno úlohou bylo určit $\hat{\Theta}$ jako takovou funkci pozorovatelné veličiny X, která minimalizuje střední hodnotu $L(\hat{\Theta}, \Theta)$. Laplace užil absolutní hodnoty rozdílu $L_L(\hat{\Theta}, \Theta) = |\hat{\Theta} - \Theta|$, zatímco Gauss dal přednost jejímu čtverci $L_G(\hat{\Theta}, \Theta) = (\hat{\Theta} - \Theta)^2$. Výsledkem byla známá teorie nejmenších čtverců. Díky své jednoduchosti a intuitivní přitažlivosti byla metoda nejmenších čtverců rychle přijata „konzumenty" teorie a stále se ještě hojně používá. Na teoretický základ se ztrátovou funkcí L se však téměř na století zapomnělo. Zájem o ně oživil až začátkem 20. století v Rusku Markov [22] a méně výrazně Edgeworth v Anglii. Dnes některé učebnice statistiky obsahují Gaussovu-Markovovu větu o nejmenších čtvercích.

Vývoj teorie testování statistických hypotéz byl pomalejší. Pokud vím, první pokus o test statistické hypotézy je přisuzován Laplaceovi [5]. Jak jsem se zmínil výše, Laplace uvažoval možnost, že komety nejsou řádní členové sluneční soustavy, ale „vetřelci" z vnějšího prostoru. Rozvážil, že je-li tomu tak, pak úhly mezi orbitálními rovinami a ekliptikou budou rovnoměrně rozdělené mezi nulou a $\pi/2$. To byla hypotéza, kterou chtěl Laplace testovat. K tomu *přijal* za kritérium aritmetický průměr pozorovaných úhlů $\bar{\Phi}$. (Slovo přijal je psáno kurzívou, aby se zdůraznil fakt, že průměr nebyl odvozen, aby byl optimální, jako tomu bylo v případě bodového odhadu.) Dále Laplace *odvodil* rozdělení průměru $\bar{\Phi}$ na základě testované hypotézy a přiřadil mu skutečnou hodnotu vypočítanou pro několik tehdy pozorovaných komet. Výsledkem tohoto porovnání bylo, že se Laplace rozhodl pracovat dále za předpokladu, že komety nejsou řádnými členy sluneční soustavy jako planety.

Čtenář si povšimne, že při konstrukci testu své hypotézy musel Laplace řešit speciální úlohu – problém rozdělení svého kritéria $\bar{\Phi}$. Problémy rozdělení stále provázejí úlohy odhadu i úlohy testování hypotéz. Pearsonův χ^2 test dobré shody také vyžadoval řešení problému rozdělení. K. Pearson jej vyřešil pro případ, kdy rozdělení, o němž se předpokládá, že odpovídá datům, je úplně určeno a neobsahuje žádné neznámé parametry. Původně Pearson sám a mnoho jeho následovníků myslelo, že přítomnost parametrů, jež je třeba odhadnout, neovlivní rozdělení χ^2. Udny Yule se však při užití simulovaných výběrů přesvědčil o opaku [23]. Řada zkoumání, počínající prací R. A. Fishera [24], potvrdila, že Yuleova domněnka je oprávněná – bylo zjištěno, že rozdělení χ^2 závisí na tzv. počtu stupňů volnosti a na metodě, které je užito pro odhad neznámých parametrů.

Metoda odhadu, kterou zavedl a systematicky rozvíjel zejména se zřetelem na své četnostní křivky K. Pearson, byla metoda momentů [7]. Alternativní metodu, metodu maximální věrohodnosti [25], prosazoval R. A. Fisher, což dalo vzniknout okázalému sporu. Obě metody byly rozvíjeny na intuitivním základě a poněkud dogmaticky. Fisherova idea byla založena na věrohodnostní funkci, již nazýval „nová míra důvěry nebo nedůvěry". Odůvodnění samo však nebylo dogmatické, ale bylo založeno na přesnosti odhadu.

Problém odhadu byl předmětem rozsáhlých studií Fisherových, který zavedl množství plodných pojmů, jako je konzistence odhadu a jeho vydatnost, postačitelnost a množství informace. Tyto pojmy upoutaly pozornost mnoha vědců. V Americe byli mezi prvními Harold Hotelling [26] a J. L. Doob [27]. Tytéž pojmy se stále ještě zkoumají z mnoha hledisek v poněkud upravené formě.

Teorii testování statistických hypotéz založili na počátku 30. let Egon S. Pearson a autor tohoto článku [28]. Řečeno moderní terminologií, jde o problém dvou rozhodnutí – při dané hypotéze H, týkající se rozdělení pozorování X, a při dané alternativě $\bar{H}$ uvažujeme vzhledem k H dvě možná rozhodnutí: přijmout předpoklad, že H neplatí (tedy že $\bar{H}$ je správná), nebo od toho upustit. Každé z těchto rozhodnutí může být nesprávné a důsledky těchto dvou chyb mohou mít různou důležitost. Chybou prvního druhu se nazývá ta z chyb, o níž se praktický statistik domnívá, že vyhnout se jí je důležitější (což je subjektivní úsudek). Prvním požadavkem kladeným na matematickou teorii je, aby odvodila testové kritérium, které by zajistilo, aby se pravděpodobnost chyby prvního druhu rovnala (nebo přibližně rovnala nebo nepřekročila) předem danému číslu α, např. $\alpha = 0{,}05$ nebo 0,01 apod. Toto číslo se nazývá hladina významnosti.

Když je určena třída $K(\alpha)$ testů, z nichž všechny zajišťují tutéž nízkou frekvenci α chyb prvního druhu, je nutné uvažovat možnost, že dojde k chybě druhého druhu. Stručně řečeno, matematická úloha spočívá v určení testu uvnitř třídy $K(\alpha)$, který minimalizuje pravděpodobnost chyby druhého druhu. Základní pojmy teorie jsou silofunkce testu, podobné oblasti, nejsilnější testy, stejnoměrně nejsilnější testy atd.

Pro mnoho případů, nyní většinou považovaných za knižní, na které patrně nenarazíme ve statistických problémech „ze života", s nimiž se setkáváme v moderních vědeckých studiích, vytvořila raná teorie požadované stejnoměrně nejsilnější testy. V jiných případech se shledalo, že takové testy neexistují, což otevřelo dveře rozmanitým „kompromisním" definicím optimality, počínaje nevychýlenými testy.

Nová forma problému odhadu, na niž upozornil autor tohoto článku, je odhad pomocí intervalu nebo obecněji množiny. Předpokládejme, že rozdělení pozorovatelné veličiny X (obvykle vektoru) závisí na parametru Θ, který je neznámý až na to, že musí ležet uvnitř určitého intervalu, třeba od nuly do jedné apod. Úkolem je přiřadit každé možné hodnotě x veličiny X interval $S(x/\gamma)$ možných hodnot Θ splňující podmínku, že $S(X/\gamma)$ pokrývá skutečnou hodnotu Θ s „velkou" pravděpodobností rovnou (nebo přibližně rovnou nebo aspoň tak velkou) předem danému číslu γ, tak blízkému 1, jak je žádoucí, řekněme $\gamma = 0{,}95$ nebo 0,99 apod. Takto vybrané číslo γ se nazývá koeficient spolehlivosti.

Když je třída takových intervalů („intervalů spolehlivosti") určena, je také nutno vybrat jeden z nich, který vyhovuje nějaké jasné podmínce optimality. V závislosti na okolnostech optimální interval spolehlivosti může být nejkratší (v určitém definovaném smyslu). Není to však obecné pravidlo. V mnoha vědeckých a technických problémech je například důležité být si jist, že hodnota neznámého Θ nepřekročí číslo $\bar{\Theta}(x)$, vypočítané z pozorování X. Jedním z příkladů tohoto druhu je případ, kdy Θ označuje počet škodlivých baktérií v jednotkovém objemu pitné vody. Zde interval spolehlivosti sahá od nuly do $\bar{\Theta}(x/\gamma)$. Užití $\bar{\Theta}(x/\gamma = 0{,}99)$ zajišťuje, že tvrzení, že hustota baktérií nepřekročí $\bar{\Theta}(x/\gamma = 0{,}99)$, bude pravdivé asi v 99% případů.

Přestože první publikace týkající se intervalů spolehlivosti pocházejí z roku 1930 [29], definitivní teorie byla vytvořena až ve dvou článcích z let 1937 a 1938 [30, 31]. Později se teorie testování hypotéz a odhadu pomocí množin stala speciálním případem obecné teorie statistických rozhodovacích funkcí, vybudované Abrahamem Waldem [32].

Problém plánování experimentů

Snad největším činem R. A. Fishera bylo zavedení a rozvoj teorie experimentování s proměnlivým materiálem. Tento čin je tak významný, že si vyžaduje aspoň stručnou zmínku. Podle mého názoru nejzákladnější z četných Fisherových myšlenek je, že experiment s proměnlivým materiálem musí být znáhodněný, aby byl spolehlivý. Význam tohoto pojmu je následující: Předpokládejme, že je pokus navržen za účelem testování vlivu několika ošetření $T_1, T_2, \ldots, T_s$. Ošetření se mají srovnávat na několika pokusných jednotkách. V zemědělském experimentu to mohou být části pokusného pole. V lékařství mohou být pokusnými jednotkami pacienti se stejnou diagnózou, při studiu počasí dny splňující určité podmínky týkající se počasí atd.

Fisherův princip randomizace vyžaduje, aby studovaná ošetření nebyla pokusným jednotkám přiřazována na základě výběru experimentátora, nýbrž na základě dobře navrženého náhodového mechanismu. Je to proto, že zřejmé rozdíly mezi ošetřeními, jak jsou vidět z pokusu, mohou být ve skutečnosti způsobeny vnějšími příčinami a nikoli skutečnými vlastnostmi ošetření. Jeden z častých případů odchylek tohoto druhu je, že experimentátoři mají citový vztah k určitým ze studovaných ošetření a možná podvědomě mají snahu přiřadit preferovaná ošetření pokusným jednotkám, které se zdají být v nějakém smyslu lepší. Častým výsledkem takových pokusů je nejprve sebeklam a potom klamání druhých.

Znáhodnění pokusu může být „neomezené" nebo může vyhovovat rozmanitým omezením. Spolu s Fr. Yatesem vytvořil Fisher množství prostředků pro efektivní randomizaci. Zde si zaslouží zmínku tři knihy: *Statistické metody pro výzkumné pracovníky* [33] a *Plánování experimentů* [34] od Fishera a *Statistické tabulky pro biologický, zemědělský a lékařský výzkum* [35] od Fishera s Yatesem, všechny mnohokrát vydané.

R. A. Fisher koncem 20. let

Všechny tři knihy ovlivnily experimentování ve všech oblastech. Obvykle však jejich vliv nebyl okamžitý. Dost často by se daly první reakce experimentátorů vyjádřit slovy: „Vlezte mi s tím Fisherem na záda; vím o experimentování a o svém materiálu všechno, co je třeba." Takové protesty se ozývaly zvláště proti randomizaci. Časem se však přístup změnil.

Rozšíření pluralistických studií ve vědě a technice jako zdroj inspirace pro rozvoj matematické teorie pravděpodobnosti a statistiky

Na začátku tohoto článku bylo poznamenáno, že ač nyní teorie pravděpodobnosti a statistika dosáhly dospělosti a žijí „svými vlastními životy", stále se vyvíjejí a obměňují v mnoha směrech. Důvodem je stále rostoucí počet reálných problémů v udivujícím množství rozmanitých oblastí, které skoro stále obsahují problémy matematické. V předcházející části jsem se stručně zmínil o deseti příkladech:

sklon řidičů autobusů k nehodám,
rozmanité problémy astronomie [36],
pokusy vyvolat déšť,
bakteriální nakažení pitné vody,
ekonometrie,
experimentování s variabilním materiálem ve všech oblastech,
faktorová analýza v psychologii,
mechanismus zhoubných nádorů,
lékařská diagnostika,
dopravní inženýrství.

K tomuto seznamu je vhodné přidat další soupis nových oblastí teoretických a praktických činností. Několik následujících příkladů jako ilustrace jistě postačí.

1. Populační genetika; od G. H. Hardyho („čistého matematika" z Cambridge, který zaslal v roce 1908 otevřený dopis časopisu Science) po R. A. Fishera, Sewalla Wrighta a moderní neodarwinistické studie vývoje [37].
2. Technika; od W. A. Shewharta a kontroly jakosti po přejímací postupy [38], operační výzkum a teorii spolehlivosti.
3. Studie týkající se znečištění ovzduší a veřejného zdraví [39].
4. Demografie; od práce Alfréda Lotky [40] ve Spojených státech, týkající se reálné situace, a vysoce „matematického" díla Vita Volterry v Itálii [41] ve 20. a 30. letech po novodobé mezinárodní starosti s přelidněním.

To jsou tedy příčiny stálého širokého výzkumu i pedagogického úsilí ve dvou spolu souvisejících odvětvích, pravděpodobnosti a statistice.

Literatura

[1] FELLER, W.: *An introduction to probability theory and its applications.* Wiley, New York, 1950, 1967.

[2] BERNOULLI, J.: *Ars conjectandi.* 1713; přetištěno ve *Wahrscheinlichkeitsrechnung,* Engelmann, Leipzig 1899.

[3] KOLMOGOROV, A. N.: *Grundbegriffe der Wahrscheinlichkeitsrechnung.* J. Springer, Berlin 1933.

[4] MISES, R. VON: *Probability, statistics and truth.* Allen and Unwin, London 1957.

[5] LAPLACE, P. S.: *Théorie analytique des probabilitiés.* Académie Française, Paris 1812.

[6] FECHNER, G. T.: *Kollektivmasslehre.* G. R. Lipps ed., Engelmann, Leipzig 1897.

[7] KENDALL, M. G.: *The advanced theory of statistics 1, 2.* Griffin, London 1947.

[8] PEARSON, K.: *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling.* Phil. Mag. Ser. *V 50,* (1900) 157—175.

[9] CRAMÉR, H.: *On the composition of elementary errors.* First paper: *Mathematical deductions.* Skand. Aktuarietidskr. *11* (1928), 13—74.

[10] MISES, R. VON: *Wahrscheinlichkeitsrechnung und ihre Anwedung in der Statistik und theoretischen Physik.* Deuticke, Leipzig 1931.

[11] KOLMOGOROV, A. N.: *Sulla determinazione empirica de una leggi di distribuzione.* Giorn. Inst. Ital. Attuari *4* (1933), 83—91.

[12] SMIRNOV, N. V.: *Ob otklonenijach empiričeskich funkcij raspredelenija.* Mat. Sbornik *6* (1939) 3—26.

[13] NEYMAN, J.: *„Smooth“ test for goodness of fit.* Skand. Aktuarietidskr. *20* (1937), 149—199.

[14] YULE, G. U., GREENWOOD, M.: *An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurence of multiple attack of disease or of repeated accidents.* J. Roy. Statist. Soc. *83* (1920), 255—279.

[15] NEYMAN, J., SCOTT, E. L.: *Statistical aspects of the problem of carcinogenesis.* Pro. 5th Berkeley Symp. Math. Statist. Prob. *4* (1967), 745—776.

[16] ZONN, W.: *Explosive events in the universe. The Copernican Heritagee Theories „Pleasing to the Mind“.* ed. J. NEYMAN, MIT Press, Cambridge, Mass., 1974.

[17] NEYMAN, J.: *Experimentation with weather control.* J. Roy. Statist. Soc. *A 130,* (1967), 285—326,

[18] PÓLYA, G.: *Sur quelques points de la théorie des probabilités.* Ann. Inst. Henri Poincaré *1* (1930), 117—161.

[19] REIERSØL, O.: *Identifiability of a linear relation between variables which are subject to error.* Econometrica *18* (1950), 375—389.

[20] HARRIS, T. E.: *The theory of branching processes.* Springer-Verlag, Berlin 1963.

[21] GAUSS, C. F.: *Abhandlungen zur Methode der kleinsten Quadrate.* Stankiewicz, Berlin 1887.

[22] MARKOV, A. A.: *Wahrscheinlichkeitsrechnung.* Teubner, Leipzig 1912.

[23] YULE, G. U.: *An application of the χ^2 method to association and contingency tables, with experimental illustrations.* J. Roy. Statist. Soc. *85* (1922), 95—104.

[24] FISHER, R. A.: *The conditions under which χ^2 measures the discrepancy between observation and hypothesis.* J. Roy. Statist. Soc. *87* (1924), 442—450.

[25] FISHER, R. A.: *On mathematical foundations of theoretical statistics.* Phil. Trans. Roy. Soc. (London) Ser. *A 222* (1921), 309—368.

[26] HOTELLING, H.: *The consistency und ultimate distribution of optimum statistics.* Trans. Amer. Math. Soc. *32* (1930), 847—859.

[27] DOOB, J. L.: *Probability and statistics.* Trans. Amer. Math. Soc. *36* (1934), 759—775.

[28] NEYMAN, J., PEARSON, E. S.: *On the problem of the most efficient tests of statistical hypotheses.* Phil. Trans. Roy. Soc. (London) Ser. *A 231* (1933), 289—337.

[29] PYTKOWSKI, W.: *The dependence of the income in small farms upon their area, the outlay and the capital invested in cows* (polsky). Monografie č. 31 řady Biblioteka Pulawska, Polsko 1932.

[30] NEYMAN, J.: *Outline of a theory of statistical estimation based on the classical theory of probability.* Phil. Trans. Roy. Soc (London) Ser. *A 236* (1937), 333—380.

[31] NEYMAN, J.: *L'estimation statistique traitée comme un probléme classique de probabilité.* Actual. Scient. Indust. *739* (1938), 25—57 (ruský překlad Usp. Mat. Nauk *10* (1949), 207—229).

[32] WALD, A.: *Statistical Decision Functions.* Wiley, New York 1950.

[33] FISHER, R. A.: *Statistical methods for research workers.* Oliver and Boyd, London (1. vyd. 1925, 12. vyd. 1954).

[34] FISHER, R. A.: *The design of experiments.* Hafner, New York (1. vyd. 1935, 8. vyd. 1966).

[35] FISHER, R. A., YATES, F.: *Statistical tables for biological, agricultural and medical research.* Hafner, New York (1. vyd. 1938, 6. vyd. 1963).

[36] NEYMAN, J., SCOTT, E. L.: *Field galaxies and cluster galaxies: abundances of morphological types and corresponding luminosity functions. Confrontation of cosmological theories with observation,* ed. M. S. LONGAIT, D. Reidel Publishing Co. Dordrecht (1974), 119—130.

[37] Proc. Sixth Berkeley Symp. Math. Statist. Prob. 5: *Darwinian, Neo-Darwinian and Non-Darwinian Evolution.* Univ. of California Press, Berkeley 1972.

[38] BOWKER, A.: *Sampling inspection by variables.* McGraw-Hill, New York 1952.

[39] Proc. Sixth Berkeley Symp. Math. Statist. Prob. 6: *Effects of pollution on health.* Univ. of California Press, Berkeley 1972.

[40] LOTKA, A. J.: *Elements of physical biology.* Williams and Wilkins, Baltimore 1925 (reedice Dover, New York 1956).

[41] VOLTERRA, V.: *Leçons sur la théorie mathématique de la Lutte pour la vie.* Gauthier-Villars, Paris 1931.