

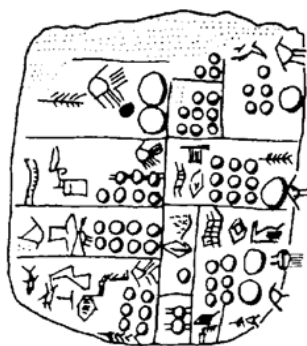
Stručný výtah z počátků statistiky a pravděpodobnosti

01DEM

Pavel Hrabák

1 Strípky z dějin popisné statistiky

Jak je běžným zvykem u podobných textů, i zde tvrdíme, že již staří Sumerové položili základy popisné statistiky. Nejstarší písemné památky se statistickou povahou skutečně pochází z oblasti Sumeru, Jednalo se o záznamy o časových intervalech, počtech osob a kusů domácího zvířectva a úrodě. Jeden z takových záznamů vidíme na obrázku 1 psaném protosumerským písmem.



Obrázek 1:

Potřeba zaznamenávat počty a stav obyvatelstva společně s majetkem jednotlivých lidí pramení z nutnosti výběru daní a to ve formě naturálních či peněžních, nebo ve formě pracovních povinností obyvatelstva. Pro tyto potřeby byly vypracované různé "statistické metodiky". Z četných písemných záznamů jsme nyní schopni si udělat představu o systému tehdejší státní správy a skladbě obyvatelstva starověkých říší počínaje mezopotámskými městskými státy.

Ve starověkém Egyptě je od roku 2850 př.n.l. pravidelně jednou za dva roky prováděn soupis dobytka a od roku 2000 př.n.l. je vybírána rovná daň z hlavy, což je spojené se zavedením aktuálního sčítání lidu. Již poměrně sofistikované metody byly používány při výpočtu daně z půdy, jejíž výše vychází ze zaměřené rozlohy a vydatnosti záplav na daném místě.

Ve starověkém Římě v republikánském období byl pravidelně prováděn tzv. census. Jednalo se o propracovaný systém evidence obyvatelstva. Kromě sčítání lidu byl prováděn soupis nemovitého majetku, později otroků a dobytka. Na základě těchto informací byla pak vypočtena výše daně pro jednotlivé občany. Jedním z posledních pravidelných censů byl proveden za císaře Augusta kolem přelomu letopočtu a je spojován s narozením Ježíše Krista.

V období raného středověku, kdy mezi lidem byla poměrně vysoká negramotnost, se ohnisky vzdělanosti staly církevní instituce, které vedly evidenci svého majetku a jeho změn. Členové církevních řádů byli zaměstnáváni aristokracií právě za tímto účelem. Nejstarší dochovaná písemnost tohoto typu na našem území je soupis majetku litoměřického kostela z roku 1058 jako součást zakládací listiny vydané knížetem Spytihněvem II.

Jsou vydávány tzv. vrchnostenské urbáře obsahující evidenci příjmů poddaných, pozemkové vlastnictví šlechty a církve, městské berní knihy a berní rejstříky. Ve 14. století se v Evropě objevují první církevní matriky, které jsou dnes cenným zdrojem informací o složení a migraci obyvatelstva v tehdejší době. V roce 1583 vypukla v českých zemích epidemie moru. Na základě toho nechal Rudolf II. zahájit šetření o zdraví populace, které mapovalo vznik a rozvoj zhoubných epidemií a mělo umožnit včasná protipatření.

16. století lze považovat za období rozvoje popisné statistiky v novodobé Evropě. Rozvíjí se společně se státovědou a teoriemi o řízení státu. V roce 1589 použil Ital Girolamo Ghilini prvně termín *statistika*. Slovo pochází z latinského *status*, což znamená stav. Tímto pojmem byl myšlen stav státu či země, pojem statistika pak označoval činnost a metody pro zjišťování tohoto stavu. Užití statistiky ve státovědě se intenzivně zabývá kniha *Německý knížecí stát* (1662), která je hojně vydávána i v průběhu 17. a 18. století.

V průběhu 17. století v Anglii vzniká odvětví statistiky nazývané *politická aritmetika*, která vycházela z údajů o narozeních a úmrtích a pokoušela se na jejich základě zkoumat vývoj obyvatelstva v delších časových obdobích. Jedná se tedy o předchůdce demografie. Zakladatelem této disciplíny je John Graunt (1620-1674), který zkoumal matriky novorozených a zemřelých v Londýně. Odhalil poměr mezi počtem mužů a žen v populaci a zjistil, že počet narozených chlapců je vyšší než počet narozených dívek. Do anglické školy patří i Edmund Halley (1656-1742), který při konstrukci úmrtnostních tabulek propojuje statistiku se vznikajícím počtem pravděpodobnosti. Úmrtnostní tabulky se používali pro odhad rizika při pojištění a objevují se i snahy o vybudování všeobecné teorie pojištění a rizik.

Poznamenejme krátce, že slovo riziko (latinsky *resicum*, italsky *risatio*, anglicky *risk*) pravděpodobně pochází z arabského *rizq*, což znamená šance, nebo štěstí. Prvně se objevuje kolem roku 1160 v janovských spisech v souvislosti s námořní dopravou a postupně se rozšířilo do oblastí ručení za půjčky, dozoru na osobní bezpečnost obchodníků a posléze i pojištnictví.

Významnou osobností evropské statistiky byl belgický matematik a astronom Adolphe Lambert Quételet (1796-1874), který vypracoval zásady moderního sčítání lidu, které byly prvně použity v roce 1846 při sčítání lidu v Belgii. Na základě těchto dat vypočítal rozměry průměrného člověka a odchylky jedinců od tohoto průměru. Objevují se již první náznaky pojmů jako je průměr, střední hodnota, rozptyl a rozdělení. V roce 1853 zorganizoval první statistickou konferenci, kde představuje tzv. Queteletův index $QI = (\text{hmotnost v kg})/(\text{výška v m})^2$. Je-li $QI > 30$, je osoba považována za obézní.

2 Zrod pravděpodobnostního počtu

Odhlédneme-li od prací indických a čínských autorů, kteří se řešením kombinatorických úloh zabývali již ve starověku, můžeme položit vznik pravděpodobnosti do 17. století, kdy Blaise Pascal (1623-1662) a Pierre de Fermat (1601-1665) v roce 1654 vyřešili tzv. *úlohu o rozdělení sázky*. Princip úlohy o rozdělení sázky lze popsat následovně. Dva hráči hrají sérii her o určitou částku s tím, že tuto sumu získá ten, kdo jako první dosáhne předem domluveného počtu vítězství. Série je však předčasně ukončena a hráči si chtějí sázku rozdělit podle pravděpodobnosti, že ten který vyhraje. Vyskytují se samozřejmě i verze úlohy pro více hráčů. Úloha o rozdělení sázky vznikla ve středověku, údajně byla objevena již v italském rukopisu z roku 1380. Připouští se, že mohla být i arabského původu.

Je známo několik forem úlohy, vyskytovala se často ve sbírkách příkladů i rukopisech. Prvními známými, nicméně neúspěšnými, řešiteli byli Luca Pacioli a Nicolo Tartaglia. Luca Pacioli ve spise *Summa Arithmetica* z roku 1494 přerušuje míčovou hru na 60 vítězných bodů za stavu 50:20 a lukostřelecké závody na šest vítězství za stavu 4:3:2. Obě řešení jsou však nesprávná. Než přejdeme k Pascalovu a Fermatovu řešení zmíníme ještě jméno Hieronyma Cardana (1501-1576), který v roce 1526 vydává spis *De ludo Aleæ* věnovaný především problémům spadajícím dnes pod teorii pravděpodobnosti. Patří mezi ně i úlohy typu: Kolika způsoby může padnout jistý počet ok při házení určitým počtem kostek. Mezi průkopníky pravděpodobnosti patří i Galileo Galilei (1564-1642), který zkoumal chyby vznikající při fyzikálních měření a považoval je za výsledky náhodných pokusů.

Za opravdový počátek teorie pravděpodobnosti je považována korespondence mezi Pascalem a Fermatem o problémech s kterým se na Pascala obrátil rytíř de Méré. Jednalo se o úlohu o rozdělení sázky a otázku, kolik hodů jednou či dvěma kostkami je třeba, aby šance, že padne alespoň jedna šestka, respektive dvě šestky, byla nadpoloviční. Na obě úlohy nalezneme odpověď ve zmíněné korespondenci.

Uvažujme úlohu pro dva hráče A, B a nízký maximální počet počet chybějících her (ozn. N). Hráčům chybí m a n výher, tedy $N = m + n - 1$. Najdou se všechny možné kombinace her s opakováním N prvků A, B, a stanoví se poměr dělení banku $p_{A:B}(m, n)$, který je podílem celkového počtu her, které by dali celou sázku hráči A ku počtu her, které by dali celou sázku hráči B. Např. pro $N = 3$ dostaneme AAA, AAB, ABA, BAA, BAB, BBA, ABB, BBB. Je-li $m = 1$ a $n = 3$, vyhrává hráč A v sedmi případech a hráč B pouze v jednu. Dělení hry je pak v poměru $p_{A:B}(1, 3) = 7 : 1$. Obecný návod dal Pascal pomocí Pascalova trojúhelníku. Narozdíl od dnešního úzu psal Pascal trojúhelník s jednou odvěsnou vodorovně a druhou svisle:

1	1	1	1	1	1	1	1
1	2	3	4	5	6	7	
1	3	6	10	15	21		
1	4	10	20	35			
1	5	15	35				
1	6	21					
1	7						
1							

Řešení úlohy pak nalezneme tak, že v trojúhelníku nalezneme diagonálu (v dnešní interpretaci řádek), která obsahuje $n + m$ čísel, sečteme v něm prvních m a posledních n čísel, poměr dělení sázky je pak poměr těchto součtů.

V dopise z 29. 6. 1654 nastiňuje Pascal i druhý přístup. Nejprve uvažuje situaci, kdy k vítězství jsou nutné nejvýše dvě hry ($m = 1, n = 2$). Uvažuje takto: celá sázka by byla moje, kdybych vyhrál a v opačném případě se budeme dělit napůl. Polovina sázky je tedy moje v každém případě a druhou si rozdělíme napůl. Dělení je tedy v poměru 3:1. Případ $m = 1, n = 3$ tímto způsobem vede opět na poměr 7 : 1. Tento princip dělení je však mnohem starší, neboť se jedná o talmudický princip dělení šatstva. Z jistého souboru oděvů nárokuje jeden účastník sporu všechno a druhý polovinu. Podle talmudického řešení první dostane 3/4 a druhý 1/4, neboť jednu polovinu nárokuje jen jeden a druhou polovinu si tedy rozdělí.

Dalším z mužů, kteří přispěli ke zrodu nové matematické disciplíny, byl Christian Huygens (1629-1695). Ten znal téma korespondence mezi Pascalem a Fermatem a v roce 1656 ve spisu *De*

raciociniis in ludo aleæ posunul Pascalovy a Fermatovy úvahy od řešení konkrétních úloh již k poloze obecných pojmů a postupů. Stal se duchovním otcem klasické definice pravděpodobnosti a zavádí střední hodnotu (i když ani jeden z těchto pojmů neužívá, nemluví o pravděpodobnosti, ale o *očekávané výhře*). Tuto teorii dále rozvinuli Abraham de Moivre (1667-1754) a především Jacob Bernoulli (1655-1705), který jako první dokázal tzv. zákon velkých čísel. Ten se objevuje v nedokončeném díle *Ars conjectandi*, kde se také objevuje důkaz pomocí matematické indukce.

Snad nejvýznamnějším klasikem teorie pravděpodobnosti byl Pierre-Simon Laplace (1749-1827). Ve svém díle o teorii pravděpodobnosti *Théorie analytique des probabilités* systematizoval veškeré poznání svých předchůdců a dalekosáhle je rozpracoval i aplikoval na téměř všechny oblasti tehdejšího vědeckého poznání od fyziky až po sociální vědy. Přichází s klasickou definicí pravděpodobnosti. Pravděpodobnost jevu A je dán jako podíl $P(A) = m/n$, kde n je počet všech výsledků náhodného pokusu a m je počet výsledků příznivých jevu A . Poznamejme však, že Laplace toto vnímal jen jako speciální případ kdy všechny výsledky jsou apriori stejně možné. Uvažoval i mnohem obecnější situace. Laplaceovo pojetí pravděpodobnosti je bayesovské - jako zobecnění logiky pro úlohy s neúplnou informací.

V této době dochází ke skloubení dvou matematických disciplín, popisné statistiky a teorie pravděpodobnosti a vzniká nová disciplína, kterou je matematická statistika. Výčet matematiků a osobností, kteří vytvářeli a formovali další vývoj pravděpodobnosti a statistiky, společně s jejich přínosem značně přesahuje rozsah této práce. Nabízí však rozsáhlé pole pro zkoumání vývoje této dnes velice rozšířené vědy, jejíž historie je bohatá na úsměvné historicky jako např. o vzniku studentova rozdělení nebo rozvoj zkoumání distribučních směsí na základě měření délky klepet krabů.

Použitá literatura

- [1] Libor Žák, Historie statistiky a pravděpodobnosti, ÚM FSI v Brně, 2006.
- [2] Ivan Saxl, Pravděpodobnost ve středověku.
- [3] <http://cs.wikipedia.org>