13. kapitola

METODA MAXIMÁLNÍ VĚROHODNOSTI ANEB O TOM, JAK ODHADNOUT POČET VOLNĚ ŽIJÍCÍCH DIVOKÝCH ZVÍŘAT

„Jest zcela nepochybným faktem,
že nemůžeme-li poznat nejpravdivější soudy,
musíme se řídit soudy nejpravděpodobnějšími".
(Descartes, Pojednání o metodě)

Fyzik vyvozuje často různé hypotézy o složení hmoty, struktuře atomu atd. Z těchto teorií jsou některé více pravděpodobné, jiné méně pravděpodobné. Fyzik přijímá hypotézu nejpravděpodobnější jako pravdě nejbližší (nejvěrohodnější). Podobně uvažuje chemik, astronom, ekonom. Princip, o nějž se opírají jejich úvahy, je prostý: Nejvěrohodnější je to, co je nejpravděpodobnější. Teorie pravděpodobnosti nás učí určovat, počítat pravděpodobnost jevu. Teorie pravděpodobnosti umožňuje odhadnout šanci pro realizaci toho jevu, a tedy v jistém smyslu předpovídat budoucnost. Mnohem větší význam má však jiné „předpovídání budoucnosti". Opírá se o následující zásadu:

Existují-li v praxi dva jevy, první s velkou, druhý s malou pravděpodobností, budeme očekávat, že nastane jev s velkou pravděpodobností. Můžeme věřit, že jev s malou pravděpodobností nenastane. Podstata tohoto postupu, zvaného metoda maximální věrohodnosti, je formulována Descartesovými slovy. Z této metody vyplývá, že v případě dvou jevů, z nichž některý potom nastal (my nevíme který), dáme přednost tomu jevu, jehož pravděpodobnost je větší. Říkáme, že je věrohodnější.

Na metodě maximální věrohodnosti je založena metoda odhadování četnosti populace volně žijících zvířat (např. počet ryb v jezeře, zubrů v pralese, zajíců v určeném prostoru atd.).

Příklad 13.1. V určitém jezeře žije neznámý počet ryb. Abychom tento počet (označme jej n) odhadli, budeme postupovat takto: Vylovíme m ryb, označkujeme je a pustíme zpátky do jezera. Počkáme, až se všechny ryby promíchají, a potom vylovíme r ryb. Toto vylovení r ryb sítí je náhodné vybírání ryb. Jeho výsledky zakódujeme r-prvkovými podmnožinami množiny všech n ryb. Jsou to kombinace r-té třídy z prvků této množiny. Prostor výsledků tedy obsahuje $\binom{n}{r}$ prvků. Je to klasický prostor. Každý výsledek je třeba pokládat za stejně pravděpodobný. Nechť A_n^k označuje jev: Z těchto r náhodně vybraných ryb bude právě k ryb označkovaných. Jevu A_n^k je příznivých tolik výsledků, kolik je možných způsobů vylovení právě k kusů označkovaných ryb při vylovení r ryb. Tento počet se rovná

$$\binom{n}{k}\binom{n-m}{r-k},$$

kde $n - m$ je počet neoznačených ryb, $\binom{n}{k}$ je počet způsobů, jak vybrat k označených ryb, $\binom{n-m}{r-k}$ je počet možností, jak k nim přidat zbývajících $r - k$ neoznačených ryb.

Z věty o klasickém prostoru vyplývá, že

$$P(A_n^k) = \frac{\binom{m}{k}\binom{n-m}{r-k}}{\binom{n}{r}}.$$

Došlo k jevu A_n^k, tj. vylovili jsme právě k označených ryb. Známe r, m a k. Ptáme se, jaké je n. Metoda maximální věrohodnosti říká, že n bude takové, pro které je $P(A_n^k)$ největší. Nyní je tedy třeba vyhledat největší člen posloupnosti $P(A_n^k)$. Je to posloupnost s kladnými členy. Maximum můžeme vyhledat tak, že zkoumáme, jaký je podíl n-tého členu se členem předcházejícím vzhledem k 1. Doporučujeme vám propočítat si to. Odpověď zní: $P(A_n^k)$ je největší pro takové n, které splňuje podmínku $\frac{m}{n} = \frac{k}{r}$. Dal se takový výsledek předpokládat? Popsaná metoda odhadování počtu ryb patří k metodám postupného odchytu s pouštěním. První ji použil Lincoln v r. 1930 k odhadu počtu volně žijících divokých zvířat.

Úloha 13.1. K odhadu počtu kaprů v rybníce bylo chyceno 300 kusů, označkováno a puštěno zpět do rybníka. Po promísení ryb bylo chyceno znovu 200 kaprů, mezi nimiž bylo 50 předtím označkovaných. Jaký je odhad počtu kaprů v tomto rybníku?

Příklad 13.2. Továrna vyprodukovala sérii n kusů určitého zboží (např. žárovek, praček, televizorů, konzerv apod.). Než se toto zboží dostane na pulty obchodů, je třeba tuto sérii zboží podrobit kontrole jakosti. Mluvili jsme o tom již v odst. 5.4. Předpokládejme, že série má n kusů a že náhodný reprezentativní vzorek jsme vybrali pomocí náhodného výběru bez vracení. Náhodně vybraných r kusů zboží podrobíme důkladné kontrole. Nechť mezi těmito r kusy je právě k kusů vadných. Předpokládejme, že v celé sérii n kusů je m kusů vadných. Označme B^k_m jev, že těch k vadných kusů bylo vybráno ze série n kusů, která má m kusů vadných. Obdobně jako v předcházejícím příkladě jsme hledali takové n, při kterém $\mathsf{P}(A^k_n)$ byla maximální, musíme nyní nalézt m, pro které je $\mathsf{P}(B^k_m)$ největší.

Úloha 13.2. Dokažte, že $\mathsf{P}(B^k_m)$ nabývá maxima pro m, které vyhovuje vztahu $\dfrac{m}{n} = \dfrac{k}{r}$.

Na základě metody maximální věrohodnosti jsme určili, že nejpravděpodobnějším počtem vadných kusů v této sérii zboží je číslo m, vyhovující vztahu $\dfrac{m}{n} = \dfrac{r}{k}$.

Je to výsledek, který jsme očekávali. Vyjadřuje, že poměr počtu m vadných kusů k n, což je počet všech kusů, je stejný jako poměr počtu k vadných kusů v náhodně vybraném vzorku k počtu r všech kusů ve vzorku. K takovému závěru nás jistě vedla i naše intuice. Náhodný výběr kusů pro reprezentativní vzorek zaručuje, že se vzorek „podobá“ celé populaci.

Význam zde prakticky popsané metody je veliký. Nechť tyto skromné příklady potvrdí, jak značná je úloha matematiky a zvláště teorie pravděpodobnosti.

Naše pátrání zakončeme citátem: „Nejpřesvědčivější argumenty pro to, jakou hodnotu matematika skutečně má, poskytuje teorie pravděpodobnosti.“ To jsou slova slavného holandského didaktika matematiky Hanse Freudenthala.