Pravděpodobnost v aplikačních úlohách

MAGDALENA HYKŠOVÁ

Fakulta dopravní ČVUT, Praha

Článek volně navazuje na statě [3], [5] a [6], které v minulých ročnících tohoto časopisu představily některé části sbírky aplikačních úloh z matematiky, jež vznikla na Katedře didaktiky matematiky MFF UK v Praze a v současné době je v nakladatelství Prometheus připravována pro tisk.[1]

Autoři samozřejmě nezpochybňují, že matematika jako vyučovací předmět by měla především rozvíjet logické myšlení. Většinu studentů je však nutné ke studiu této disciplíny výrazně motivovat a pomoci jim zbavit se nejrůznějších předsudků a obav. A právě to je hlavním cílem *Sbírky aplikačních úloh ze středoškolské matematiky*, která se snaží ukázat, že matematika je užitečným a nepostradatelným nástrojem pro řešení skutečných problémů z reálného světa. Sbírka je určena především studentům a učitelům středních škol, ale také všem zájemcům o řešení úloh z praxe.

Sbírka začíná kapitolou věnovanou základním matematickým poznatkům s důrazem na finanční matematiku, procenta a poměry, následující kapitola je zaměřena na úlohy vedoucí na rovnice, nerovnice a jejich soustavy a úlohy související s funkcemi a posloupnostmi. Třetí kapitola obsahuje úlohy geometrické týkající se planimetrie, stereometrie a trigonometrie. Sbírku pak uzavírá kapitola věnovaná pravděpodobnosti; několik úloh z této části bych ráda představila v tomto článku.

Úlohy věnované pravděpodobnosti

Podstatná část studentů považuje pravděpodobnost za nejméně zajímavou, užitečnou a pochopitelnou partii matematiky, která řeší problémy vzdálené od reálného života. Přitom je tomu právě naopak – s pravděpodobností se každý z nás setkává téměř neustále, ať už hovoříme o počasí, dopravních nehodách, nemocích, průzkumech veřejného mínění, finančních

[1] Autory sbírky jsou J. Hromadová, M. Hykšová, O. Odvárko, P. Pavlíková, J. Robová a A. Slavík.

trzích, detekci spamů, hazardních hrách, vadách materiálů, složení hornin, o rostlinách, tkáních či samotných lidských bytostech aj. Teorie pravděpodobnosti si tedy zaslouží, aby jí byla věnována větší pozornost a aby byl co nejvíce využit její potenciál pro motivaci výuky matematiky i pro všeobecnou finanční gramotnost. Mladí lidé by si ze školy měli odnést alespoň základní „pravděpodobnostní“ uvažování, aby nepodlehli hráčské vášni tam, kde by stejně nemohli vyhrát, a aby si uměli vytvořit správnou představu o údajích, jimiž je budou zahrnovat politici, firmy provádějící průzkumy veřejného mínění, lékaři, farmaceutické firmy, biologové či provozovatelé heren, kasin a loterií.[2]

1. Hazardní hry

První část kapitoly věnované pravděpodobnosti se zabývá hazardními hrami a kurzovými sázkami. Na jednoduché úloze týkající se loterie je nejprve objasněn pojem *střední* či *očekávané hodnoty* určité veličiny, např. výhry či zisku. Potom jsou zkoumány různé hazardní hry (hrací automaty, keno, ruleta a různé hry s kostkami) a vždy je ukázáno, že střední hodnota zisku hráče je záporná – to znamená, že jediný, kdo na hře z dlouhodobého hlediska může vydělat, je provozovatel herny či kasina.

Tyto úvahy jsou zároveň přirozenou motivací pro připomenutí skutečnosti, že zisk majitelům výherních zařízení zaručuje *zákon velkých čísel*, který lze zjednodušeně zformulovat takto: Opakuje-li se daný náhodný pokus dostatečně mnohokrát, pak se relativní četnost výskytu určitého jevu blíží jeho teoretické pravděpodobnosti s libovolnou přesností. Budeme-li tedy například dostatečně dlouho házet nevychýlenou mincí, padne panna přibližně v polovině případů. Bude-li v nějaké hře pravděpodobnost výhry v jedné partii menší než 1/2 a hráč u hry vytrvá dostatečně dlouho, pak prodělá a kasino se obohatí. Jediné, co si tedy kasino musí zajistit, je to, aby šance v každé jednotlivé hře byly lehce vychýlené v jeho prospěch. Ne příliš mnoho, aby podmínky hráče neodradily, ale dostatečně na to, aby podnik slušně vydělával.

V této souvislosti ve sbírce rovněž upozorňujeme, že mnoho hráčů má o zákonu velkých čísel poněkud zkreslenou představu a domnívá se, že

[2] Zajímavé motivační příklady lze nalézt v řadě publikací. V českém jazyce jsou dostupné například knihy [1], [4] a [7] autorů J. Anděla, E. a M. Kaplanových a J. S. Rosenthala; z cizojazyčných knih uveďme alespoň knihu [2] G. Gigerenzera. V naší sbírce jsme se snažili především o to, abychom učitelům poskytli materiál bezprostředně použitelný ve výuce. Úlohy jsou řazeny tak, aby studenty postupně seznamovaly se základními pojmy a jejich vlastnostmi a zároveň ukázaly jejich praktické využití.

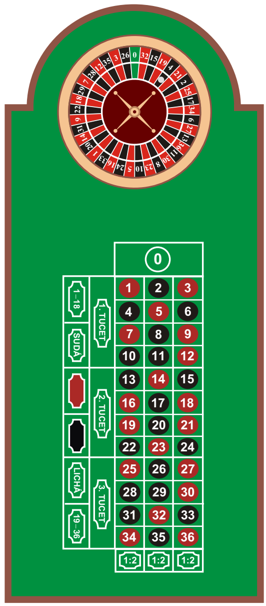
padne-li například desetkrát po sobě červená, musí dříve nebo později padnout častěji černá, protože zákon velkých čísel vyžaduje, aby obě barvy padaly stejně často. Kulička v ruletě se však nedívá zpět, neví, kolikrát za sebou přistála na černém či na červeném políčku, a šance na výhru je stále stejná. Četnosti obou barev se obvykle vyrovnají až v mnohem delším časovém období.

Pro ilustraci zde uveďme jednu řešenou úlohu z této části.

Ruleta

Základem rulety[3] je otáčivé zařízení tvořené dvěma soustřednými koly, z nichž větší je nehybné a menší se otáčí. Vnitřní kolo je rozděleno do 37 políček očíslovaných od 0 do 36; políčko s číslem 0 je zelené, ostatní jsou střídavě červená a černá.[4] Při hře posílá krupiér kuličku po vnějším kole proti směru otáčení vnitřního kola. Kulička postupně zpomaluje, až spadne do vnitřního kola a usadí se na jednom z políček.

Sázka	**Výplatní poměr**
Černá čísla	1 : 1
Červená čísla	1 : 1
Malá čísla (1–18)	1 : 1
Sousední sloupce	1 : 2
První sloupec	2 : 1
Sousední tucty	1 : 2
První tucet (1–12)	2 : 1
Šest čísel (sousední řádky)	5 : 1
Čtyři čísla (tvořící čtverec)	8 : 1
Tři čísla (řádek)	11 : 1
Dvě sousední čísla	17 : 1
Jedno číslo	35 : 1



[3]Ruleta pochází z Francie (název *roulette* znamená doslova *malé kolo*) a v dnešní podobě se hraje přinejmenším od roku 1796.

[4]*Americká ruleta* obsahuje navíc ještě jedno zelené políčko označené jako 00; ve většině evropských kasin se však používá tzv. *francouzské kolo*, které je popsané v textu.

Umísťováním žetonů na herní plán hráč sází na to, kde kulička zůstane ležet po příštím roztočení. Vsadí-li například S Kč na červenou barvu, pak v případě, že kulička skončí na červeném políčku, získá zpět svou sázku a k tomu navíc dalších S Kč (v tomto případě se říká, že sázka je vyplácena v poměru $1 : 1$), zůstane-li však kulička na černém nebo zeleném políčku, hráč o svou sázku přijde. Některé další možnosti sázek jsou uvedeny v následující tabulce. Je-li udán poměr $a : b$, pak v případě výhry hráč dostane ke vsazené částce navíc její (a/b)násobek.

Příklad: Ruleta – zaručený výdělek pro kasina

V kasinu je provozována ruleta popsaná výše.

a) Jaká je pravděpodobnost výhry při sázce na červenou barvu? Jaký bude náš očekávaný zisk, vsadíme-li 10 Kč na červenou barvu?

b) Jaká je pravděpodobnost výhry při sázce na druhý a třetí sloupec? Jaký bude náš očekávaný zisk, vsadíme-li 10 Kč na druhý a třetí sloupec?

Řešení

a) Na kole rulety je celkem 37 políček, z toho 18 je červených. Pravděpodobnost výhry je proto $18/37 \doteq 0{,}486$.

Vsadíme-li 10 Kč na červenou barvu, pak v průměru v 18 z 37 případů vyhrajeme 10 Kč a ve zbývajících 19 případech o 10 Kč přijdeme. Očekávaný zisk je proto

$$\left(10 \cdot \frac{18}{37} - 10 \cdot \frac{19}{37}\right) \text{ Kč} = -\frac{10}{37} \text{ Kč} \doteq -0{,}270 \text{ Kč}.$$

b) Druhý a třetí sloupec obsahují dohromady 24 z celkového počtu 37 čísel. Pravděpodobnost výhry je proto $24/37 \doteq 0{,}649$.

Vsadíme-li 10 Kč na druhý a třetí sloupec, pak v průměru ve 24 případech vyhrajeme 5 Kč a v ostatních 13 případech přijdeme o 10 Kč. Očekávaný zisk je proto

$$\left(5 \cdot \frac{24}{37} - 10 \cdot \frac{13}{37}\right) \text{ Kč} = -\frac{10}{37} \text{ Kč} \doteq -0{,}270 \text{ Kč}.$$

Někteří hráči samozřejmě mohou mít štěstí a odejít s velkou výhrou. Vypočítané hodnoty však napovídají, že v delším časovém horizontu hráči více peněz prohrají než vyhrají, a jediný, kdo má zaručený trvalý zisk, je provozovatel kasina či herny.

2. Kurzové sázky

Další skupina úloh sbírky se zabývá kurzovými sázkami. Čtenáři se zde kromě jiného mohou seznámit s pojmem *návratnosti sázky* (která je pro sázejícího vždy menší než 100 %), pochopit souvislost mezi vypsanými kurzy a odhady pravděpodobností daných událostí a uvědomit si, kdo na těchto hrách může dlouhodobě vydělat.

3. Podmíněné pravděpodobnosti

Úlohy v následující části jsou věnovány podmíněným pravděpodobnostem. Na jednotlivých příkladech nejprve připomínáme pojem podmíněné pravděpodobnosti $P(\mathsf{A}|\mathsf{B})$ a ilustrujeme rozdíl mezi podmíněnou a nepodmíněnou pravděpodobností a také mezi podmíněnými pravděpodobnostmi $P(\mathsf{A}|\mathsf{B})$ a $P(\mathsf{B}|\mathsf{A})$.[5]

V následujících ukázkách je hledaná pravděpodobnost odhadována pomocí relativních četností.

Příklad: Falešně pozitivní testy

Před nástupem do nového zaměstnání musel David podstoupit preventivní prohlídku, jejíž součástí byl test na HIV. Podle údajů výrobce test odhalí přítomnost viru u nemocné osoby s pravděpodobností 99,90 % a s pravděpodobností 99,99 % dá negativní výsledek u osoby zdravé. Po vyhodnocení lékař Davidovi sdělil, že mu test vyšel pozitivní a že musí znovu na odběr krve, aby se výsledek ověřil.

Předpokládejme, že v České republice je virem HIV infikováno přibližně 0,01 % obyvatel.

a) David se z hlediska rizika nákazy virem HIV považuje za průměrného Čecha. Jaká je po výsledku prvního testu pravděpodobnost, že má skutečně HIV?

b) Představme si, že David žije od dětství spořádaným životem v malé vesnici na Vysočině, kde je podle statistik čtyřikrát nižší výskyt HIV než v celé České republice. Jaká je nyní pravděpodobnost, že má skutečně HIV?

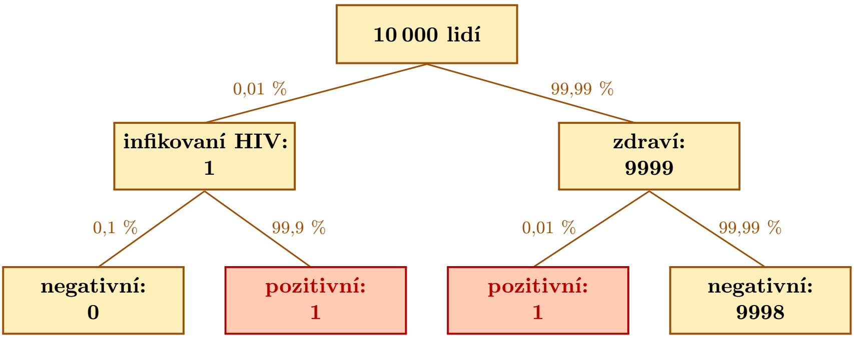
---

[5] Poslední zmíněný rozdíl je dobře patrný již z následujícího jednoduchého příkladu. Uvažujme náhodně vybranou osobu, která žila v USA před rokem 2013. Zřejmě platí: $P(X$ byl muž $|$ X byl prezident USA$) \neq P(X$ byl prezident USA $|$ X byl muž$)$, neboť v USA byli všichni dosavadní prezidenti muži, zdaleka ne každý muž se však stal prezidentem.

c) Nyní si představme, že David v minulosti injekčně užíval drogy. Na internetu najde statistiky, podle nichž je mezi injekčními uživateli drog 1 % HIV pozitivních. Jaká je v tomto případě pravděpodobnost, že je skutečně infikován virem HIV?

Řešení

a) Pravděpodobnost, že bude mít David pozitivní test, je-li zdráv, je velmi nízká: $P(\text{pozitivní test} \mid \text{zdravý}) \doteq 0,01\ \%$. Jakmile však test pozitivní vyjde, pak pravděpodobnost $P(\text{nemocný} \mid \text{pozitivní test})$, že je skutečně infikován, není tak vysoká, jak by se možná na první pohled mohlo zdát. Uvažujme například 10 000 Čechů; podle zadání mezi nimi bude v průměru jeden infikovaný, jemuž vyjde téměř jistě pozitivní výsledek, a 9999 zdravých, z nichž přibližně jednomu vyjde test „falešně pozitivní“ (v úlohách tohoto typu budeme vždy zaokrouhlovat na celé počty osob):



Pozitivní výsledek tedy vyjde průměrně dvěma osobám, jedné zdravé a jedné infikované. Pravděpodobnost, že je David nakažený, vyšel-li mu pozitivní test, je proto přibližně rovna $1/2$, tj. 50 %.

Podobně můžeme postupovat i v částech b) a c). Ve sbírce jsou uvedeny všechny obrázky, v tomto článku si však dovolím ponechat jejich vytvoření čtenářům a jen stručně popsat výsledky.

b) Je-li výskyt HIV v dané populaci čtyřikrát nižší než v celé České republice, pak bude infikován přibližně jeden člověk ze 40 000. Místo 10 000 osob, uvažovaných v předchozí části úlohy, si tedy představme 40 000 obyvatel Vysočiny. V průměru mezi nimi bude jeden infikovaný, jemuž opět vyjde téměř jistě pozitivní výsledek, a 39 999 zdravých. Přitom přibližně čtyřem zdravým osobám, tj. 0,01 % ze 39 999, vyjde falešně pozitivní test.

Pozitivní test tedy vychází v průměru pěti osobám, z nichž čtyři jsou zdravé a jen jedna infikovaná. Pravděpodobnost, že je David skutečně infikovaný, je proto přibližně rovna 1/5, tj. 20 %.

c) Je-li výskyt HIV v dané populaci roven 1 %, pak bude infikován přibližně jeden člověk ze 100. Aby nám vycházely celočíselné počty osob, uvažujme raději větší skupinu, například 1 000 000 osob. V průměru mezi nimi bude 10 000 infikovaných, z nichž 10 bude mít falešně negativní test, a 990 000 zdravých, z nichž přibližně 99 bude mít test falešně pozitivní. Pozitivní test tedy vychází v průměru 10 089 osobám, z nichž 99 je zdravých a 9990 je infikovaných. Pravděpodobnost, že je David skutečně infikovaný, je nyní přibližně $9990/10\,089 \doteq 0{,}9902$, tj. 99,02 %.

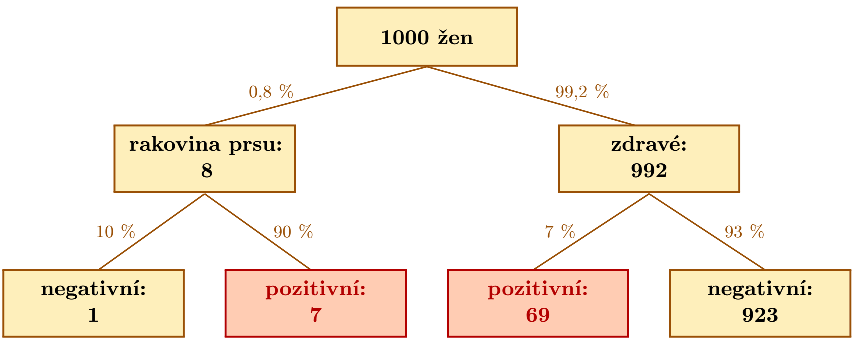
Poznámka. Povšimněte si, že hledaná pravděpodobnost závisí nejen na kvalitě testu, ale do velké míry také na výskytu dané nemoci. Je-li nemoc vzácná, pak je pravděpodobnost falešně pozitivního výsledku poměrně vysoká. S narůstajícím výskytem však tato pravděpodobnost rychle klesá.

Příklad: Mamografické vyšetření

Pravděpodobnost, že má žena rakovinu prsu, je 0,8 %. Pokud ji má, pak pravděpodobnost, že mamogram bude pozitivní, je 90 %. Pokud ji nemá, pak pravděpodobnost, že mamogram bude i tak pozitivní, je 7 %.

a) Představte si ženu, jejíž mamogram je pozitivní. Jaká je pravděpodobnost, že má skutečně rakovinu prsu?

b) Představte si ženu, jejíž mamogram je negativní. Jaká je pravděpodobnost, že rakovinu prsu nemá?



Řešení

a) Uvažujme například skupinu 1 000 žen. V průměru osm z nich má rakovinu prsu a přibližně sedm z těchto osmi žen bude mít pozitivní mamogram. Z ostatních 992 žen, které rakovinu nemají, jich bude mít zhruba 69 rovněž pozitivní mamogram. Celkem tedy vyjde pozitivní výsledek přibližně 76 ženám, z nichž jen 7 má rakovinu, viz obr. na str. 177. Hledaná pravděpodobnost je proto přibližně $7/76 \doteq 0{,}092$ tj. 9,2 %.

b) Ze stejného obrázku je patrné, že negativní mamogram vyjde celkem 924 ženám, z nichž jen jedna má rakovinu. Hledaná pravděpodobnost je proto přibližně $923/924 \doteq 0{,}9988$, tj. 99,89 %.

Poznámka. V předchozí úloze znamenal pozitivní výsledek poměrně nízkou, avšak nezanedbatelnou pravděpodobnost, že pacientka trpí rakovinou prsu, zatímco negativní výsledek tuto nemoc s velmi vysokou pravděpodobností vyloučil. Podobné testy tedy umožňují vyčlenit v prvním kole většinu zdravých pacientů a dále pak vyšetřovat jen ty, u nichž je podezření, že danou nemoc mají. K tomu se obvykle používají metody, které jsou sice přesnější, ale také dražší či znamenají výraznější zásah do organismu, a proto není vhodné je používat plošně na celou skupinu testovaných.

4. Posuzování věrohodnosti

Poslední typ úloh, pro jehož představení v tomto článku zbývá prostor, se týká situací, kdy je třeba prokázat, že určitý výsledek nastal například díky pozitivnímu působení jistého léku či procedury, zvláštním schopnostem jedince apod. Ve sbírce je nejprve objasněn pojem *p-hodnoty* jako pravděpodobnosti, že by ke stejnému nebo ještě výraznějšímu výsledku (např. stejně nebo ještě více uzdravených pacientů) došlo vlivem náhody za předpokladu, že by žádný pozitivní účinek neexistoval; je-li tato pravděpodobnost nižší než stanovená hranice (nejčastěji 5 %, někdy se požaduje jen 1 % nebo ještě méně), považuje se obvykle výsledek za statisticky významný, tj. prokazující, že k němu nedošlo pouhou náhodou.[6]

Příklad: Je lék skutečně účinný?

Nemoc, která se vyskytuje velmi vzácně a dosud se na ni nepodařilo nalézt účinnou léčbu, vede podle údajů Ministerstva zdravotnictví k úmrtí ve 30 % případů. Farmaceutická firma vyvine lék a oznámí, že snižuje

[6] Připomeňme ještě, že hypotéza, že žádný pozitivní účinek neexistuje, se obvykle označuje jako *nulová*, zatímco hypotéza, podle níž k pozitivnímu účinku dochází, se označuje jako *alternativní*.

úmrtnost na tuto chorobu. Pro ověření je objednána odborná studie. Předpokládejme, že k prokázání účinnosti je požadována p-hodnota menší než 5 %.

Lék byl podán dvaceti pacientům, z nichž sedmnáct přežilo a tři zemřeli. Prokazuje takováto studie účinnost léku?

Řešení

Předpokládejme, že lék je neúčinný, a vypočítejme pravděpodobnost, že by i tak náhodou přežilo alespoň 17 z 20 pacientů, tj. že by se stejný nebo ještě výraznější výsledek objevil pouhou náhodou. Tato pravděpodobnost je součtem pravděpodobností následujících jevů: přežije právě 17 pacientů a zbývající tři zemřou (přitom existuje $\binom{20}{3}$ možností, jak vybrat tři pacienty, kteří nepřežijí), přežije právě 18 pacientů a zbývající dva zemřou, přežije právě 19 pacientů a jeden zemře, přežije všech 20 pacientů, tedy

$$\binom{20}{3} \cdot 0{,}7^{17} \cdot 0{,}3^3 + \binom{20}{2} \cdot 0{,}7^{18} \cdot 0{,}3^2 + 20 \cdot 0{,}7^{19} \cdot 0{,}3^1 + 0{,}7^{20} \doteq 0{,}107.$$

Výsledná pravděpodobnost činí 0,107, tj. 10,7 %, což je více než požadovaná hranice 5 %. Studie proto účinnost léku neprokazuje.

Poznámka. Pětiprocentní hranice je ve vědách široce používána a je považována za rovnocennou s právním vyjádřením „nade vší pochybnost“. Připouští však, že z dvaceti studií bude v průměru jedna chybná. Někdy je proto tato hranice snížena například na jedno procento.

Literatura

[1] *Anděl, J.*: Matematika náhody, Matfyzpress, Praha, 2007.

[2] *Gigerenzer, G.*: Calculated Risks. How to Know When Numbers Deceive You, Simon & Schuster, New York, 2002.

[3] *Hromadová, J.*: Aplikační úlohy z geometrie, Matematika–Fyzika–Informatika, roč. 22 (2013), s. 17–24.

[4] *Kaplan, M., Kaplanová, E.*: Chances are… Adventures in Probability. Viking Penguin, New York, 2006 [český překlad: Čtrnáct, M.: Šance je, že… Dobrodružství v pravděpodobnosti, Praha, Triton, 2008].

[5] *Pavlíková, P., Robová, J., Slavík, A.*: Fahrenheit, Celsius a americký cent. Matematika–Fyzika–Informatika, roč. 20 (2011), s. 385–392.

[6] *Pavlíková, P., Robová, J., Slavík, A.*: Úlohy s dopravní tematikou. Matematika–Fyzika–Informatika, roč. 20 (2011), s. 454–461.

[7] *Rosenthal, J. S.*: Struck by Lightning: The Curious World of Probabilities, Toronto, Harper Collins, 2005 [český překlad: Hykšová, M.: Zasažen bleskem. Podivuhodný svět pravděpodobností. Praha, Academia, 2008].