4 Počet pravděpodobnosti jako základ statistického usuzování

K analýze dat přistupujeme z několika hledisek. Dosud jsme probrali explorační a popisnou analýzu dat, jimiž dokážeme přehledně shrnout informace, jež se týkají právě těch objektů, které jsme pozorovali nebo změřili. Jestliže jsme však data získali na základě dobře navrženého výzkumného plánu, můžeme provádět zobecňující úsudky o chování sledovaných proměnných a jejich parametrech v celé uvažované populaci. Metody takového statistického usuzování se opírají o počet pravděpodobnosti. Proto jsou základy počtu pravděpodobnosti tématem této kapitoly.

Metody statistického testování a odhadovaní vyžadují data získaná náhodným výběrem nebo metodou znáhodněného experimentu. Statistické usuzování spočívá na kladení otázek typu: „Jak často tato metoda dá správnou odpověď, pokud ji použiji mnohokrát?" Pokud si nemůžeme představit, že proces sběru dat lze opakovat (např. tím, že z populace vybereme jiný výběr), statistické usuzování nemá smysl. Jestliže však využijeme při získání dat náhodu, můžeme použít teorii pravděpodobnosti pro zodpovězení otázky: „Jak často nastane určitý jev, pokud experiment nebo výběr provedeme mnohokrát?"

Látka probíraná v této kapitole je klíčová pro porozumění většině postupů, jež uvedeme v dalších částech knihy. Nesmírně důležitý je pojem náhodné proměnné a rozdělení náhodné proměnné. V závěru této kapitoly se dostaneme k další problematice, která je pro statistickou analýzu rozhodující – k pravděpodobnostnímu chování statistik vypočítaných z dat, a uvedeme základní teoretická pravděpodobnostní rozdělení, jež jsou vhodná pro popis variability statistik. Tato část tvoří základ pro rozvinutí principů teorie statistického usuzování, kterými se budeme zabývat v příští kapitole.

Aplikace počtu pravděpodobnosti a příslušné teorie pronikly do četných vědních oborů i oblastí praktické činnosti. Historici dovozují, že vývoj počtu pravděpodobnosti neprobíhal nijak jednoduše. Jak Řekové, tak první křesťané neměli důvod se zabývat kvantifikací náhody. Řekové si uvědomovali působení náhody, ale věřili, že není správné matematicky spojovat to, co se stalo, a to, co by se *mělo stát*, protože by šlo o překrývání „pozemského plánu" a „nebeského plánu". Navíc u Řeků hrál roli jejich antiempiricismus. Znalost se nemohla získat experimentováním, ale pouze logickou cestou. Tyto dva momenty jim bránily zabývat se problémem náhody v souvislosti s predikcí jevů. Pro první křesťany zase něco jako náhoda nemohlo vůbec existovat. Každá událost byla přímým svědectvím božského působení.

Nejdříve se prvky teorie pravděpodobnosti uplatňovaly při výpočtech šance na výhru v hazardních hrách. Mezi první, kdo se zabýval pravděpodobnostními problémy, patřil Blaise Pascal (1623–1662). Hazardní hry v době, o níž mluvíme, měly za sebou historii dlouhou nejméně dva tisíce let, protože již Řekové a Římané byli vášnivými hráči. Nejpopulárnější hra Pascalovy doby se nazývala „hazard", jejíž pojmenování pochází z arabského *al zhar*, což znamená „kostka". Pascal si dal za úkol zodpovědět řadu otázek, které mu položil jeho přítel António Gombard rytíř de Mere. Například – proč je výhodné vsadit na čtyři šestky při čtyřech hodech a proč není výhodné vsadit při dvojnásobném hodu na dvě šestky v 24 pokusech? Vedl o tomto problému korespondenci s jiným významným vědcem té doby Pierre de Fermatem.

Nejstarší knihou o pravděpodobnosti bylo dílo holandského matematika Christiana Huygense *De Ratiociniis in Ludo Aleæ* (Výpočty v hrách náhody) z roku 1657. Po padesát let sloužila jako standardní učební text o pravděpodobnosti. Právě jejího autora považují mnozí historici za zakladatele teorie pravděpodobnosti. Pierre Simon Laplace (1749–1827), autor přehledného pojednání o teorii pravděpodobnosti, prohlásil: „Je obdivuhodné, že počtu pravděpodobnosti, jenž vznikl při úvahách o hazardních hrách, bylo určeno stát se nejdůležitější složkou lidského vědění."

Rozvoj teorie pravděpodobnosti si vyžádal kromě přemýšlení i fyzické úsilí, o čem svědčí pokusy statistika K. Pearsona, který v roce 1900 hodil 24 000krát mincí, aby se přesvědčil, zda relativní četnost jevu, že padne „orel", konverguje k číslu 0,5. Jeho pokus vedl k četnosti „orla" 12012.

4.1 Základní pojmy a výpočty

Vysvětlíme stručně pouze základy počtu pravděpodobnosti, které budeme potřebovat v této knize. Musíme si přitom uvědomit, že matematická teorie pravděpodobnosti nemůže objasnit podstatu náhodnosti a pravděpodobnosti. Je pouze vhodným formálním popisem situací, v nichž se náhodnost, resp. nejistota projevuje; umožňuje o nich uvažovat.

4.1.1 Náhodné jevy, pravděpodobnost

Náhodnost vede k tomu, že jevy, které nás zajímají, se za daných podmínek mohou nebo nemusí vyskytnout. Například při házení mincí sledujeme jev, že padne „orel". V daném hodu můžeme predikovat jeho výsledek pouze vyjádřením pravděpodobností možností, jež mohou nastat. Pravděpodobnost, že padne „orel", vyjadřujeme číslem, které má určitý význam. Slova pravděpodobný nebo nepravděpodobný, jež se vyskytují v běžné řeči, vyjadřují nejistotu kvalitativně. Vztah tohoto vyjádření k matematickému pojmu pravděpodobnost je dán kontextem.

Určitý fenomén považujeme za náhodný, jestliže jeho výskyt je nejistý, ale zároveň pozorujeme v dlouhé řadě situací určitou pravidelnost v rozdělení jeho výskytu.

PŘÍKLAD 4.1

Použití teorie pravděpodobnosti pro modelování jevů

Počet pravděpodobnosti lze využít pro modelování nejrůznějších situací. Uvedeme jednoduchý modelový příklad z oblasti sportu, který lze řešit pomocí pravděpodobnostního počtu. Jeden z přístupů k řešení popíšeme na konci tohoto odstavce (s. 122).

Jana má ve svém repertoáru dva druhy tenisového podání, tvrdý a měkký servis. Její tvrdý servis je v poli v 50 % podání a v 75 % pak uhraje míč. Měkký servis Jana nezkazí v 75 %, ale míč pak uhraje jenom v 50 %. Jakou má Jana zvolit strategii při svém podání, pokud lze předpokládat, že během utkání se tyto charakteristiky nezmění? Má hrát obě podání tvrdě nebo měkce? Nebo má začít tvrdým podáním a po chybě podávat měkce? Je snad pro ni lepší zahrát první podání měkce a druhé podání tvrdě? Jak by se měla rozhodovat, aby v průměru dosahovala nejlepších výsledků, jestliže předpokládáme, že uvedené relativní četnosti jsou platné bez ohledu na průběh utkání?

Existuje mnoho různých definic pravděpodobnosti: definice axiomatická; definice pravděpodobnosti jako kvantitativní míry jistoty; klasická definice, jež pojem pravděpodobnosti převádí na pojem stejné možnosti. Uvedeme **statistickou definici pravděpodobnosti**.

Mluvíme o **náhodném pokusu**, jestliže při pokusu lze dostat různé možné výsledky a přitom:

1. nelze předem určit, který z těchto výsledků získáme;
2. pokus lze libovolně často opakovat, aniž se jednotlivá opakování vzájemně ovlivňují.

Množina všech možných výsledků náhodného pokusu tvoří prostor náhodných výsledků (E). Například při hodu mincí tvoří prostor jevů v jednom hodu „panna" a „orel".

Vymezená množina výsledků je **náhodný jev**. Všechny možné náhodné jevy tvoří **pole jevů**. Jev, jenž se skládá pouze z jednoho výsledku, se nazývá elementární jev. Jev, který nastává, jestliže dostaneme více možných výsledků, se nazývá jev složený.

Jev jistý obsahuje všechny možné výsledky náhodného pokusu. Pro pole náhodných jevů lze použít vztahy teorie množin.

Symbolem $A \cup B$ označujeme jev, že nastane jev A nebo nastane jev B nebo že nastanou oba dva. Současný výskyt jevu A a jevu B označujeme symbolem $A \cap B$. Případ, že $A \cap B$ je prázdná množina, znamená vzájemně se vylučující jevy.

Také říkáme, že tyto jevy jsou disjunktní. Jev doplňkový $\overline{A}$ (nebo také opačný) k jevu A je jev, který nastane, když nenastane v pokusu jev A.

Pravděpodobnost náhodného jevu A je číslo $P(A)$, k němuž se blíží relativní četnost jevu A, jestliže pokus dostatečně často opakujeme. Jestliže jsme provedli n pokusů a v m z nich nastal jev A, pak názorně vyjádřena:

$$\lim_{n\to\infty} \frac{m}{n} = P(A)$$

Tuto hodnotu pravděpodobnosti považujeme v teorii pravděpodobnosti za danou. Výrok „Pravděpodobnost jevu A je rovna hodnotě p" znamená, že $P(A) = p$.

Pravděpodobnost náhodného jevu je tedy číslo mezi 0 a 1, které popisuje relativní četnost, s jakou se jev vyskytne ve velmi dlouhé řadě opakování situace, kdy tento jev může nastat. Pravděpodobnosti popisují pouze to, co se stane v dlouhé řadě pokusů. Krátké série náhodných jevů, jako házení mincí nebo střelba na koš, často nevypadají náhodně, protože neukazují pravidelnost, jež se ve skutečnosti může prosadit jenom při mnoha opakováních.

Pravděpodobnost má tyto základní vlastnosti:

1. Pravděpodobnost jevu, který je jistý, se rovná 1.
2. Pravděpodobnost jevu nemožného je rovna 0.
3. Lze-li náhodný jev rozložit na několik vzájemně se vylučujících (disjunktních) jevů, pak se jeho pravděpodobnost rovná součtu pravděpodobností těchto jevů.

Pro výpočet pravděpodobnosti jevu A často používáme pravidlo, které je východiskem definice pravděpodobnosti na základě stejné možnosti: Jestliže náhodný pokus může vést k r různým elementárním jevům, jež jsou stejně pravděpodobné, pak pravděpodobnost jevu A je

$$P(A) = \frac{\textit{počet elementárních jevů, které vedou k A}}{r}.$$

PŘÍKLAD 4.2

Elementární a složený náhodný jev a jejich pravděpodobnosti

Při házení kostkou platí, že prostor náhodných výsledků E je (1; 2; 3; 4; 5; 6). Příkladem elementárního jevu je jev, že padne číslo 5. Počet všech elementárních jevů je 6. Jev A, že padne sudé číslo, je jevem složeným – tvoří jej 3 elementární jevy. Proto je pravděpodobnost padnutí sudého čísla $P(A) = 3/6 = 1/2$. Relativní četnost tohoto jevu se tedy se vzrůstajícím počtem hodů blíží k číslu 0,5.

Často používáme pravděpodobnosti spojení a průniku jevů A a B nebo pravděpodobnost doplňku jevu. Pravidlo 3 lze napsat obecněji pomocí rovnice

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Rozšíření tohoto pravidla na tři jevy má tvar

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C).$$

Ze tří základních vlastností pravděpodobnosti plynou již všechny vlastnosti další. Uvedeme ty nejvýznamnější:

1. Pro libovolný jev A platí: $0 \le P(A) \le 1$.
2. Je-li jev $\overline{A}$ doplňkový k jevu A, pak $P(\overline{A}) = 1 - P(A)$.
3. Je-li jev A částí jevu B, pak $P(A) \le P(B)$.

PŘÍKLAD 4.3

Výpočet pravděpodobnosti různých jevů

Statistické šetření ukázalo u 1000 dotázaných občanů volební preference, které uvádí tabulka 4.1. Jestliže z této skupiny náhodně vybereme jedince, jaká bude pravděpodobnost jevu:

a) bude se jednat o ženu, která nepreferuje ODS;
b) osoba bude ženského pohlaví nebo chce volit „ostatní".

Obě úlohy nejsou složité, ale musíme si promyslet přesně obsah otázky.
V první úloze je odpověď dána zlomkem (530 – 220)/1000.
Abychom vyřešili druhou otázku, musíme si uvědomit, že se jedná o výpočet pravděpodobnosti sjednocení jevu (*žena*) ∪ (*ostatní*). Z toho plyne

$$P((\textit{žena}) \cup (\textit{ostatní})) = P(\textit{žena}) + P(\textit{ostatní}) - P((\textit{žena}) \cap (\textit{ostatní}))$$
$$= 530/1000 + 303/1000 - 157/1000.$$

Tab. 4.1 Modelová data – výsledky průzkumu volebních preferencí

Preferovaná politická strana	Ženy	Muži	Celkem
ČSSD	153	130	283
ODS	220	194	414
Ostatní	157	146	303
Celkem	530	470	1000

4.1.2 Podmíněná pravděpodobnost, Bayesova formule

Často závisí pravděpodobnost výskytu určitého jevu na tom, zda nastal či nenastal nějaký jiný jev. Takovým pravděpodobnostem říkáme podmíněné a značíme je $P(A|B)$, což čteme: pravděpodobnost jevu A za předpokladu, že nastal jev B. Pro podmíněné pravděpodobnosti lze dokázat všechna základní pravidla, která jsme uváděli u nepodmíněné pravděpodobnosti, tedy zejména $0 \leq P(A|B) \leq 1$. Platí dvě ekvivalentní rovnice, pokud $P(B)$ nemá nulovou hodnotu:

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

Rovná-li se podmíněná pravděpodobnost pravděpodobnosti nepodmíněné, tedy když $P(A|B) = P(A)$, říkáme, že jevy A a B jsou **statisticky nezávislé**. Výskyt jevu B nemá v tomto případě vliv na pravděpodobnost výskytu jevu A v dané situaci. Jsou-li jevy A a B statisticky nezávislé, pak$P(A \cap B) = P(A) \cdot P(B)$. Platí: Jsou-li jevy A, B statisticky nezávislé, pak jsou statisticky nezávislé i dvojice jevů $\overline{A}, B$; $A, \overline{B}$; $\overline{A}, \overline{B}$.

Jestliže jev B nastává vždy s některým jevem $A_1, \ldots, A_n$, přičemž A_i jsou jevy disjunktní, pak

$$P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i).$$

Tento vztah nazýváme vzorec pro úplnou pravděpodobnost.

Při pravděpodobnostních úvahách se často používá **Bayesova formule**. Slouží k vypočítání podmíněné pravděpodobnosti $P(A|B)$ za předpokladu, že známe pravděpodobnosti $P(B|A)$ a $P(A)$. Pomáhá nám např. při výpočtech, které provádíme při hodnocení diagnostických testů binárního typu v medicínské a psychologické diagnostice. Uvedeme její jednoduchou podobu:

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\overline{A})P(B|\overline{A})}$$

V čitateli této formule je pravděpodobnost, že současně nastane jev A a jev B, ve jmenovateli je vzorec pro úplnou pravděpodobnost jevu B.

PŘÍKLAD 4.4

Aplikace Bayesova přístupu pro studium vlastností diagnostických testů

Významné použití nachází Bayesova formule při hodnocení diagnostických testů, jež mohou nabývat pouze dvou hodnot (pozitivní, negativní). Takové hodnocení se provádí i u uměle dichotomizovaných kvantitativních testových výsledků (např. normální výsledek, výsledek nad normální mezí testu). Popíšeme stručně tuto situaci. Pacient může, nebo nemusí mít danou chorobu ($D+$, $D-$). Provedený diagnostický test může, nebo nemusí tuto chorobu indikovat

($T+$, $T-$). Záleží to na jeho specificitě a senzitivitě. **Senzitivita diagnostického testu** *Se* je podmíněná pravděpodobnost $P(T+|D+)$ toho, že výsledek testu bude pozitivní, když pacient má chorobu. **Specificita diagnostického testu** *Sp* je podmíněná pravděpodobnost $P(T-|D-)$, že za předpokladu, že pacient nemá danou chorobu, test bude negativní. **Prediktivní hodnota pozitivního testu** $P+$ je podmíněná pravděpodobnost $P(D+|T+)$, že pacient má chorobu, pokud byl test pozitivní. **Prediktivní hodnota negativního testu** $P-$ je podmíněná pravděpodobnost $P(D-|T-)$, že pacient nemá danou chorobu, když test byl negativní. *Prevalence* $P(D+)$ je pravděpodobnost choroby v populaci. Uvedené pravděpodobnosti se odhadují pomocí statistické evidence výsledků v medicínských databázích a zvlášť zaměřeného výzkumu diagnostické věrohodnosti diagnostického testu. Podle výsledků testu se sestavuje čtyřpolní tabulka s četnostmi (tabulka 4.2). Například četnost a je počet výsledků nemocných jedinců, kteří měli pozitivní test. Pro odhad uvedených charakteristik se četnosti z tabulky použijí takto:

$$\begin{aligned} Se &= P(T+|D+) = a/(a+b) \\ Sp &= P(T-|D-) = d/(c+d) \\ P+ &= P(D+|T+) = a/(a+c) \\ P- &= P(D-|T-) = b/(b+d) \end{aligned}$$

To však lze provést pouze v případě, že získáváme výsledky pro jedince vybraného zcela náhodným způsobem. Častější je případ, kdy jsou k dispozici předem dané skupiny jedinců s diagnózou nebo bez ní a provedeme u obou skupin posuzovaný test. Odhad senzitivity a specificity je v pořádku, ale odhad pravděpodobností $P+$ a $P-$ pomocí četností z tabulky je zkreslený. Musíme nejdříve získat informaci o výskytu uvažované nemoci v populaci. Proto zjišťujeme prevalenci $P(D+)$ u různých subpopulací. Podle Bayesovy formule následně spočítáme prediktivní hodnotu pozitivního testu

$$P+ = \frac{Se P(D+)}{Se P(D+) + (1 - Sp)(1 - P(D+))}$$

nebo prediktivní hodnotu negativního testu

$$P- = \frac{Sp(1 - P(D+))}{Sp(1 - P(D+)) + (1 - Se)P(D+)}.$$

Ve vzorcích použijeme prevalenci $P(D+)$ podle toho, z které subpopulace jedinec pochází.

Tab. 4.2 Čtyřpolní tabulka s četnostmi

Skutečná diagnóza	Výsledek testu	
	$T+$	$T-$
$D+$	a	b
$D-$	c	d

4.1.3 Šance

Často používáme výraz, že šance vítězství fotbalového mužstva v daném zápase je 1 : 4 nebo 2 : 1. V prvním případě považujeme vítězství našeho klubu za málo pravděpodobné, ve druhém případě se domníváme, že pravděpodobnost vítězství $P(\mathrm{V})$ je dvojnásobně větší než pravděpodobnost prohry $P(\mathrm{P})$. Tedy šance na vítězství mého klubu se rovná $P(\mathrm{V})/P(\mathrm{P}) = 2:1$. Protože vítězství V a prohra P jsou vzájemně se vylučující jevy, můžeme šanci na vítězství zapsat takto:

$$\textit{šance na vítězství} = \frac{P(\mathrm{V})}{1 - P(\mathrm{V})}$$

Formálně je šance ve prospěch nějakého jevu A definována poměrem

$$\textit{šance ve prospěch}\ \mathrm{A} = \frac{P(\mathrm{A})}{P(\bar{\mathrm{A}})} = \frac{P(\mathrm{A})}{1 - P(\mathrm{A})},$$

kde $P(\mathrm{A})$ je pravděpodobnost jevu A.

Jestliže počítáme pravděpodobnosti pomocí vypočítaných relativních četností, pak platí, že jmenovatele ve zlomcích pro relativní četnost se vyruší. Proto lze vypočítat šanci pouze pomocí četností jevů A a $\bar{\mathrm{A}}$, které se realizovaly v daném sledování:

$$\textit{šance ve prosěch}\ \mathrm{A} = \frac{\textit{počet výskytů jevu}\ \mathrm{A}}{\textit{počet případů, kdy jev}\ \mathrm{A}\ \textit{nenastal}}.$$

Jestliže známe šanci ve prospěch A, lze samozřejmě zpětně vypočítat pravděpodobnost jevu A v dané situaci.

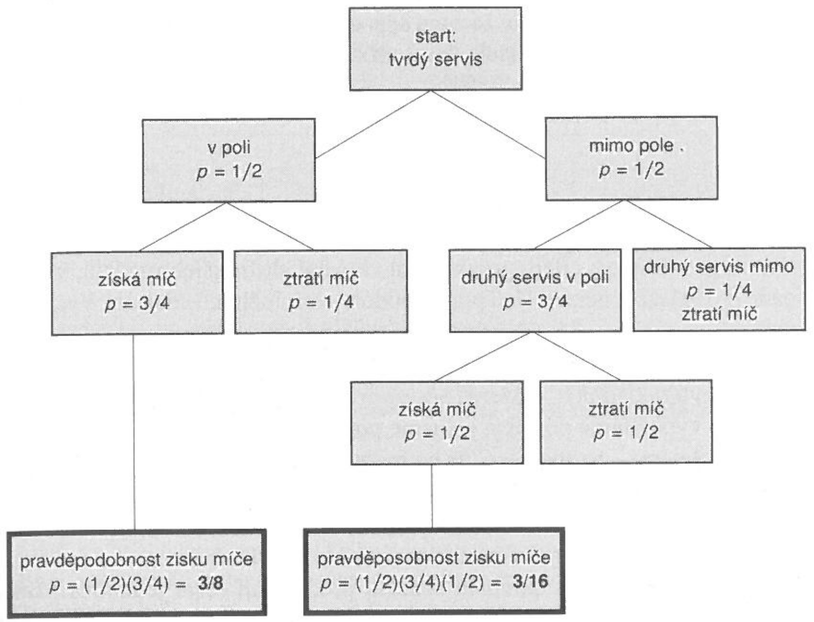
Řešení příkladu 4.1 ze stránky 117

Naším úkolem je navrhnout pro tenistku Janu, jak má využívat své podání. Hledané řešení by mělo vést v průměru k největšímu počtu vyhraných míčů. Tvrdý servis se jí podaří v 50 %, míč pak vyhraje v 75 %. Měkký servis se jí podaří sice v 75 %, ale pak ho vyhraje pouze v 50 %. Jaké má zvolit pořadí obou typů servisů, aby dosáhla v průměru nejlepšího výsledku?

Úkol lze řešit sestrojením stromového grafu pro znázornění různých možností průběhu hry. Pomocí něj vypočítáme pravděpodobnosti výhry, které pak srovnáme.

Uvažujeme jako první možnost, že Jana začne tvrdým servisem a pokud se jí nepodaří, servíruje měkce. První servis může nebo nemusí být v poli s pravděpodobností 1/2. To lze reprezentovat dvěma hranami, které vycházejí z prvního uzlu grafu (obr. 4.1). Jestliže první servis je v poli, pravděpodobnost výhry míče je 3/4 a ztráty je 1/4. Jestliže první míč je v autu a Jana servíruje podruhé měkce, je pravděpodobnost podařeného servisu 3/4 a následná pravděpodobnost výhry míče se rovná 1/2. Jestliže i druhý servis je v autu, Jana míč ztrácí. Celková pravděpodobnost výhry, když Jana začne tvrdým servisem a v případě jeho zkažení pokračuje měkkým servisem, se spočte jako součet pravděpodobností výhry ve dvou nezávislých větvích grafu, jež vedou k výhře 3/8 + 3/16 = 9/16. Poznamenejme,

Obr. 4.1 Stromový graf jedné varianty strategie podání (první podání tvrdé, druhé měkké)



Tab. 4.3 Příklad pravděpodobnostního modelu z oblasti sportu

První podání	Druhé podání	Pravděpodobnost výhry
tvrdé	měkké	9/16 = 0,563
tvrdé	tvrdé	9/16 = 0,563
měkké	tvrdé	15/32 = 0,469
měkké	měkké	15/32 = 0,469

že ostatní pravděpodobnosti zisku míče v grafu jsou podmíněné pravděpodobnosti, kde podmínka je daná předpokladem, že nastaly jevy, které uvažovanému ziskovému uzlu předcházely.

Podobný způsobem odvodíme pravděpodobnosti výhry pro další tři možnosti. Výsledky znázorníme tabulkou 4.3. Vidíme, že Jana by měla nejdříve podávat tvrdě. Jestliže se jí podání nepodaří, je jedno, zda bude druhý servis podávat měkce nebo tvrdě. Můžeme jenom doufat, že ho dá do pole a vyhraje.

4.1.4 Využití simulace pro odhad pravděpodobností

Simulace provádíme s cílem prozkoumat chování složitějších modelů, v našem případě pro nalezení neznámých pravděpodobností složitých událostí. Vycházíme přitom ze znalosti pravděpodobností základních jevů, které jsou pro celý proces směrodatné. Pokud určíme tyto pravděpodobnosti a mechanismus, jak nové jevy vznikají z jevů základních, můžeme postupovat dvěma cestami. Buď nové pravděpodobnosti vypočítáme postupy, jež jsme popsali v předchozím příkladu, anebo proces simulujeme – to znamená, že ho mnohokrát opakujeme a jednotlivé pravděpodobnosti odhadujeme pomocí relativních četností jevů, které nás zajímají. Těmto metodám se také říká metody Monte Carlo. Při modelování nám obvykle pomáhá počítač, ale někdy lze model realizovat manuálně.

Simulace a počítání relativních četností představují často jednodušší cestu, jak odhadnout neznámé pravděpodobnosti. Naše simulace povede k validním výsledkům, pokud jsme model dobře sestavili. Simulace v kontextu statistiky lze definovat jako využívání tabulek náhodných čísel nebo generování náhodných čísel počítačem s cílem napodobit reálné procesy.

PŘÍKLAD 4.6

Odhad pravděpodobnosti pomocí simulace

Házíme mincí a chceme zjistit pravděpodobnost, že v deseti hodech třikrát za sebou padne orel nebo třikrát za sebou padne panna. Popíšeme kroky sestavování modelu a realizaci simulace pomocí tabulky náhodných čísel z přílohy B.

Krok 1. Zadání pravděpodobností. Náš model má dvě části:

a) v každém hodu je pravděpodobnost „orla" 0,5;
b) jednotlivé hody jsou na sobě nezávislými pokusy – to znamená, že výsledek jednoho hodu neovlivňuje pravděpodobnosti výsledku jiného hodu.

Tab. 4.4 Příklad využití simulace pro odhad pravděpodobnosti

1	9	2	2	3	9	5	0	3	4
P	P	O	O	P	P	P	O	P	O

0	5	7	5	6	2	8	7	1	3
O	P	P	P	O	O	O	P	P	P

9	6	4	0	9	1	2	5	3	1
P	O	O	O	P	P	O	P	P	P

Krok 2. Jednotlivým základním jevů přiřadíme číselné označení. V tabulce náhodných čísel (tab. I přílohy B) budou jednotlivé číslice zastupovat výsledek hodu mincí. Každá číslice v tabulce má pravděpodobnost 0,1 a jejich následné uspořádání je v tabulce nezávislé. Jedna číslice znamená jeden hod. Sudá číslice reprezentuje jev „orel" (O), lichá „panna" (P).

Krok 3. Simulace mnoha opakování daného pokusu. Deset číslic v tabulce náhodných čísel za sebou představuje jeden pokus (deset hodů). Zaznamenáme mnoho skupin po deseti číslicích a zjistíme v každé z nich jevy „O" a „P". Určíme, v kolika případech z nich se vyskytl sledovaný jev (3 panny nebo orli za sebou neboli 3 sudá nebo lichá čísla za sebou). Z tabulky náhodných čísel v příloze B zjistíme např. pro první tři skupiny po deseti číslicích údaje, které uvádí tabulka 4.4. Ve všech těchto třech simulovaných opakováních pokusu deseti hodů mincí se realizoval jev tří za sebou stejných hodnot. Jestliže vyhledáme 25 skupin po deseti číslicích, získáme počet realizovaných jevů $m = 25$. Tedy relativní četnost jako odhad hledané pravděpodobnosti má hodnotu $m/n = 20/25 = 0{,}80$. Pro přesnější odhad musíme opakovat celý pokus mnohem vícekrát pomocí počítače, který umí generovat náhodná čísla. Pak dospějeme k číslu přibližně 0,826.

Příklad ilustroval mnoho pravděpodobnostních problémů z praxe. Ty se vyznačují tím, že se provádějí nezávislé pokusy, v nichž sledovaný jev má stejnou pravděpodobnost. Velmi důležitou roli zde hraje nezávislost jednotlivých pokusů (v našem příkladu nezávislost jednotlivých hodů mincí). Nezávislost pokusů lze verifikovat pozorováním mnoha realizací pokusu.