Pravděpodobnost a statistika

poznámky k přednáškám

Radim Bělohlávek

radim.belohlavek@acm.org

Katedra informatiky
Univerzita Palackého v Olomouci

2016

Text je určen studentům informatických oborů na Přírodovědecké fakultě Univerzity Palackého v Olomouci. Ja zamýšlen jako doprovodný text k předmětu Pravděpodobnost a statistika, který přednáším. Jde o text pracovní, je tudíž neúplný. Zjištěné chyby a případné další připomínky mi prosím zasílejte na adresu radim.belohlavek@acm.org.

Obsah

1 Úvod

1.1 Co je pravděpodobnost a statistika?

Pravděpodobnost a statistika patří mezi nejužitečnější oblasti matematiky.

Proč? Každá oblast matematiky je přeci užitečná, tak proč zrovna pravděpodobnost a statistika?

Dávný ideál: řešíme-li nějaký problém, máme mít úplné informace, které jsou jisté. Pak používáme deduktivní usuzování, z jistých informací tedy odvozujeme jisté informace. Například: pacient P má příznaky x, y, z; vím, že platí: když má pacient příznaky x, y, z, pak má nemoc n. Tedy usoudíme (MP), že P má nemoc n.

Realita je ale v naprosté většině případů odlišná od tohoto ideálu, informace jsou v naprosté většině případů neúplné a nejisté. Tak například to, že má pacient příznak x může být jisté, ale že má y může být poněkud nejisté, protože přístroj, který probíhalo pacientovo vyšetření se mohl dopustit chyby. Také pravidlo „když má pacient příznaky x, y, z, pak má nemoc n“ nemusí být jisté, mohou existovat výjimky (uvést konkrétní příklad).

Přesto o nich provádíme úsudky (tj. přemýšlíme o nich a děláme závěry), na základě takových informací se rozhodujeme apod. A lze říci, že lidé jsou schopni o nejistých informacích usuzovat úspěšně.

V naprosté většině případů tak člověk provádí na základě intuice, popř. zdravého/selského rozumu . Příklad: vybíráme meloun a ze čtyř náhodně vybraných a rozkrojených jsou tři nahnilé, usoudíme, že melouny jsou špatné a meloun si nekoupíme. Další příklady, které člověk běžně řeší: jízda autem a píchnutí penumatiky, uzavírání pojištění.

Člověk tedy pracuje s neúplnými a nejistými informacemi, bez toho, že by prošel speciálním školením (stejně tak jako používá deduktivní logiku bez toho, že by prošel speciálním školením). Ale: intuice nás velmi často dosti zásadně klame (to je rozdíl od deduktivního usuzování, kde takovéto riziko není tak velké). Proto není radno se na intuici spoléhat. Literatura obsahuje řadu příkladů, chytáků a paradoxů, které ukazují úskalí intuitivního usuzování.

Z tohoto pohledu by bylo vhodné mít k dispozici vědu o tom, jak pracovat s nejistotou. Právě takovou vědou je teorie pravděpodobnosti.

Podobné je to se statistikou. Zabývá se zpracováním rozsáhlých souborů dat, a tedy rozsáhlých informací. Zabývá se rozmanitými otázkami takového zpracování. Např.: jaký je průměr, popř. další charakteristiky, které by nám několika málo čísly o ohromném souboru čísel řekly něco užitečného. Nebo: Jak navrhnout dostatečně spolehlivý průzkum veřejného mínění (nelze se ptát všech, kolika se zeptat, aby byl spolehlivý a reprezentativní)? Nebo: Lze ve získaných datech sledovat trendy? Lze sledovat v datech nějaké struktury? Lze podat vysvětlení (faktorová analýza)?

Shrňme tedy: teorie pravděpodobnosti se nezabývá něčím speciálním, s čím přichází do styku je úzká skupina odborníků (jako např. teorie diferenciálních rovnic, která je matematikou, bez které se neobejdou fyzikové, popř. inženýři). Zabývá se tím, s čím se na každém kroku dnes a denně potkává každý člověk: nejistota a nutnost se za podmínek nejistoty rozhodnout (rozhodování se nelze vyhnout, není to něco, čemu se můžeme nebo nemusíme věnovat).

Do určité míry je to podobné se statistikou. Pokud se chceme umět orientovt ve světě, ve kterém žijeme, je znalost statistiky velmi užitečná. Umožňuje nám popisovat realitu kvantitativně, konkrétně, umožňuje racionální debatu. Politici.

Součást vzdělání každého přirodovědně nebo technicky vzdělaného člověka.

Základy pravděpodobnosti a statistiky, základní úvahy, atd. by ale měl projít každý vzdělaný člověk, jak je jasné z výše uvedeného.

V tom je jiná než jiné oblasti matematiky.

V tomto kurzu se dozvíme, kromě základních pojmech a metodách pravděpodobnosti a statistiky, i o použití v informatice. Typické příklady: analýza časové složitosti v průměrném případě, generování náhodných čísel, očekávaný počet struktur daného typu v datech, apod.

Již v úvodu je dobré podotknout, že teorie pravděpodobnosti má i významnou filozofickou složku (náboj). V přirozeném jazyce běžně používáme výrazu jako „je pravděpodobné, že ...“ apod. Co to ale je pravděpodobnost? Zkuste se dobrat k rozumné definici, která by byla obecná, tj. dala by se použít v co největším počtu případů, kdy v přirozeném jazyce výraz pravděpodobnost používáme.

Teorie pravděpodobnosti: přesná věda o náhodnosti a nejistotě. Zdánlivý paradox, ale není. Náhodnost a nejistotu je ale potřeba matematicky uchopit, tak, abychom o náhodnosti a nejistotě mohli uvažovat tak, jak je v matematické obvyklé, tj. mohli definovat pojmy, o těchto pojmech dokazovat tvrzení (matematické věty), na jejich základě vytvářet metody apod.

1.2 Omyly intuice

Uvedeme několik příkladů a k některým z nich se později vrátíme.

Příklad 1.1. Co má větší pravděpodobnost při hodu dvěma kostkami, že padne součet 7 nebo součet 8?

Chybná úvaha: Máme 3 možnosti pro 8: $2+6$, $3+5$, $4+4$; máme 3 možnosti pro 7: $1+6$, $2+5$, $3+4$. Obě možnosti (padne 7, padne 8) jsou stejně pravděpodobné.

Správná úvaha: Chyba předchozí úvahy je v tom, že ve skutečnosti zahrnuje možnost $2+6$ (tj. výsledek, že padne 2 a 6), dva výsledky. První je, že na první kostce padne 2 a na druhé kostce 6, druhá, že na první kostce padne 6 a na druhé 2. Každý výsledek bychom tedy správně měli chápat jako uspořádanou dvojici $\langle x, y\rangle$, kde x je číslo, které padlo na první kostce, a y číslo na druhé kostce. Výsledků příznivých tomu, že padne součet 7, je 6, totiž

$$\langle 1,6\rangle, \langle 6,1\rangle, \langle 2,5\rangle, \langle 5,2\rangle, \langle 3,4\rangle, \langle 4,3\rangle,$$

zatímco výsledků příznivých součtu 8 je 5, totiž

$$\langle 2,6\rangle, \langle 6,2\rangle, \langle 3,5\rangle, \langle 5,3\rangle, \langle 4,4\rangle.$$

Protože výsledků příznivých součtu 7 je více, je možnost, že padne součet 7, pravděpodobnější než možnost, že padne součet 8.

Příklad 1.2. [2] Paule je 31 let. Je svobodná, přímočará, chytrá. Studovala filozofii. Když byla studentkou, horlivě podporovala práva původních obyvatel Ameriky (indiánů) a účastnila se protestů proti obchodnímu domu, který neměl zařízení pro kojící matky. Očíslujte následující tvrzení podle jejich pravděpodobnosti od 1 (nejpravděpodobnější) po 6 (nejméně pravděpodobné):

__ (a) Paula je aktivní feministka.

__ (b) Paula je bankovní úřednice.

__ (c) Paula pracuje v malém knihkupectví.

__ (d) Paula je bankovní úřednice a aktivní feministka.

__ (e) Paula je bankovní úřednice a aktivní feministka, která cvičí jógu.

__ (f) Paula pracuje v malém knihkupectví a je aktivní feministka, která cvičí jógu.

Slavný příklad, kteří studovali psychologové Amos Tversky a Daniel Kahneman (v roce 2002 mu byla udělena Nobelova pamětní cena za ekonomii, tzv. Nobelova cena za ekonomii). Zjistili, že vetšina lidí považuje za nejpravděpodobnější (f). Lidé často seřadili tvrzení takto: (f), (e), (d), (a), (c), (b).

To je ovšem nesprávné. Pravděpodobnost (f) musí být vždy $\leq$ pravděpodobnost (a) i než pravděpodobnost (c), protože z (f) vyplývá (a) i (c), rozvést. Uvědomme si to: Kdykoli platí (f), platí i (a), tedy (a) musí být aspoň tak pravděpodobné jako (f), tj. $P(a) \geq P(f)$.

Tverskyho a Kahnemanovo zjištění může znamenat, že lidé neusuzují v souladu se zákony pravděpodobnosti a že jsou v tomto smyslu iracionální. Může to ale znamenat i něco jiného, totiž to, že jsou nepozorní a že ve skutečnosti odpovídají na jinou otázku. Na jakou?: Které tvrzení je to nejpřesnější, nejužitečnější (nejvíce informativní) a zároveň dostatečně pravděpodobné?

Nejpravděpodobnější tvrzení totiž může být velmi málo užitečné. Například tvrzení „Zítra bude teplota $\leq 20°$C nebo $> 20°$C“ je jisté, a má tedy nejvyšší možnou pravděpodobnost (1), ovšem není užitečné (je to tautologie, platí vždy a nedává žádnou informaci).

V každém případě jsme i tímto příkladem varování, že při úvahách o pravděpodobnosti a nejistotě musíme být po všech stránkách opatrní.

Příklad 1.3. Kolo rulety má 38 políček: 18 černých, 18 červených a 2 zelená. Vsadím-li 100 Kč na černé a černé skutečně padne (tj. po roztočení kola se kulička zastaví na černém poli), vyhraju 200 Kč. Vsadím-li 100 Kč na červené a červené skutečně padne, vyhraju 200 Kč. Jinak prohraju.

Předpokládejme, že patnáctkrát po sobě padlo černé. Hráč vsadí peníze na červené, protože uvažuje následovně. Brzy musí padnout červené. Kolo je totiž spravedlivé, tj. pravděpodobnost, že se na políčku kulička zastaví, je stejná pro všechna políčka. Protože červené už dlouho nepadlo, je pravděpodobnost, že teď nebo brzy padne, větší než pro políčko černé. Je tato úvaha správná?

Úvaha správná není. Příslušná chyba, která se nazývá „the gambler's fallacy“, spočívá v tom, že vycházíme z toho, že předchozí výsledky (tj. kam spadla kulička) ovlivňují výsledky, které po nich následují. Ve skutečnosti jsou tyto výsledky nezávislé (někdy se v tomto případě říká, že ruleta nemá paměť).

Příklad 1.4. Předpokládejme, že v sázecí soutěži si účastník zvolí 6 čísel ze 49. Pokud budou vylosována právě čísla, která zvolil, vyhraje první cenu (ta se dělí mezi případných více výherců). Můžete si zdarma vybrat jeden z následujících lístků.

A: 1, 2, 3, 4, 5, 6

B: 5, 12, 25, 29, 39, 42

Který si vyberete?

Někteří lidé zvolí B, protože jim připadá málo pravděpodobné, že by byla vylosovaná zrovna takto pravidfelná posloupnost čísel. Ale to je omyl. Každá možnost je stejně pravděpodobná.

Jedna z nich ale skutečně může být výhodnější s ohledem na dělení hlavní výhry mezi více výherců. Úvaha: témeř nikdo nezvolí A, tak to zvolím já. Pokud náhodou vyhraju (pravděpodobnost, že vyhraju, je stejná jako u jakékoli jiné volby čísel), pak se výhra rozdělí mezi méně lidí. Není ovšem radno spoléhat se na to, že tímto způsobem ostatní přechytračím, jak ukazuje příklad z Velké Británie. Když tam v druhé polovině dvacátého století byla státní loterie zavedena, mnoho lidí právě z výše uvedeného důvodu volilo možnost 1, 2, 3, 4, 5, 6 [2].

Příklad 1.5. [2] Lékař považuje za velmi pravděpodobné, že pacient má streptokokový zánět nosohltanu, ale není si jistý. Vezme pacientovi vzorky a pošle je do laboratoře. Je známo, že laboratorní test není zcela spolehlivý: Ma-li pacient zánět, test dopadne pozitivně (tj. výsledek je diagnoáza zánětu, A) v 70% případů, negativně (N) ve 30% případů. Nemál-li pacient zánět, text je v 90% případů negatinví, v 10% pozitivní. Lékař poslal do laboratoře pět vzorků daného pacienta a dostal zpět výsledky

A, N, A, N, A.

Co z toho vyplývá?:

(a) Z výsledků nelze nic usoudit.

(b) Je pravděpodobnější, že pacient nemá zánět.

(c) Je pravděpodobnější, že pacient má zánět.

(c) Je monohem pravděpodobnější, že pacient má zánět, než že ho nemá.

Varianta 1: pravděpodobnost, že pacient má nemoc, je dle lékaře 0.9.

Varianta 2: Lékař neví nic (pacienta neviděl, k dispozici jsou jen pacientovy vzorky).

Příklad 1.6. Problém Montyho Halla. Televizní soutež spočívá v následujícím. Ve studiu jsou troje zavřené dveře. Za jedněmi je auto, za každými ze zbývajícíh pak koza. Soutěžící ukáže na dveře, řekněme první. Moderátor, který ví, co je za dveřmi, otevře třetí dveře. Za nimi se objeví koza. Zeptá se soutežícího, zda trvá na prvních dveřích nebo chce volbu změnit na dveře druhé. Pokud soutežící zvolí dveře, za kterými je auto, pak to auto vyhraje. Je rozumné volbu změnit?

Úloha je založena na televizní show, kterou moderoval Monty Hall a která vzbudila ohromný ohlas. Když byla v jistém časopise zveřejněna spolu se správným řešením, redakce časopisu obdržela kolem 10 000 dopisů čtenářů, z nichž naprostá většina s uvedeným řešením nesouhlasila. Traduje se dokonce, že jeden z nejproduktivnějších matematiků, Paul Erdős, nebyl zveřejněným řešením přesvědčen, dokud neuviděl výsledky počítačové simulace, která řešení potvrdila.

Jaká je tedy správná odpověď? Překvapivě snadná: Poté, co moderátor otevře dveře s kozou, je pravděpodobnost, že auto je za prvními dveřmi stejná, jako že je za druhými dveřmi. Je tedy jedno, na jaké dveře soutěžící ukazuje, a není tedy důvod volnu měnit. . . . Problém je v tom, že opak je pravdou. Soutěžící by měl volbu změnit a ukázat na druhé dveře.

1.3 Stručně z historie

Od 17. století. Vrátíme se k tomu.

1.4 Obsah textu

2 Potřebné pojmy

2.1 Množiny

Základní pojmy v rozsahu úvodního kurzu. Viz např. Bělohlávek R., *Úvod do informatiky*, UP Olomouc, 2008, kap. 2.2, `http://belohlavek.inf.upol.cz/vyuka/UvodDoInformatiky.pdf`.

Stručně shrňme:

- Množiny chápeme intuitivně jako soubory prvků. Množiny značíme zpravidla velkými písmeny, jejich prvky malými. $\omega \in A$ znamená, že prvek ω patří do množiny A, $\omega \notin A$ pak, že tam nepatří.
- $A \subseteq B$ (množina A je podmnožinou množiny B) znamená, že pro každý prvek x platí: pokud $x \in A$, pak $x \in B$. Obvykle předpokládáme, že všechny prvky se berou z nějaké základní množiny, tzv. univerza, kterou značíme např. U.
- Základní operace s množinami:

$$\begin{aligned} A \cup B &= \{x \in U \mid x \in A \text{ nebo } x \in B\} \text{ (sjednocení)} \\ A \cap B &= \{x \in U \mid x \in A \text{ a } x \in B\} \text{ (průnik)} \\ \overline{A} &= \{x \in U \mid x \notin A\} \text{ (doplněk, komplement)} \\ A \times B &= \{\langle x, y\rangle \mid x \in A \text{ a } x \in B\} \text{ (kartézský součin)} \end{aligned}$$

- De Morganovy zákony: $\overline{A \cup B} = \overline{A} \cap \overline{B}$ a $\overline{A \cap B} = \overline{A} \cup \overline{B}$.

2.2 Kombinatorika

Základní pojmy v rozsahu úvodního kurzu. Viz např. Bělohlávek R., *Úvod do informatiky*, UP Olomouc, 2008, kap. 4, `http://belohlavek.inf.upol.cz/vyuka/UvodDoInformatiky.pdf`:

- pravidlo součtu, pravidlo součinu
- permutace, permutace s opakováním
- variace, variace s opakováním
- kombinace

2.3 Teorie míry

Zatím nic.

2.4 Diferenciální a integrální počet

Úvod v rozsahu základního vysokoškolského kurzu.

3 Základní pojmy z pravděpodobnosti

3.1 Pojem pravděpodobnosti

3.1.1 Co je to pravděpodobnost?

Pravděpodobnost je pravým průvodcem našeho života.
Marcus Tullius Cicero (106 př. Kr.–43 př. Kr.), *De Natura*

Základní otázka; chceme-li budovat teorii pravděpodobnosti, je třeba začít touto otázkou. Podrobnější zamyšlení ukáže, že otázka není jednoduchá, lze na ni odpovědět několika odlišnými způsoby, které vedou k různým teoriím pravděpodobnosti.

V běžné mluvě:

(a) Je pravděpodobné, že vlak v 10:35 nestihnu.

(b) Pravděpodobnost, že padne sudé číslo je větší než pravděpodobnost, že padne číslo 5.

(c) Pravděpodobnost, že bude dnes pršet, je vysoká.

(d) Pravděpodobnost, že bude dnes pršet, je 60%.

(e) Pravděpodobnost, že padne sudé číslo je 0.5.

(f) Pravděpodobnost, že vyhynutí dinosaurů způsobil náraz meteoritu, je aspoň 80%.

V (a) je „je pravděpodobné“ použito jako modalita (unární spojka, podobně jako „je možné, že“, „není pravda, že“). V (b): porovnávající tvrzení. (a) i (b) neobsahují hodnotu pravděpodobnosti, obě jsou kvalitativní tvrzení o pravděpodobnosti, obě jsou užitečná. (c): hodnota je popsána slovně („vysoká“), ne pomocí číselné hodnoty. (d)–(f) je použita hodnota. Poznámka: říct 60% je stejné jako říct 0.6, je to věc zvyku, v matematické teorii pravděpodobnosti se používá 0.6.

Tedy: o pravděpodobnosti je možné pronášet tvrzení kvalitativní (např. (a), (b)), i kvantitativní (např. (d)–(f)), i jiné (např. (c)). Tyto různé možnosti vedou k různým koncepcím, jak budovat teorii pravděpodobnosti. Takových koncepcí bylo vypracováno pozoruhodné množství a tyto koncepce jsou spojeny s velkými jmény v oblasti matematiky, logiky a filozofie. Velmi dobrý přehled podává kniha [1].

My se budeme věnovat koncepci kvantitativní. Jaký význam ale mají tvrzení (d), (e), (f), tj. co to znamená, že *pravděpodobnost nečeho je* 0.6 (nebo, cheme-li 60%)? Existují dva základní přístupy k této otázce, *frekvenční* (angl. frequency) a *doměnkový* (angl. belief).

Frekvenční: „pravděpodobnost φ je 0.6“ v zásadě znamená, že při opakovaném pozorování situace, ve které φ může nebo nemusí nastat, φ nastává s relativní frekvencí 0.6 (např. v 6 případech z 10, 60 z 100, 300 z 500 atd.). Typický příklad, kde pravděpodobnost má tento význam, je (e) výše. Doměnkový: „pravděpodobnost φ je 0.6“ v zásadě znamená, že 0.6 vyjadřuje míru naší důvěry v φ; typický příklad je (f) výše. Oba přístupy však spíše představují směry, v rámci každého z nich bylo vypracováno několik konkrétních teorií. Poznamenejme také, že frekvenční a doměnový aspekt se prolínají a doplňují. Uvažme (d): Frekvenční pohled je tento: z opakovaných pozorování v minulosti plyne, že když byly klimatické podmínky podobné, pak v 60% případů pršelo. Doměnkový: Moje znalost a zkušenosti, mě vedou k 0.6 jako rozumné míře důvěry v to, že bude pršet. Máme tedy dva různé výklady. Často, jako v tomto příkladě, lze však doměnkový přístup redukovat nebo aspoň částečně redukovat na frekvenční: Nejsou ty znalosti a zkušenosti dány jen nebo v určující míře tím, co se dělo v minulosti, tj. nejsou podloženy frekvenčním pohledem? U příkladu (f) takto redukovat nelze, resp. si lze podobnou argumentaci představit stěží.

My se budeme věnovat teorii pravděpodobnosti, kterou vytvořil Andrej N. Kolmogorov (1903–1987) [3]. Ta je kvantitativní a vychází z frekvenčního přístupu. Na druhou stranu je to obecná teorie a frekvenční pohled je pouze jedna, byť hlavní, z interpretací této teorie.

3.1.2 Intuitivní přístupy k definici pravděpodobnosti

Jde nám tedy o tvrzení jako „pravděpodobnost něčeho je 0.6“. Za prvé, co je to *něco*, čemu pravděpodobnost připisujeme (přiřazujeme)? V běžném jazyce to jsou výroky, např. „padne šestka“, „padne sudé číslo“, „vybraný výrobek je vadný“ apod. Obecně máme tedy výrok φ (φ je tedy třeba některý ze tří uvedených výroků) a hovoříme o jeho pravděpodobnosti; řekneme třeba „Pravděpodobnost φ je 0.6.“, popř. „Pravděpodobnost, že φ je 0.6“, což můžeme zapsat symbolicky takto: $P(\varphi) = 0.6$.

Výrok φ popisuje nějakou událost. V tomto smyslu také hovoříme o pravděpodobnosti této události. Do jisté míry je jedno, zda mluvíme o pravděpodobnosti výroku nebo události. Mluvit o události má ale jisté výhody, které nyní ozřejmíme.

Jak se vypořádat s otázkou „Jaká je pravděpodobnost, že při hodu kostkou padne sudé číslo?“ Zdravý rozum nás vede k následující úvaze (tak se k výuce pravděpodobnosti přistupuje na středních školách). Při hodu kostkou může padnout šet čísel, existuje šest možných výsledků. Události „padne sudé číslo“ odpovídají 3 z těchto 6 možností. Protože každá z těchto možností je stejně pravděpodobná, je pravděpodobnost, že padně sudé číslo $\frac{3}{6} = 0.5$. Obecné pravidlo pro výpočet pravděpodobnosti $P(A)$ události A, které jsme při tom použili zní:

$$\text{jsou-li všechny možné výsledky stejně pravděpodobné, pak}$$
$$P(A) = \frac{\text{počet možných výsledků, při kterých } A \text{ nastane}}{\text{počet všech možných výsledků}} \tag{1}$$

Toto pravidlo se někdy nazývá *klasická definice pravděpodobnosti*. Vytvořil ji jeden z průkopníků pravděpodobnosti, Pierre-Simon Laplace (1749–1827). Je založeno na intuitivním chápání pojmu událost a výsledek. Jak ukážeme v poznámce 3.3, je klasická definice pravděpodobnosti speciálním případem obecné definice, která se v moderní teorii pravděpodobnosti používá.

Dalším intuitivním přístupem k definici pravděpodobnosti nějaké události je přístup známý jako *geometrická pravděpodobnost*. Princip je následující.

Terč je kruh o poloměru $r = 25$ cm. Uprostřed terče je namalován čtverec o straně $l = 5$ cm. Na terč je vystřelen náboj (předpokládejme pro jednoduchost, že je to ideální hmotný bod, tj. má nulový rozměr). Náboj zasáhne terč a pravděpodobnost zásahu libovolného bodu v terči je stejná. Jaká je pravděpodobnost, že náhoj zasáhne zmíněný čtverec? Zdravý rozum nás vede k tomu, že pravděpodobnost $P(A)$ zásahu oblasti A je rovna podílu

$$P(A) = \frac{\text{obsah oblasti } A}{\text{obsah celé oblasti}} \tag{2}$$

V našem případě je $P(A) = \frac{l^2}{\pi \cdot r^2} = \frac{5\cdot}{\pi \cdot 25^2} \approx 0.013$, tj. asi 1.3%.

Výše uvedené procesy, tj. házení kostkou a střelba na terč se v teorii pravděpodobnosti nazývají *náhodné pokusy*. Termín „náhodný pokus“ je ovšem *terminus technicus*, který může označovat skutečný pokus, např. pokus v laboratoři (smícháme dvě látky a jde o to, jestli dojde k výbuch, nebo ne), může označovat něco, co lze s trochou dobré vůle nazvat pokusem (např. hod kostkou), ale i něco, co bychom v běžné mluvě pokusem nenazvali (losujeme zemi, ve které proběhne mistrovství světa ve fotbale; náhodně vybereme obyvatele České republiky a zjistíme jeho hmotnost). Každý náhodný pokus má několik možných výsledků. Množina Ω možných výsledků může být konečná, jako v případě hodu kostkou, kde je $\Omega = \{1, 2, 3, 4, 5, 6\}$. Může být ale nekonečná, jako v případě střelby na terč, kde $\Omega = \{\langle x, y\rangle \mid x^2 + y^2 \leq r^2\}$ je množina všech souřadnic bodů v daném kruhu se středem v bodě 0 a poloměrem r. Události je pak možné přirozeně reprezentovat podmnožinami množiny Ω. Například událost „padne sudé číslo“ reprezentujeme množinou $A = \{2, 4, 6\}$, tj. množinou výsledků, při kterých událost nastává, neboli množinou výsledků příznivých dané události. Událost „čtverec bude zasažen“ ve stejném duchu reprezentujeme množinou $A = \{\langle x, y\rangle \mid -5 \leq x, y \leq 5\}$. Termíny „událost“ (intuitivní) a „jev“ (přesně definovaný) tak můžeme volně zaměňovat, většinou ale budeme nadále hovořit o jevech.

3.1.3 Pravděpodobnostní usuzování (usuzování za nejistoty), induktivní logika.

rozdíl od deduktivní logiky, dříve byly knihy o logice na dvě části,

Historická poznámka o vývoji. Boole a jeho úvahy o pravděpodobnosti. POZDEJI DOPLNIT

3.2 Pravděpodobnostní prostory

…hodnota teorie pravděpodobnosti spočívá v tom, že náhodné události, když je pozorujeme souhrnně a ve velkém měřítku, vykazují nenáhodnou pravidelnost.
Andrej N. Kolmogorov (1903–1987)

Přistupme nyní k formalizaci. Množinu všech možných výsledků nějakého náhodného pokusu budeme značit Ω, její prvky budeme značit ω a nazývat je *elementárními jevy* (angl. elementary event), popř. *výsledky* (angl. outcome).

Jev je libovolná podmnožina množiny Ω. Jen některé jevy však budeme uvažovat, těm budeme říkat *pozorovatelné jevy*, popř. jen jevy, bude-li to v daném kontextu jasné. Hraniční případy jevů: Ω – *jev jistý* (vždy nastane), $\emptyset$ – *jev nemožný* (nikdy nenastane).

Definujeme-li jev jako množinu elementárních jevů, provádíme abstrakci v následujícím smyslu. Na jednu stranu můžeme každý přirozený jev, včetně jevů popsaných výše, reprezentovat jemu odpovídající množinou, tj. např. jev „padne sudé číslo“ množinou $\{2, 4, 6\}$. Na druhou stranu každou podmnožinu, kterou z nějakého důvodu nevyloučíme (viz dále), musíme chápat jako množinu reprezentující nějaký jev. To může být někdy krkolomné (např. množinu $\{1, 4, 5\}$ asi nelze chápat jinak než jako formalizaci poněkud krkolomně popsané události „padne 1 nebo 4 nebo 5“), ale v principu to nevadí.

Příklad 3.1. Uveďme příklady náhodných pokusů a odpovídajících množin elementárních jevů. Hod mincí: $\Omega = \{H, T\}$ (H zmanená „panna“ (head), T znamená „orel“ (tail)). Hod kostkou: $\Omega = \{1, 2, 3, 4, 5, 6\}$. Výběr koule z osudí s 50 koulemi: $\Omega = \{\omega_1, \dots, \omega_{50}\}$.

Odkud se ale vezme množina Ω? Je nějakým jednoznačným způsobem určena popisem úlohy, kterou máme vyřešit? Není (říci „příklady náhodných pokusů a odpovídajících množin elementárních jevů“ v předchozím příkladu bylo tedy zavádějící). Přitom otázka, jak zvolit množinu Ω, je zásadní. Její správná volba může výpočet pravděpodobnosti, popř. vyřešení jiného daného problému, usnadnit, ale i zkomplikovat. Volba Ω je prvním krokem, který musíme učinit, abychom mohli použít pojmy a metody teorie pravděpodobnosti pro řešení daného problému. Prvním krokem, který musíme učinit, abychom „napasovali“ pojmy teorie pravděpodobnosti na zadaný problém.[1]

Příklad 3.2. Házíme dvěma konstami. Jaká je pravděpodobnost, že padne součet 8? Jak určit množinu Ω elementárních jevů? Ukážeme tři možnosti.

(a) Možnosti výsledného součtu jsou 2 (padnou dvě jedničky), 3 (jednička a dvojka), $\ldots$, 12 (dvě šestky). Tato úvaha nás vede k volbě $\Omega = \{2, \ldots, 12\}$. Ω má 11 prvků.

(b) Za elementární jev můžeme považovat (neuspořádnou) dvojici $\{i, j\}$, která vyjadřuje, že na jedné kostce padne i, na druhé j a my nerozlišujeme, na které. Tedy $\Omega = \{\{1, 1\}, \{1, 2\}, \{1, 3\}, \ldots, \{6, 6\}\}$. Ω má 21 prvků (15 dvojic, které je možné vybrat z 6 možností: $15 = \binom{6}{2}$; 6 jednoprvkových množin $\{1, 1\}, \ldots, \{6, 6\}$).

(c) Za elementární jev budeme považovat dvojici čísel $\langle i, j\rangle$, která vyjadřuje, že na první kostce pade i a na druhé j. Pak $\Omega = \{\langle 1, 1\rangle, \langle 1, 2\rangle, \langle 2, 1\rangle, \ldots, \langle 6, 6\rangle\}$ a Ω má 36 prvků.

O žádné z těchto možností nelze říct, že to je „ta správná“. Záleží na tom, jak budeme se zvolenou množinou Ω pracovat. Pojďme tedy spočítat pravděpodobnost, že padne součet 8. Použijme Laplaceovo pravidlo, tj. stanovme zmíněnou pravděpodobnost, p, jako podíl počtu výsledků příznivých jevu „padne součet 8“ a počtu všech výsledků, tj. počtu prvků množiny Ω.

Při volbě (a) vychází $p = \frac{1}{11} \approx 0.091$, protože z 11 elementárních jevů v Ω je jen jeden příznívá, totiž 8. Při volbě (b) jsou příznivé $\{2, 6\}$, $\{3, 5\}$, a $\{4, 4\}$, tedy $p = \frac{3}{21} \approx 0.143$. Při volbě (c) jsou příznivé $\langle 2, 6\rangle$, $\langle 6, 2\rangle$, $\langle 3, 5\rangle$, $\langle 5, 3\rangle$, $\langle 4, 4\rangle$. Dostáváme tedy $p = \frac{5}{36} \approx 0.139$.

Jak je to možné? Úvahy týkající se případů (a) a (b) jsou totiž chybné. Problém je v tom, že jednotlivé elementární jevy v množinách Ω nejsou stejně pravděpodobné, jak to vyžaduje Laplaceovo pravidlo. Uvažme například elementární jevy $\{2, 6\}$ a $\{4, 4\}$ v případu (b). $\{2, 6\}$ představuje ve skutečnosti dva možné výsledky: na první kostce padne 2, na druhé 6, druhý výsledek je, že na první kostce padne 6, na druhé 2; $\{4, 4\}$ představuje jen jeden takový výsledek: na první i na druhé kostce padne 4. $\{2, 6\}$ má tedy dvakrát větší pravděpodobnost než $\{4, 4\}$, a použít Laplacovo pravidlo tedy nelze.

[1] Že je v matematice často možné dojít k výsledku různými postupy, je známá věc. Někdy jde pouze o to, že existuje více postupů výpočtu (např. výpočtu řešení rovnice). Zde se ovšem jedná o uchopení problému, které musíme provést, než začneme počítat.

Při volbě množiny elementárních jevů je tedy třeba přihlížet k úloze kterou máme řešit. Jako rozumná se však ukazuje zásada volit jako elementární jevy neredukovatelné, neagregované entity, jak jsme to udělali v případu (c) Příkladu 3.2.

Definujeme-li jev jako podmnožinu množiny Ω elementárních jevů, vzniká otázka, zda je rozumné za jev považovat každou podmnožinu Ω. Není. Důvod spočívá v tom, že ne každá podmnožina musí odpovídat nějakému přirozenému jevu, který můžeme pozorovat. Uvažme například hod dvěma kostkami a množinu $\Omega = \{\langle 1,1\rangle, \langle 1,2\rangle, \langle 2,1\rangle, \dots, \langle 6,6\rangle\}$. Mají-li kostky stejný vzhled, a nejsme-li je tedy schopni pohledem rozlišit, pak padne-li na první kostce 2 a na druhé 6, vnímáme to stejně, jako když padne na první kostce 6 a na druhé 2. Jsme tedy schopni pozorovat jev „padne 2 a 6“, ale nejsme schopni pozorovat jev „na první kostce padne 2 a na druhé 6“, ani jev „na první kostce padne 6 a na druhé 2“. Z tohoto pohledu je tedy rozumné mít mezi jevy, které uvažujeme, jev $\{\langle 2,6\rangle, \langle 6,2\rangle\}$, který odpovídá popisu „padne 2 a 6“, ale nemít mezi uvažovanými jevy ani jev $\{\langle 2,6\rangle\}$, ani $\{\langle 6,2\rangle\}$, které odpovídají popisům „na první kostce padne 2 a na druhé 6“ a „na první kostce padne 6 a na druhé 2“. Množina $\mathcal{A}$ jevů, které uvažujeme, nemusí tedy být rovna množině 2^Ω všech jevů, ale může být jen její podmnožinou, tj. $\mathcal{A} \subseteq 2^\Omega$. Takto vybraným jevům říkáme *pozorovatelné jevy* (angl. observable events); není-li to nezbytné, přívlastek „pozorovatelný“ vynecháváme. Kromě výše uvedeného důvodu však existuje další důvod, který je technicky zásadnější a který vysvětlíme níže (poznámka 3.1 (c)).

Je přirozené požadovat, aby množina pozorovatelných jevů splňovala určité podmínky. Je-li například A pozorovatelný jev („sudé číslo“, „prvočíslo“ apod.), měl by být pozorovatelným i jeho komplement, tj. jev $\overline{A} = \{\omega \in \Omega \mid \omega \notin A\}$ („liché číslo“, „číslo, které není prvočíslem“ apod.). Umíme-li totiž rozpoznat, že výsledek ω je některým z výsledků v A, pak bychom měli umět rozpoznat, že výsledek není v A. Podobně lze požadovat, aby množina pozorovatelných jevů byla uzavřena na operace sjednocení a průniku, tj. aby $A_1 \cup A_2 \in \mathcal{A}$ a $A_1 \cap A_2 \in \mathcal{A}$ pro každé $A_1, A_2 \in \mathcal{A}$. Protože množina elementárních jevů může být nekonečná (např. množina $\mathbb{R}$ reálných čísel), je vhodné požadavek na uzavřenost formulovat i pro nekonečná sjednocení a průniky. Množina $\mathcal{A}$ pozorovatelných jevů by tedy měla splňovat výše uvedené podmínky, a ty zajišťuje následující definice.

Definice 3.1. *σ-algebra* na neprázdné množině Ω je neprázdná podmnožina

$\mathcal{A} \subseteq 2^{\Omega}$ splňující následující podmínky:

(a) je-li $A \in \mathcal{A}$, pak $\overline{A} \in \mathcal{A}$;

(b) jsou-li $A_1, A_2, \dots \in \mathcal{A}$, pak $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

Poznámka 3.1. (a) σ-algebry tedy představují rozumné systémy pozorovatelných jevů. Dvojice $\langle \Omega, \mathcal{A} \rangle$, kde $\mathcal{A}$ je σ-algebra na Ω se někdy nazývá experiment (považuje se tedy za matematizaci neformálního pojmu experiment); v teorii míry se nazývá *měřitelný prostor*.

(b) Je-li Ω konečná, volíme obvykle $\mathcal{A} = 2^{\Omega}$. Výše uvedený příklad ale ukazuje, že to tak nemusí být vždy.

(c) Je-li Ω nekonečná, pak požadavek, aby $\mathcal{A} = 2^{\Omega}$ může být nesplnitelný. Je například známo následující.[2] Nechť $\Omega = [0, 1]$ a předpokládejme, že za pozorovatelné jevy vezmeme všechny podmnožiny Ω. Zajímá nás, zda existuje přiřazení pravděpodobností pozorovatelným jevům, které je rozumné v tom smyslu, že splňuje axiomy pravděpodobnosti, které uvedeme níže. Chceme navíc, aby toto přiřazení P splňovalo přirozenou podmínku $P([a, b]) = b - a$. Lze ukázat, že žádné takové přiřazení však neexistuje. Problém spočívá v tom, že chceme, aby takové přiřazení bylo definováno pro každou podmnožinu Ω.

Důležitým tvrzením je následující věta:

Věta 3.1. *Ke každé podmnožině $\mathcal{A} \subseteq 2^{\Omega}$ existuje nejmenší σ-algebra na Ω, která obsahuje $\mathcal{A}$. Taková algebra se značí $\sigma(\mathcal{A})$ a nazávý se σ-algebra generovaná systémem $\mathcal{A}$.*

Důkaz. Systém všech σ-algeber na Ω obsahujících $\mathcal{A}$ je neprázdný (obsahuje 2^{Ω}) a uzavřený na libovolné průniky (snadno se ověří). Odtud plyne tvrzení. Za prvé, 2^{Ω} je σ Podrobněji na přednášce. $\square$

Zvláštní důležitost má σ-algebra generovaná systémem otevřených intervalů v $\mathbb{R}$, tj. systémem $\mathcal{A} = \{(a, b) \mid a, b \in \mathbb{R}, a < b\}$. Ta se nazývá *$\sigma$-algebra všech borelovských podmnožin* množiny $\mathbb{R}$. Její prvky se nazývají borelovské množiny. Představuje tedy nejmenší rozumný systém jevů, po kterém požadujeme, aby jevem byl každý otevřený interval. Poznamenejme, že stejná σ-algebra vznikne, pokud místo otevřených budeme uvažovat

[2] I. P. Natanson, Teorija funkcij veščestvennoj peremennoj. Gosudarstvennoje izd. techniko-teoretičeskoj literatury,Moskva, 1957.

intervaly zleva uzavřené, zprava uzavřené, uzavřené, popř. intervaly zleva neomezené nebo zprava neomezené.

Každá dvojice $\langle \Omega, \mathcal{A} \rangle$, kde $\mathcal{A}$ je σ-algebra na Ω, tedy popisuje toto: Proces (náhodný pokus), který se snažíme popsat, má nějké možné výsledky a ty tvoří množinu Ω. Díky schopnostem a přostředkům, které máme (kognitivní schopnosti, přístroje), jsme schopni pozorovat určité jevy – to jsou prvky množiny $\mathcal{A}$. Abychom ale popsali náhodnou povahu daného procesu, musíme popsat pravděpodobnosti jevů, tj. musíme zadat ke každému jevu $A \in \mathcal{A}$ jeho pravděpodobnost $P(A)$. Trojici $\langle \Omega, \mathcal{A}, P \rangle$ pak nazveme pravděpodobnostní prostor. Jaké vlastnosti by ale takové přiřazení P mělo mít? Takové vlastnosti popsal zakladatel moderní teorie pravděpodobnosti Kolmogorov a jsou předmětem definice pravděpodobnostního prostoru (definice 3.2).

Než k této definici přistoupíme, ukažeme si, jak lze k těmto vlastnostem přirozeně dojít pomocí frekvenční interpretace pravděpodobnosti. Předpokládejme, že daný pokus opakovaně provádíme, řekněme n-krát, a že výsledky pokusu jsou postupně $\omega_1, \ldots, \omega_n$. Pozorujme vlastnosti relativní frekvence jevů. Relativní frekvenci jevu $A \subseteq \Omega$ označme $f(A)$ a definujme takto:

$$f(A) = \frac{\text{počet výsledků } \omega_i \text{ příznivých jevu } A}{n}, \text{ tj.}$$

$$f(A) = \frac{|\{\omega_i\,;\, 1 \leq i \leq n \text{ a } \omega_i \in A\}|}{n}.$$

Zřejmě platí $f(\emptyset) = 0$ a $f(\Omega) = 1$. Dále platí, že pokud $A \subseteq B$ (tj. jev B je obecnější než jev A neboli B vyplývá z A) pak $f(A) \leq f(B)$. Jsou-li A a B disjunktní, tj. $A \cap B = \emptyset$, pak $f(A \cup B) = f(A) + f(B)$. Frekvenční interpretace tedy nabízí určité vlastnosti a například ty výše uvedené se zdají být přijatelnými požadavky pro každé rozumné přiřazování pravděpodobností jevům. Jsou skutečně základem Kolmogorovovy definice pravděpodobnostního prostoru, ke které nyní přistoupíme.

Definice 3.2. Nechť $\langle \Omega, \mathcal{A} \rangle$ je σ-algebra na Ω. *Pravděpodobnostní prostor* je uspořádaná trojice $\langle \Omega, \mathcal{A}, P \rangle$, kde P je funkce splňující:

(a) $P(A) \geq 0$ pro každý $A \in \mathcal{A}$,

(b) $P(\Omega) = 1$,

(c) $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ pro každou posloupnost jevů $A_1, A_2, \ldots$, které jsou po dvou disjunktní, tj. $A_i \cap A_j = \emptyset$ pro $i \neq j$.

Poznámka 3.2. (a) Funkce P z Definice 3.2 se nazývá *pravděpodobnostní míra* (z hlediska teorie míry je to míra na $\mathcal{A}$, která splňuje dodatečnou podmínku $P(\Omega) = 1$; míra dle definice splňuje podmínky (a), (c) a $P(\emptyset) = 0$; poslední podmínku každá pravděpodobnostní míra splňuje, viz níže), někdy také *pravděpodobnostní rozdělení* (rozděluje pravděpodobnost mezi jevy), popř. pouze *pravděpodobnost*. Číslo $P(A)$ se nazývá *pravděpodobnost jevu* A.

(b) Rovnost v bodě (c) se nazývá σ-aditivita míry P. V součtu $\sum_{i=1}^{\infty} P(A_i)$ nezáleží na pořadí členů ve sčítané posloupnosti. To plyne z toho, že jednotlivé členy $P(A_i)$ jsou nezáporné dle (a); jde tedy o součet absolutně konvergentní řady, a takový součet nezáleží na pořadí sčítaných prvků. (řada $\sum a_i$ se nazývá absolutně konvergentní, jestliže je konvergentní řada $\sum |a_i|$, tj. jestliže konverguje posloupnost částečných součtů $s_n = |a_1| + \cdots + |a_n|$).

Speciálně tedy, je-li $A = \{\omega_{i_1}, \omega_{i_2}, \dots\}$ spočetná množina elementárních jevů, pak $P(A) = \sum_{j=1}^{\infty} P(\{\omega_{i_j}\})$, tedy součet $\sum_{j=1}^{\infty} P(\{\omega_{i_j}\})$ existuje a nezáleží v něm na pořadí členů.

(c) Přímým důsledkem σ-aditivity P je, že pro disjunktní jevy A a B je $P(A \cup B) = P(A) + P(B)$ (snadno se vidí). Tato podmínka se nazývá aditivita. Je-li Ω konečná, pak tato podmínka je ekvivalentní podmínce (c).

(d) Poznamenejme také, že pravděpodobnost může být axiomatizována mnoha způsoby. Některé axiomatizace za základní považují pojem podmíněké pravděpodobnosti, se kterým se seznámíme níže (ovšem jako s pojmem odvozeným, nikoli základním). Autorem prvních axiomů (lépe řečeno základních pravidel) týkajících se pravděpodobnosti byl Christiaan Huygens (1629–1695). Huygens ve skutečnosti nehovořil o pravděpodobnosti, jeho úvahy se týkaly hazardních her a zabývaly se očekávanou hodnotou výhry v dané hazardní hře. Výše uvedené Kolmogorovovy axiomy pocházejí z práce [3].

Věta 3.2. *Pro jevy libovolné A a B v pravděpodobnostním prostoru platí:*

(a) $P(\overline{A}) = 1 - P(A)$

(b) $P(\emptyset) = 0$

(c) je-li $A \subseteq B$, *pak* $P(A) \leq P(B)$

(d) $0 \leq P(A) \leq 1$

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

(f) princip inkluze a exkluze pro pravděpodobnosti

Důkaz. (a) Jevy A a $\overline{A}$ jsou disjunktní a platí tedy $1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A})$, odkud tvrzení ihned plyne.

(b) Dle (a) je $P(\emptyset) = P(\overline{\Omega}) = 1 - P(\Omega) = 1 - 1 = 0$.

(c) Když $A \subseteq B$, pak $B = A \cup B - A$. Protože jevy A a $B - A$ jsou disjunktní, je $P(B) = P(A) + P(B - A)$. Protože $P(B - A) \geq 0$, je $P(B) \geq P(A)$.

(d) $0 \leq P(A)$ je podmínka (a) z Definice 3.2. Protože $A \subseteq \Omega$, plyne z (c), že $P(A) \leq P(\Omega)$. Protože z Definice 3.2 plyne $P(\Omega) = 1$, je tvrzení dokázáno.

(e) Je $A = (A-B) \cup (A \cap B)$, $B = (B-A) \cup (A \cap B)$ a $A \cup B = (A-B) \cup (A \cap B) \cup (B - A)$, přitom $A - B$, $B - A$ a $A \cap B$ jsou po dvou disjunktní. Z aditivity P tedy plyne $P(A \cup B) = P((A - B) \cup (A \cap B) \cup (B - A)) = P(A - B) + P(A \cap B) + P(B - A)$. Platí tedy

$$\begin{aligned} P(A \cup B) &= P(A-B) + P(A \cap B) + P(B - A) = \\ &= P(A-B) + P(A \cap B) - P(A \cap B) + P(A \cap B) \\ &\quad + P(B - A) + P(A \cap B) - P(A \cap B) = \\ &= P((A - B) \cup A \cap B) + P((B - A) \cup (A \cap B)) - P(A \cap B) = \\ &= P(A) + P(B) - P(A \cap B). \quad \square \end{aligned}$$

Pravděpodobnostní míra na konečné množině Ω jednoznačně určena pravděpodobnostmi elementárních jevů. Je-li $\Omega = \{\omega_1, \ldots, \omega_n\}$, pak označíme-li $p_i = P(\{\omega_i\})$, platí díky aditivitě P pro libovolný jev $A \in \mathcal{A}$

$$P(A) = \textstyle\sum_{\omega_i \in \mathcal{A}} p_i. \tag{3}$$

Pravděpodobnosti elementárních jevů splňují

$$0 \leq p_i \leq 1 \text{ pro každé } i = 1, \ldots, n, \quad \text{a} \quad \textstyle\sum_{i=1}^{n} p_i = 1. \tag{4}$$

Při daném pořadí pvků Ω, např. $\omega_1, \ldots \omega_n$, je tedy míra P je tedy jednoznačně určena n-ticí $\langle p_1, \ldots, p_n \rangle$ a tato n-tice splňuje (4). Na druhou stranu každá n-ticí $\langle p_1, \ldots, p_n \rangle$ reálných čísel splňujících (4) určuje pravděpodobnostní míru dle (3). Zadat n-tici $\langle p_1, \ldots, p_n \rangle$ je jednodušší než zadat míru P, a proto se v případě konečné množiny zadává pravděpodobnostní prostor obvykle tím, že zadáme n-tici $\langle p_1, \ldots, p_n \rangle$. Ta se potom obvykle nazývá *pravděpodobnostní rozdělení*.

Poznámka 3.3 (klasická pravděpodobnost). Podívejme se znovu, tentokrát z hlediska pojmu pravděpodobnostní prostor, na Laplaceovu klasickou definici pravděpodobnosti (1). Odpovídá jí pravděpodobnostní prostor, ve kterém $\Omega = \{\omega_1, \ldots, \omega_n\}$ je množina všech možných výsledků a $p_i = P(\{\omega_i\}) = \frac{1}{n}$. Z aditivity P je $P(A) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} \frac{1}{n} = |A| \cdot \frac{1}{n} = \frac{|A|}{n}$. To je ale právě výsledek, který dává Laplaceův vzorec (1), neboť výsledky, při kterých nastane daný jev A, jsou právě prvky množiny A a těch je $|A|$. Klasická definice pravděpodobnosti je tedy speciálním případem pravděpodobnosti definované podle Kolmogorovovy definice pravděpodobnostního prostoru.

Příklad 3.3. Balíček mariášových karet obsahuje karty 4 barev (červené, zelené, žaludy, kule) a 8 hodnot (od sedmičky po eso), celkem 32 karet. Náhodně vytáhneme tři karty. Jaká je pravděpodobnost, že:

(a) jako druhá karta bude vytaženo červené eso;

(b) jako druhá karta bude vytažena červená nebo žaludy;

(c) budou vytažena 3 esa;

(d) bude vytaženo aspoň jedno eso?

(a): Potřebujeme vzít v úvahu pořadí karet. Za elementární jevy (výsledky, prvky množiny Ω) tedy považujme uspořádané trojice $\langle k_1, k_2, k_3 \rangle$, kde k_i označuje i-tou vybranou kartu. Takových trojic je $32 \cdot 31 \cdot 30$ (jde o variace 3 z 32), tj. máme $|\Omega| = 32 \cdot 31 \cdot 30$. Výsledky příznivé danému jevu A (druhá karta bude červené eso) jsou právě trojice, pro které k_2 je červené eso. Těch je $31 \cdot 30$ (dle pravidla součinu: k_1 lze zvolit 31 způsoby, k_3 pak nezávisle na tom 30 způsoby). Pravděpodobnost výběru každé trojice je stejná: $\frac{1}{32 \cdot 31 \cdot 30}$. Jde tedy o klasickou pravděpodobnost (viz poznámka 3.3), a proto $P(A) = \frac{|A|}{|\Omega|} = \frac{31 \cdot 30}{32 \cdot 31 \cdot 30} = \frac{1}{32}$.

Jinou úvahou: Je zřejmé, že pro každou kartu je pravděpodobnost p, že bude vytažena jako druhá, stejná. Protože karet je 32 a protože součet těchto pravděpodobností je 1, je $p = \frac{1}{32}$.

(b): Zvolme Ω stejně jako v (a). Jev A, tj.

$$A = \{\langle k_1, k_2, k_3 \rangle \in \Omega \mid k_2 \text{ je červená nebo žaludy}\},$$

má $(8 + 8) \cdot 31 \cdot 30$ prvků. Totiž, na pozici k_2 lze zvolit libovolnou z 8 červených nebo libovolnou z 8 žaludských karet. Zbylé karty lze po volbě karty k_2 zvolit, stejně jako v (a), $31 \cdot 30$ způsoby. Je zřejmé, že tímto způsobem dostaneme všechny trojice, které jsou příznivé jevu A. Dostáváme tedy $P(A) = \frac{|A|}{|\Omega|} = \frac{16 \cdot 31 \cdot 30}{32 \cdot 31 \cdot 30} = \frac{1}{2}$.

Pokud bychom příznivé výsledky popisovali od volby první karty, mohli bychom postupovat takto. k_1 lze zvolit 32 způsoby. Nyní je třeba rozlišit, zda k_1 je, nebo není jednou z červených nebo žaludských karet. Pokud je, tj. k_1 je jednou z 16 červených nebo žaludských, lze k_2 zvolit 15 způsoby (zbývá 15 červených nebo žaludských), poté pak zvolit k_3 jedním ze 30 zbývajících způsobů; tím dostáváme $16 \cdot 15 \cdot 30$ výsledků. Pokud není, tj. k_1 je jednou z 16 zelených nebo kulových, pak lze k_2 zvolit 16 způsoby, poté k_3 30 způsoby; dostáváme $16 \cdot 16 \cdot 30$ dalších způsobů. Tyto způsoby jsou navzájem různé, celkem tedy máme $16 \cdot 16 \cdot 30 + 16 \cdot 15 \cdot 30 = 16 \cdot (16+15) \cdot 30 = 16 \cdot 31 \cdot 30$ způsobů. Došli jsme tedy ke stejnému počtu výsledků příznivých jevu A, ovšem bylo to složitější. Vidíme tedy, že je důležité se vhodným způsobem na stituaci podívat.

(c) Zvolme Ω opět jako v (a). Protože jsou celkem 4 esa, je celkem $4 \cdot 3 \cdot 2$ trojic, ve kterých jsou 3 esa, tj. výsledků příznivých danému jevu A. Je tedy $P(A) = \frac{4 \cdot 3 \cdot 2}{32 \cdot 31 \cdot 30} = 0.0008$.

Jiný pohled. Považujme za výsledek množinu tří vybraných karet, tj. nikoli uspořádanou trojici $\langle k_1, k_2, k_3 \rangle$ jako výše, ale neuspořádanou trojici neboli tříprvkovou množinu $\{k_1, k_2, k_3\}$. Takových trojic je $\binom{32}{3}$. Trojic obsahujících jen esa je $\binom{4}{3}$. Konkrétně máme $\Omega = \{\{k_1, k_2, k_3\} \mid k_1 \neq k_2 \neq k_3 \neq k_1\}$ a

$$A = \{\{e_1, e_2, e_3\}, \{e_1, e_2, e_4\}, \{e_1, e_3, e_4\}, \{e_2, e_3, e_4\}\},$$

kde $e_1, \ldots, e_4$ označují čtyři esa. Hledaná pravděpodobnost je tedy

$$P(A) = \frac{\binom{4}{3}}{\binom{32}{3}} = \frac{\frac{4!}{3!}}{\frac{32!}{29!3!}} = \frac{4 \cdot 3 \cdot 2}{32 \cdot 31 \cdot 30},$$

a to je stejná hodnota, ke které jsme došli prvním způsobem.

(d) Tento příklad je jedním z častých případů, kdy je snadnější určit pravděpodobnost jevu $\overline{A}$, který je komplementární k danému jevu A. Zvolme Ω jako v (a). $\overline{A}$ je jev, který nastane, když nepadne žádné eso, tj. každá karta v trojici $\langle k_1, k_2, k_3 \rangle$ bude některou z 28 karet, které nejsou esa. Takových trojic je $28 \cdot 27 \cdot 26$. Dostáváme tedy $P(\overline{A}) = \frac{28 \cdot 27 \cdot 26}{32 \cdot 31 \cdot 30} \approx 0.66$ a $P(A) = 1 - P(\overline{A}) \approx 0.34$.

Ke stejnému výsledku opět dojdeme, budeme-li za elementární jevy brát tříprvkové množiny $\{k_1, k_2, k_3\}$. Trojic neobsahujících esa je $\binom{28}{3}$, tedy $P(\overline{A}) = \frac{\binom{28}{3}}{\binom{32}{3}}$, což po úpravách dává opět $\frac{28 \cdot 27 \cdot 26}{32 \cdot 31 \cdot 30}$.

Pokud bychom chtěli určit pravděpodobnost jevu A přímo, mohli bychom postupovat takto. Za výsledky berme opět neuspořádané trojice. Jev A (aspoň jedno eso) lze chápat jako sjednocení tří navzájem disjunktních jevů, A_1 (právě jedno eso), A_2 (právě dvě esa), A_3 (tři esa). Počet trojic v A_1 je $4 \cdot \binom{28}{2}$, protože trojice s právě jedním esem dostaneme tak, že zvolíme eso ($\binom{4}{1} = 4$ způsoby), zvolíme neuspořádanou dvojici karet, která jesou esy ($\binom{28}{2}$ způsobů), a eso do této dvojice přidáme. Podle pravidla součinu tak získáme $4 \cdot \binom{28}{2}$ různých trojic, tedy $|A_1| = 4 \cdot \binom{28}{2}$. Trojic s právě dvěma esy je podle podobné úvahy (zvolíme dvojici es a pak znývající kartu) $\binom{4}{2} \cdot \binom{28}{1} = \binom{4}{2} \cdot 28$. Trojic s třemi esy je $\binom{4}{3}$. Je tedy

$$\begin{aligned} |A| &= |A_1 \cup A_2 \cup A_3| = |A_| \cup |A_2| \cup |A_3| = \\ &= 4 \cdot \binom{28}{2} + \binom{4}{2} \cdot 28 + \binom{4}{3} = 4 \cdot 378 + 6 \cdot 28 + 4 = 1684. \end{aligned}$$

Je tedy

$$P(A) = \frac{|A|}{|\Omega|} = \frac{1684}{\binom{32}{3}} = \frac{1684}{4960} \approx 0.34.$$

Došli jsme tedy ke stejné hodnotě, ale složitěji.

Poznámka 3.4. Mnoho příkladů z pravděpodobosti pochází z hazardních her. Není to proto, že by tyto příklady byly nejužitečnějšími ukázkami použití teorie pravděpodobnosti, ale proto, že jsou jednoduché a všeobecně srozumitelné. Představují prototypy jednoduchých úloh, které se v různých obměnách vyskytují v jiných praktických úlohách. Je ale třeba říct, že problémy o hazardních hrách motivovaly rozvoj teorie pravděpodobnosti v jejích počátcích a že provozování hazardních her, loterií a podobně se bez teorie pravděpodobnosti neobejde. Pro provozovatele je cílem navrhnout hru tak, aby byla pro hráče pravděpodobnost výhry menší než pravděpodobnost prohry, a přitom tak, aby to nebylo patrné a hra byla lákavá. Hazard a loterie se provozují od nepaměti a jsou předmětem mnoha citátů a rčení. Uveďme dva z nich:

The best throw of the dice is to throw them away.
anglické přísloví

By gaming we lose both our time and treasure:
two things most precious to the life of man.
Owen Feltham (1602–1668)

Poznámka 3.5 (geometrická pravděpodobnost). Podobně jako klasická, je speciálním případem. Zde je $P(A)$ normovaný obsah oblasti A, tj. $P(A) = \mu(A)/\mu(\Omega)$, kde μ je vhodná míra v $\mathbb{R}^2$.

Příklad 3.4. Příklad na geometrickou pravděpodobnost.

Příklad 3.5. Jaká je pravděpodobnost, že náhodně vybrané k-ciferné číslo v desítkové soustavě představuje číslo v osmičkové soustavě?

V tomto případě lze zvolit $\Omega = \{c_1 \dots c_k \mid c_i \in \{0, \dots, 9\}\}$ a použít klasickou definici pravděpodobnosti. Elementárních jevů $c_1 \dots c_k$ je 10^k (lze je chápat jako variace k z 10 s opakováním. Jev, o který jde, je $A = \{d_1 \dots d_k \mid c_i \in \{0, \dots, 7\}\}$ a obsahuje 8^k elementárních jevů. Jeho pravděpodobnost je tedy $P(A) = \frac{|A|}{|\Omega|} = \frac{8^k}{10^k} = (\frac{4}{5})^k$.

Příklad 3.6. Uvažujme abecedu $\Sigma = \{0, \dots, 9, a, \dots, z\}$ (36 znaků). Jaká je pravděpodobnost, že v náhodně zvoleném slově sestaveném z 4 znaků abecedy Σ se znaky neopakují?

Zvolme $\Omega = \{a_1 \dots a_4 \mid a_i \in \Sigma\}$. Je zřejmě $|\Omega| = 36^4$. Jev A, který obsahuje slova s neopakujícími se znaky, obsahuje $36 \cdot 35 \cdot 34 \cdot 33$ slov (variace 4 z 36 bez opakování). Protože elementární jevy jsou stejně pravděpodobné, je $P(A) = \frac{|A|}{|\Omega|} = \frac{1413720}{1679616} \approx 0.84$.

Příklad 3.7. V dodávce zboží je 85 výrobků bezchybných a 15 vadných. Náhodně vybereme 10 výrobků. Jaká je pravděpodobnost, že mezi nimi bude aspoň jeden vadný? Protože na pořadí mezi 10 vybranými výrobky nezáleží, vezměme za Ω množinu všech 10-tiprvkových podmnožin množiny Ω. Jejich počet je roven počtu kombinací 10 z 100, tj. počtu způsobů, jak vybrat 10 prvků ze 100, což je $\binom{100}{10}$. Místo určení pravděpodobnosti daného jevu A, tj. „aspoň jeden vadný", je v tomto případě snazší určit pravděpodobnost jevu $\overline{A}$, tj. „žádný vadný". $\overline{A}$ obsahuje výsledky, tj. 10-ti prvkové podmnožiny, ve kterých jsou všechny výrobky bezvadné, a těch je $\binom{85}{10}$. Je tedy

$$P(\overline{A}) = \frac{\binom{85}{10}}{\binom{100}{10}} = \frac{\frac{85!}{75!10!}}{\frac{100!}{90!10!}} = \frac{85!90!}{75!100!} \approx 0.18$$

. Tedy $P(A) = 1 - P(\overline{A}) \approx 0.82$

Příklad 3.8. V dodávce je n výrobků, z nich je k bezchybných a $n - k$ vadných. Jaká je pravděpodobnost, že z p náhodně vybraných výrobků bude právě r vadných?

Za elementární jev považujme množinu p vybraných výrobků. Takových množin je $\binom{n}{p}$. Protože r vadných lze vybrat $\binom{n-k}{r}$ způsoby a zbývajících $p-r$ bezchybných lze vybrat $\binom{k}{p-r}$ způsoby, existuje $\binom{n-k}{r} \cdot \binom{k}{p-r}$ elementárních jevů, které jsou obsaženy v daném jevu A (r vadných a $p-r$ bezchybných). Je tedy

$$P(A) = \frac{\binom{n-k}{r}\binom{k}{p-r}}{\binom{n}{p}}.$$

Příklad 3.9. V krabici je n míčků, z nich je k bílých a $n - k$ černých. Míčky postupně vybíráme jeden za druhým. Jaká je pravděpodobnost, že první černý míček bude vytažen jako p-tý?

Výsledkem je situace, kdy jsou z krabice vybrány všechny míčky a je známo pořadí. Za elementární jevy tedy považujme uspořádané n-tice $\langle m_1, \ldots, m_n \rangle$, kde m_i je i-tý vybraný míček. Je tedy $n!$ elementárních jevů a všechny jsou stejně pravděpodobné. Jev A, jehož pravděpodobnost máme určit, sestává ze všech n-tic, které mají na prvních $p-1$ místech bílé míčky, na místě p černý a na zbylých místech míčky rozmístěné bez dalšího omezení. Vybrat prvních $p-1$ bílých míčků lze $\frac{k!}{(k-(p-1))!}$ způsoby (variace $p-1$ z k), vybrat pak černý míček lze $n-k$ způsoby, pak zbývá $(n-p)!$ možností, jak rozmístit zbývajících $n-p$ míčků. A má tedy $\frac{k!}{(k-p+1)!} \cdot (n-k) \cdot (n-p)!$ elementárních jevů. Pravděpodobnost jevu A je tedy

$$P(A) = \frac{k!(n-k)(n-p)!}{(k-p+1)!n!}.$$

Příklad 3.10. n různých objektů náhodně umísťujeme do k přihrádek. Jaká je pravděpodobnost, že daná přihrádka (tj. třeba první přihrádka) bude obsahovat právě p objektů?

Každé takové umístění si lze představit jako n-tici $\langle c_1, \ldots, c_n \rangle$ čísel $1 \leq c_i \leq k$, kde c_i je číslo přihrádky, do které bude umístěn i-tý objekt. První objekt můžeme umístit do libovolné z k přihrádek, druhý také atd., celkem tedy existuje k^n umístění. Počet umístění, pro které je v první přihrádce právě p objektů, tj. počet výsledků příznivých danému jevu A, určíme takto. Kterých p objektů bude v první přihrádce, lze zvolit $\binom{n}{p}$ způsoby. Zbývajících $n-p$ objektů lze umístit libovolně do zbylých $k-1$ přihrádek, což lze provést $(k-1)^{n-p}$ způsoby. Jev A tedy obsahuje $\binom{n}{p} \cdot (k-1)^{n-p}$ elementárních jevů, a jeho pravděpodobnost je tedy

$$P(A) = \binom{n}{p} \frac{(k-1)^{n-p}}{k^n}.$$

Cvičení.

1. Z balíčku 52 whistových karet (4 barvy, 12 hodnot) taháme ....
2. Jaká je pravděpodobnost, že čtyřciferné číslo zapsané v desítkové soustavě je:
   (a) sudé;
   (b) dělitelné pěti;
   (c) násobkem sta?
3. Jaká je pravděpodobnost, že čtyřmístné číslo vytvořené z cifer 1, 3, 5, 6 a 8 sudé, za předpokladu, že se jeho cifry nemohou opakovat? A jaká je za předpokladu, že se mohou opakovat?

3.3 Podmíněná pravděpodobnost

Podmíněná pravděpodobnost je jedním z nejdůležitějších pojmů teorie pravěpodobnosti. My tento pojem zavedeme jako odvozený z pojmu (nepodmíněná) pravděpodobnost, tj. z pojmu, který jsme zavedli jako základní. Někteří odborníci ale považují za základní pojem podmíněné pravděpodobnosti a pojem neprodmíněné pravděpodobnosti za odvozený. Tento postoj shrnují do hesla „neexistuje nic jiného než podmíněná pravděpodobnost" a myslí tím, že když hovoříme o pravděpodobnosti toho, že dnes bude pršet, máme ve skutečnosti na mysli pravděpodobnost, že dnes bude pršet, za podmínky, že v uplynulých dnech a hodinách bylo počasí takové a takové. Podobně místo pravděpodobnosti, že z balíčku karet jako druhou kartu vytáhneme eso, vlastně mluvíme o pravděpodobnosti, že jako druhou kartu vytáhneme eso za podmínky, že balíček je důkladně zamíchaný.

Předpokládejme, že nás zajímá pravděpodobnost jevu A a že, byť výsledek – rozhodující o tom, zda jev A nastal – neznáme, víme, že nastal jev B. Intuitivně se zdá být zřejmé, že naše očekávání jevu A bez informace o tom, že nastal jev B, je obecně jiné než očekávání jevu A za podmínek, kdy navíc víme, že nastal jev B. To první očekávání odpovídá námi dosud uvažované pravděpodobnosti $P(A)$, to druhé pak tomu, co se nazývá pravděpodobnost jevu A za předpokladu (nebo za podmínky), že nastal jev B, a co budeme značit $P(A|B)$.

Příkladem může být situace, kdy nás zajímá pravděpodobnost, že náhodně vybraný zaměstnanec z množiny Ω všech 50 zaměstnanců dané firmy má řidičský průkaz. Označne tento jev A. Za předpokladu, že pravděpodobnost výběru každého zaměstnance je stejná, je pravděpodobnost jevu A rovna podílu počtu zaměstnanců s řidičským průkazem ku počtu všech zaměnstnanců, tj. $P(A) = \frac{|A|}{|\Omega|}$. Je-li tedy zaměstnanců s řidičským průkazem 30, je

$$P(A) = \frac{30}{50} = 0.6.$$

Nechť B je jev popsaný jako „vybraný zaměstnanec je muž". Předpokládejme, že mužů je mezi zaměstnanci 20, tj. $|B| = 20$, a že 15 z nich řidičský průkaz má. Podmíněnou pravděpodobnost $P(A|B)$, tj. symbolicky $P($„má řidičský průkaz"$|$„muž"$)$, se zdá být přirozené definovat jako

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{15}{20} = 0.75, \tag{5}$$

neboť zaměstnanci, které nyní uvažujeme a vzhledem ke kterým pravděpodobnost posuzujeme, nejsou všichni zaměstnanci z Ω, ale jen mužští zaměstnanci, tj. prvky Ω, a z nich jsou přízniví zadanému jevu jen ti s řidičským průkazem, tj. ti z množiny $A \cap B$.

Vztah (5) můžeme upravit tak, aby neobsahoval výrazy $|A \cap B|$ a $|B|$, které jsou odrazem předpokladu, že výběr každého zaměstnance má stejnou pravděpodobnost. Dostaneme tak

$$P(A|B) = \frac{|A \cap B|}{|B|} = \frac{|A \cap B|/|\Omega|}{|B|/|\Omega|} = \frac{P(A \cap B)}{P(B)}. \tag{6}$$

Dostali jsme tak vzorec použitelný bez zmíněného předpokladu. Dospěli jsme tak k pojmu podmíněná pravděpodobnost:

Definice 3.3. Podmíněná pravděpodobnost jevu A za podmínky B je

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{7}$$

je-li $P(B) \neq 0$ a je nedefinována jinak.

Poznámka 3.6. (a) V případě klasické pravděpodobnosti má $P(A|B)$, která je daná vztahem (6), následující význam. Zatímco $P(A)$ je relativní podíl příznivých případů (příznivých danému jevu A) v množině všech výsledků, je $P(A|B)$ relativní podíl příznivých případů v množině výsledků, při kterých nastává B.

Význam lze ilustrovat Vennovým diagramem. DOPLNIT

(b) V obecném případě lze $P(A|B)$ chápat jako pravděpodobnost jevu $A \cap B$ „normalizovanou" pravděpodobností jevu B.

(c) Je tedy

$$P(A \cap B) = \begin{cases} P(B) \cdot P(A|B) & \text{pokud } P(B) > 0, \\ P(A) \cdot P(B|A) & \text{pokud } P(A) > 0, \\ 0 & \text{jinak.} \end{cases} \tag{8}$$

(d) Výše uvedený příklad ukázal, že může nastat případ, kdy $P(A) < P(A|B)$. Může ovšem nastat i opačný případ. Reprezentuje-li B jev „vybraný zaměstnanec je žena", pak snadno vidíme, že $P(A|B) = \frac{15}{30} = 0.5 < 0.6 = P(A)$.

Pojem podmíněná pravděpodobnost umožňuje zajímavé úlohy:

Příklad 3.11. V krabici je 1000 mikroprocesorů, z nich 700 je vyrobeno firmou x a 300 firmou y. Je známo, že 10% mikroprocesorů firmy x je vadných, od firmy y pak 5%. Jaká je pravděpodobnost, že náhodně vybraný vadný mikroprocesor pochází od firmy x? A jaká, že pochází od firmy y?

Za Ω zvolíme množinu všech mikroprocesorů v krabici. Uvažujme tyto jevy:

$$\begin{aligned} X &= \{\omega \in \Omega \mid \text{mikroprocesor } \omega \text{ je od firmy } x\}, \\ Y &= \{\omega \in \Omega \mid \text{mikroprocesor } \omega \text{ je od firmy } y\}, \\ C &= \{\omega \in \Omega \mid \text{mikroprocesor } \omega \text{ je vadný}\}. \end{aligned}$$

Máme tedy určit podmíněnou pravděpodobnost $P(X|C)$, což dle vzorce je $\frac{P(C \cap X)}{P(C)}$. Protože předpokládáme, že každý procesor má stejnou pravděpodobnost být vybrán, a protože $|C| = 0.1 \cdot 700 + 0.05 \cdot 300 = 85$, je $P(C) = \frac{85}{1000} = 0.085$. $C \cap X$ je množina vadných mikroprocesorů od firmy x a těch je dle předpokladu $0.1 \cdot 700 = 70$. Proto je $P(C \cap X) = \frac{70}{1000} = 0.07$. Dostáváme tedy

$$P(X|C) = \frac{P(C \cap X)}{P(C)} = \frac{70/1000}{85/1000} = \frac{70}{85} \approx 0.82.$$

Pravděpodobnost $P(Y|C)$ je pak dle stejné úvahy

$$P(Y|C) = \frac{P(C \cap Y)}{P(C)} = \frac{15}{85} \approx 0.18.$$

Pravděpodobnost $P(C \cap X)$ v příkladě 3.11 jsme získali díky předpokladu stejně pravděpodobných elementárních jevů. Jiný, obecnější způsob je tento: Skutečnost, že 10% výrobků od firmy x je vadných, vlastně znamená, že $P(C|X) = 0.1$; podobně $P(C|Y) = 0.05$. Potřebnou pravděpodobnost $P(C \cap X)$ tedy můžeme určit podle vzorce (8), $P(C \cap X) = P(C|X) \cdot P(X)$, který je přímým důsledkem definice podmíněné pravděpodobnosti, takto: $P(C \cap X) = P(C|X) \cdot P(X) = 0.1 \cdot 0.7 = 0.07$. zobecnění tohoto vztahu podává následující důsledek definice podmíněné pravděpodobnosti.

Důsledek 3.1. *Pokud pro jevy $A_1, \ldots, A_n$ platí $P(A_1 \cap \cdots \cap A_{n-1}) > 0$, pak*

$$P(A_1 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \cdots \cap A_{n-1}). \tag{9}$$

Důkaz. Uvědomme si nejdříve, že $P(A_1 \cap \cdots \cap A_{n-1}) > 0$ zajišťuje, že je i $P(A_1 \cap \cdots \cap A_i) > 0$ pro každé $i = 1, \ldots, n-1$, tj. každá z podmíněných pravděpodobností $P(A_{i+1}|A_1 \cap \cdots \cap A_i)$ je definována. Rozvnos pak dostaneme rozepsáním faktorů podle vzroce podmíněné pravděpodobnosti a zkrácením. Ukažme to pro tři jevy: $P(A) \cdot P(B|A) \cdot P(C|A \cap B) = P(A) \cdot \frac{P(A \cap B)}{P(A)} \cdot \frac{P(A \cap B \cap C)}{P(A \cap B)} = P(A \cap B \cap C)$. □

Následující tvrzení ukazuje důležitou skutečnost. Zafixujeme-li podmínku B, dostaneme pravděpodobnostní míru.

Věta 3.3. *Nechť $\langle \Omega, \mathcal{A}, P \rangle$ je pravděpodobnostní prostor a $B \in \mathcal{A}$ jev, pro který $P(B) > 0$. Pak funkce $P(\cdot|B)$ přiřazující každému $A \in \mathcal{A}$ hodnotu $P(A|B)$ je pravděpodbnostní míra na $\mathcal{A}$, tj. splňuje podmínky (a)–(c) definice 3.2.*

Důkaz. Označme danou funkci P_B, tj. $P_B(A) = P(A|B)$.

(a) plyne z toho, že $P(A|B)$ je podílem dvou nezáporných čísel.

(b): $P_B(\Omega) = P(\Omega|B) = \frac{P(\Omega \cap B)}{P(B)} = \frac{P(B)}{P(B)} = 1$.

(c): Jsou-li jevy A_i po dvou disjunktní, pak jsou zřejmě po dvou disjunktní i jevy $A_i \cap B$, a tedy

$$P_B(\bigcup_i A_i) = P(\bigcup_i A_i|B) = \frac{P((\bigcup_i A_i) \cap B)}{P(B)} = \frac{P(\bigcup_i (A_i \cap B))}{P(B)} =$$

$$= \frac{\sum_i P(A_i \cap B)}{P(B)} = \sum_i \frac{P(A_i \cap B)}{P(B)} = \sum_i P(A_i|B) = \sum_i P_B(A_i).$$

$\square$

Poznámka 3.7. (a) Hned je vidět, že $P(\Omega|B) = 1$; obecněji: je-li $A \supseteq B$, pak $P(A|B) = 1$. Na druhou stranu $P(\emptyset|B) = 0$; obecněji: je-li $A \cap B = \emptyset$, pak $P(A|B) = 0$.

(b) Uvedená věta říká, že $\langle \Omega, \mathcal{A}, P(\cdot|B) \rangle$ je pravděpodobnostní prostor. Je zřejmé, že pravděpodobnostním prostorem je i $\langle B, \mathcal{B}, P(\cdot|B) \rangle$, kde $\mathcal{B} = \{A \in \mathcal{A} \mid A \subseteq B\}$.

Cvičení.

1. Čtyři firmy, A, B, C, D, se v soutěži ucházejí o zakázku. Z minulosti je známo, že pravděpodobnosti získání podobných zakázek jsou pro tyto firmy postupně 0.35, 0.15, 0.3 a 0.2. Před rozhodnutím o vítězi firma B oznámí, že odstupuje ze soutěže. Jaké jsou nové pravděpodobnosti získání zakázky pro firmy A, C a D?

2. DOPLNIT

3.4 Nezávislost jevů

Jak jsme viděli, pravěpodobnost $P(A)$ může být menší než podmíněná pravděpodobnost $P(A|B)$, může být větší, ale může být i stejná.[3] Poslední možnost, tj. $P(A) = P(A|B)$, lze interpretovat následovně. Výskyt jevu A za obecných podmínek je stejně pravděpodobný jako v situaci, kdy víme, že nastal jev B. Dodatečná informace o tom, že nastal jev B, tedy nemá na pravděpodobnost jevu A žádný vliv. V takovém případě říkáme, že jevy A a B jsou nezávislé. Je ale takový vztah skutečně symetrický? Ano, protože pokud $P(A) = P(A|B)$, pak vzhledem k $P(A|B) = \frac{P(A\cap B)}{P(B)}$ je $P(B) = \frac{P(A\cap B)}{P(A|B)} = \frac{P(B\cap A)}{P(A)} = P(B|A)$. Pokud tedy výskyt jevu B nemá vliv na pravděpodobnost jevu A, pak také platí, že výskyt jevu A nemá vliv na pravděpodobnost jevu B a je přirozené říci, že A a B jsou navzájem nezávislé. Podmínku nezávislosti $P(A) = P(A|B)$ a s ní ekvivalentní podmínku $P(B) = P(B|A)$ lze však vyjádřit symetrickou podmínkou $P(A \cap B) = P(A) \cdot P(B)$. Platí-li totiž $P(A) = P(A|B)$, pak $P(A \cap B) = P(A|B) \cdot P(B) = P(A) \cdot P(B)$. Ze vztahu $P(A \cap B) = P(A) \cdot P(B)$ naopak plyne $P(A|B) = \frac{P(A\cap B)}{P(B)} = \frac{P(A)\cdot P(B)}{P(B)} = P(A)$. Proto definujeme:

[3] Pro zbytek tohoto odstavce pro jednoduchost předpokládejme, že $P(A) > 0$ a $P(B) > 0$.

Definice 3.4. Jevy A a B se nazývají *nezávislé*, někdy *statisticky* nebo *stochasticky nezávislé*, pokud

$$P(A \cap B) = P(A) \cdot P(B). \tag{10}$$

Poznámka 3.8. (a) Definovat nezávislost podmínkou $P(A) = P(A|B)$ je asi přirozenější. Výhodou podmínky (10) je, že je symetrická. Tato podmínka je vždy použitelná, na rozdíl od $P(A) = P(A|B)$, která vyžaduje $P(B) > 0$.

(b) Je-li A nemožný, tj. $P(A) = 0$, jsou A a B vždy nezávislé, protože v tom případě je $P(A) \cdot P(B) = 0$ a vzhledem k $P(A \cap B) \leq 0$, což plyne z $P(A \cap B) \leq P(A) = 0$, také $P(A \cap B) = 0$. Stejně tak, je-li nemožný jev B. Podobně platí, že Ω a B jsou vždy nezávislé jevy.

(c) Jsou-li A a B disjunktní, tj. $A \cap B = \emptyset$, pak nezávislost A a B implikuje, že $P(A) = 0$ nebo $P(B) = 0$.

(d) Platí-li, že A a A jsou nezávislé, pak musí být $P(A) = 0$, nebo $P(A) = 1$, neboť nezávislost v tomto případě znamená $P(A) = P(A)^2$.

(e) Nezávislost jevů však není tranzitivní vztah, tj. může se stát, že A a B jsou nezávislé, B a C jsou nezávislé, ale A a C nejsou nezávislé.

Věta 3.4. *Jsou-li A a B nezávislé, pak jsou nezávislé taky tyto dvojice jevů: A a $\overline{B}$, $\overline{A}$ a B, $\overline{A}$ a $\overline{B}$.*

Důkaz. Dokažme nezávislost jevů A a $\overline{B}$:

$$\begin{aligned} P(A) \cdot P(\overline{B}) &= P(A) \cdot (1 - P(B)) = P(A) - P(A) \cdot P(B) \\ &= P(A) - P(A \cap B) = P(A - A \cap B) = P(A \cap \overline{B}). \end{aligned}$$

Přitom jsme využili nezávislost jevů A a B, tj. $P(A) \cdot P(B) = P(A \cap B)$, a dále rovnost $P(A) - P(A \cap B) = P(A - A \cap B)$, která platí díky aditivitě pravděpodobnosti (zdůvodněte). □

Příklad 3.12. Z balíčku 32 mariášových karet vytáhneme náhodně jednu kartu. Jsou jevy „barva karty je zelená“ a „hodnota je 7, 8 nebo 9“ nezávislé?

Uvědomme si, že zadání je přísně vzato neúplné. Není jasné, o jaké jevy se jedná, neboť není popsána množina Ω, ani jevy, jejichž nezávislost máme ověřit. Přirozeně se však nabízí tento výklad: Ω je množina všech 32 karet, a dané jevy jsou representovány množinou A všech osmi zelených karet a množinou B všech 12 karet s uvedenými hodnotami. Předpokládáme také, že pravděpodobnosti vytažení je stejná pro každou kartu, tj. $\frac{1}{|\Omega|} = \frac{1}{32}$.

Pak $P(A) = \frac{8}{32} = \frac{1}{4}$, $P(B) = \frac{12}{32} = \frac{3}{8}$ a $P(A \cap B) = \frac{|A \cap B|}{|\Omega|} = \frac{3}{32}$. Zřejmě je $P(A \cap B) = P(A) \cdot P(B)$, tedy jevy jsou nezávislé. To je v souladu s intuicí: barva a hodnota jsou nezávislé atributy karet.

Příklad 3.13. Házíme jednou kostkou. Jsou jevy „padne sudé číslo" a „padne číslo vetší než 3" nezávislé?

Rozumí se, že $\Omega = \{1, 2, 3, 4, 5, 6\}$ a že dané jevy jsou reprezentovány množinami $A = \{2, 4, 6\}$ a $B = \{4, 5, 6\}$. Zřejmě je $P(A) = 0.5$, $P(B) = 0.5$ a $P(A \cap B) = P(\{4, 6\}) = \frac{1}{3}$. $P(A \cap B) = \frac{1}{3} \neq \frac{1}{4} = P(A) \cdot P(B)$, jevy tedy nejsou nezávislé. To je opět v souladu s intuicí: Padne-li číslo větší než 3, je pravděpodobnost, že to je číslo sudé, větší než bez této podmínky.

Definici nezávislosti lze rozšířit na n jevů takto:

Definice 3.5. Jevy $A_1, \ldots, A_n$ jsou *vzájemně nezávislé*, pokud pro každé $2 \leq k \leq n$ a libovolné indexy $i_1, \ldots, i_k \in \{1, \ldots, n\}$ platí

$$P(A_{i_1} \cap \cdots \cap A_{i_n}) = P(A_{i_1}) \cdot \cdots \cdot P(A_{i_n}). \tag{11}$$

Poznámka 3.9. (a) Definice nezávislosti dvou jevů speciálním případem definice 3.5.

(b) Platí analogie věty 3.4: Jsou-li $A_1, \ldots, A_n$ vzájemně nezávislé, pak pro libovolné $B_i \in \{A_i, \overline{A_i}\}$ jsou jevy $B_1, \ldots, B_n$ také vzájemně nezávislé.

Je zřejmé, že jsou-li jevy $A_1, \ldots, A_n$ vzájemně nezávislé, pak jsou každé dva jevy A_i a A_j, kde $i \neq j$, nezávislé dle definice 3.4. Naopak to však neplatí. Následující příklad dokonce ukazuje, žepodmínku (11) je skutečně třeba požadovat pro libovolnou podmnožinu jevů $A_1, \ldots, A_n$. Už v případě tří jevů, A, B, a C, se totiž může stát, že $P(A \cap B \cap C) = P(A)P(B)P(C)$, ale $P(A \cap B) \neq P(A)P(B)$, $P(A \cap C) \neq P(A)P(C)$ i $P(B \cap C) \neq P(B)P(C)$. Na druhou stranu může být $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$ i $P(B \cap C) = P(B)P(C)$, ale $P(A \cap B \cap C) \neq P(A)P(B)P(C)$.

Příklad 3.14. Házíme dvěma kostkami. Zvolme $\Omega = \{\langle i, j \rangle \mid 1 \leq i, j \leq 6\}$ a předpokládejme že pravděpodobnost každého $\langle i, j \rangle \in \Omega$ je $1/36$.

Uvažujme jevy A, B a C reprezentující po řadě „na první kostce padne 1, 2, nebo 3", „na první kostce padne 3, 4, nebo 5", „součet na kostkách je

9". Pak je

$$\begin{aligned} A \cap B &= \{\langle 3,1\rangle, \langle 3,2\rangle, \langle 3,3\rangle, \langle 3,4\rangle, \langle 3,5\rangle, \langle 3,6\rangle\}, \\ A \cap C &= \{\langle 3,6\rangle\}, \\ B \cap C &= \{\langle 3,6\rangle, \langle 4,5\rangle, \langle 5,4\rangle\}, \\ A \cap B \cap C &= \{\langle 3,6\rangle\}. \end{aligned}$$

Platí tedy

$$P(A \cap B \cap C) = {}^1\!/_{36} = P(A)P(B)P(C),$$

ale

$$\begin{aligned} P(A \cap B) = {}^1\!/_6 &\neq {}^1\!/_4 = {}^1\!/_2 \cdot {}^1\!/_2 = P(A)P(B), \\ P(A \cap C) = {}^1\!/_{36} &\neq {}^1\!/_{18} = {}^1\!/_2 \cdot {}^1\!/_9 = P(A)P(C), \\ P(B \cap C) = {}^1\!/_{12} &\neq {}^1\!/_{18} = {}^1\!/_2 \cdot {}^1\!/_9 = P(B)P(C). \end{aligned}$$

Uvažujme nyní jevy A, B a C reprezentující po řadě „na první kostce padne 1, 2, nebo 3", „na druhé kostce padne 4, 5, nebo 6", „součet na kostkách je 7". Pak je

$$\begin{aligned} A \cap B &= \{\langle 1,4\rangle, \langle 1,5\rangle, \langle 1,6\rangle, \langle 2,4\rangle, \langle 2,5\rangle, \langle 2,6\rangle, \langle 3,4\rangle, \langle 3,5\rangle, \langle 3,6\rangle\}, \\ A \cap C &= B \cap C = A \cap B \cap C = \{\langle 1,6\rangle, \langle 2,5\rangle, \langle 3,4\rangle\}. \end{aligned}$$

Platí tedy

$$\begin{aligned} P(A \cap B) = {}^1\!/_4 &= {}^1\!/_2 \cdot {}^1\!/_2 = P(A)P(B), \\ P(A \cap C) = {}^1\!/_{12} &= {}^1\!/_2 \cdot {}^1\!/_6 = P(A)P(C), \\ P(B \cap C) = {}^1\!/_{12} &= {}^1\!/_2 \cdot {}^1\!/_6 = P(B)P(C), \end{aligned}$$

ale

$$P(A \cap B \cap C) = {}^1\!/_{12} \neq {}^1\!/_{24} = {}^1\!/_2 \cdot {}^1\!/_2 \cdot {}^1\!/_6 = P(A)P(B)P(C).$$

Příklad 3.15. Adam a Bedřich neumí dobře počítat. Pravděpodobnost, že Adam dostane správný výsledek, je ${}^1\!/_8$, v případě Bedřicha jen ${}^1\!/_{12}$. Počítají nezávisle, tj. každý zvlášť. V případě, že oba počítají špatně, dostanou stejný výsledek jen s pravděpodobností ${}^1\!/_{1001}$. Předpokládejme, že dostali stejný výsledek. Jaká je pravděpodobnost, že je výsledek správný?

Označme A, B a C jevy popsané po řadě takto: „oba dostanou správný výsledek“, „oba dostanou stejný výsledek“, „oba dostanou stejný, ale nesprávný výsledek“. Zajímá nás podmíněná pravděpodobnost $P(A|B)$. Dle definice je třeba určit $P(A \cap B)$ a $P(B)$, ale vzhledem k tomu, že $A \cap B = A$, stačí určit $P(A)$ a $P(B)$.

Protože pracují nezávisle, je z popisu úlohy rozumné předpokládat, že $P(A) = 1/8 \cdot 1/12$. Totiž: $A = A_1 \cap A_2$, kde A_1 a A_2 jsou „Adam dostne správný výsledek“ a „Bedřich dostne správný výsledek“, a tedy z nezávislosti je $P(A) = P(A_1)P(A_2) = 1/8 \cdot 1/12$.

Dále je $B = A \cup C$ a vzhledem k $A \cap C = \emptyset$ je $P(B) = P(A) + P(C)$.

Zbývá tedy určit $P(C)$. Uvědomme si, že $1/1001$ je vlastně pravděpodobností jevu B za podmínky $\overline{A_1} \cap \overline{A_2}$, tj. $1/1001 = P(B|\overline{A_1} \cap \overline{A_2})$, a že $C = B \cap \overline{A_1} \cap \overline{A_2}$. Podle vztahu (8) je $P(B \cap (\overline{A_1} \cap \overline{A_2})) = P(\overline{A_1} \cap \overline{A_2}) \cdot P(B|\overline{A_1} \cap \overline{A_2})$. Vzhledem k nezávislosti A_1 a A_2 jsou dle věty 3.4 nezávislé také $\overline{A_1}$ a $\overline{A_2}$, a tedy $P(\overline{A_1} \cap \overline{A_2}) = P(\overline{A_1}) \cdot P(\overline{A_2})$. Je tedy

$$P(B \cap (\overline{A_1} \cap \overline{A_2})) = P(\overline{A_1} \cap \overline{A_2}) \cdot P(B|\overline{A_1} \cap \overline{A_2}) = P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(B|\overline{A_1} \cap \overline{A_2}).$$

Celkem tedy dostáváme

$$P(C) = P(B \cap (\overline{A_1} \cap \overline{A_2})) = P(\overline{A_1}) \cdot P(\overline{A_2}) \cdot P(B|\overline{A_1} \cap \overline{A_2}) = 7/8 \cdot 11/12 \cdot 1/1001.$$

Pro výslednou pravděpodobnost $P(A|B)$ tedy platí

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{P(A)}{P(A) + P(C)} \approx 0.929.$$

3.5 Bayesova věta

Bayesova věta je jednou z důležitých tvrzení teorie pravděpodobnosti, které má řadu praktických použití. Začněme ale jiným tvrzením, které je samo o sobě významné, a které v Bayesově větě použijeme.

Předpokládejme, že množina Ω výsledků (elementárních jevů) je rozdělena do jevů $B_1, \dots, B_m$, tj. že systém $\{B_1, \dots, B_m\}$ tvoří rozklad množiny Ω.[4] Systém takových jevů B_i se nazývá *úplný systém jevů*.

[4] To znamená, že B_i jsou neprázdné podmnožiny množiny Ω, jsou pod dvou disjunktní, tj. $B_i \cap B_j = \emptyset$ pro $i \neq j$, a jejich sjednocení je Ω, tj. $\bigcup_{i=1}^{m} B_i = \Omega$.

Věta 3.5 (o celkové pravděpodobnosti)**.** *Tvoří-li* $B_1, \dots, B_m$ *úplný systém jevů pravděpodobnostního prostoru* $\langle \Omega, \mathcal{A}, P \rangle$ *a platí-li pro každé* $i = 1, \dots, m$*, že* $P(B_i) > 0$*, pak pro každý jev* $A \in \mathcal{A}$ *je*

$$P(A) = \sum_{i=1}^{m} P(A|B_i) \cdot P(B_i). \tag{12}$$

Důkaz. Protože

$$A = A \cap \Omega = A \cap (B_1 \cup \cdots \cup B_m) = (A \cap B_1) \cup \cdots \cup (A \cap B_m)$$

a protože jevy $A \cap B_i$ jsou po dvou disjunktní, je

$$P(A) = P(A \cap B_1) + \cdots + P(A \cap B_m) = \sum_{i=1}^{m} P(A \cap B_i).$$

Dle (8) je $P(A \cap B_i) = P(A|B_i) \cdot P(B_i)$, dosazením tedy získáme (12). □

Poznámka 3.10. (a) Vzorec (12) je užitečný v situacích, kdy nelze pravděpodobnost $P(A)$ určit přímo, kdy ale lze situaci rozdělit na po dvou disjunktní případy, reprezentované jevy B_i, pro které umíme určit jak jejich pravděpodobnosti $P(B_i)$, tak podmíněné pravděpodobnosti $P(A|B_i)$ v těchto případech.

(b) Přechod od původní množiny Ω elementárních jevů k množině $\{B_1, \dots, B_m\}$ lze však také chápat jako proces abstrakce: Místo elementárních jevů $\omega \in \Omega$, které reprezentovaly výsledky náhodného pokusu, máme nyní jevy B_i, které reprezentují výsledky s menší rozlišovací schopností. Každý B_i totiž reprezentuje celou množinu původních výsledků. Tyto nové základní výsledky mají své pravděpodobnosti $P(B_i)$. Podmíněné pravděpodobnosti $P(A|B_i)$ lze pak chápat jako pravděpodobnost toho, že výsledek B_i je příznivý jevu A, zobecňující bivalentní vztah „elementární jev ω_i je příznivý, tj. patří do jevu A.“ Uvědomme si totiž, že pokud je Ω konečná a $P(\{\omega\})$ pro každý $\omega \in \Omega$, tvoří systém jednoprvkových množin $\{\omega\}$, $\omega \in \Omega$, úplný systém jevů splňující podmínky věty 3.5. Pak je $P(A|B_i) = P(A|\{\omega\}) = \frac{P(A \cap \{\omega\})}{P(\{\omega\})}$, a tedy platí

$$P(A|\{\omega\}) = \begin{cases} 1 & \text{pro } \omega \in A, \\ 0 & \text{pro } \omega \notin A. \end{cases}$$

Z tohoto pohledu je $P(A|B_i)$ pravděpodobnostním zobecněním vztahu „výsledek je příznivý jevu“.

(c) Pro $m = 2$ dostáváme tento speciální případ: Je-li B jev takový, že $P(B) > 0$ i $P(\overline{B}) > 0$, pak

$$P(A) = P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B}). \tag{13}$$

Příklad 3.16. Předpokládejme, že pravděpodobnost mého úspěchu u zkoušky je 0.8, pokud se budu učit, a 0.2, pokud se učit nebudu. Budu mít sice další plány, ale pravděpodobnost, že se učit budu, je vysoká, řekněme 0.85. Jaká je pravděpodobnost, že u zkoušky uspěju?

Označme A a B jevy „uspěju u zkoušky" a „budu se učit". Z popisu je známo $P(A|B) = 0.8$, $P(A|\overline{B}) = 0.1$ a $P(B) = 0.85$. Dle (13) je tedy

$$P(A) = P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B}) = 0.8 \cdot 0.85 + 0.1 \cdot 0.15 = 0.695.$$

Příklad 3.17. Softwarová firma provozuje zákaznickou linku, kde poskytuje svým třem velkým zákazníkům konzultace ohledně firemního informačního systému. Je známo, že 15% příchozích hovorů je od první firmy, 35% od druhé a 50% od třetí. Je dále známo, že pravděpodobnost, že příchozí hovor nebude schopen vyřídit operátor a bude muset dotaz předat technikovi, je pro první, druhou a třetí firmu 0.01, 0.05 a 0.02. Jaká je pravděpodobnost, že dotaz z příchozího hovoru bude muset operátor předat k vyřízení technikovi?

Označme A jev „{"dotaz bude muset předat technikovi} a B_i jevy „{"příchozí hovor je od i-té firmy}. Jevy B_i tvoří úplný systém jevů. Dle věty o celkové pravděpodobnosti je tedy

$$\begin{aligned} P(A) &= P(A|B_1) \cdot P(B_1) + P(A|B_2) \cdot P(B_2) + P(A|B_3) \cdot P(B_3) = \\ &= 0.01 \cdot 0.15 + 0.05 \cdot 0.35 + 0.02 \cdot 0.5 = 0.029. \end{aligned}$$

Věta 3.6 (Bayesova). *Nechť $B_1, \ldots, B_m$ je úplný systém jevů a A libovolný jev v pravděpodobnostním prostoru $\langle \Omega, \mathcal{A}, P \rangle$. Pokud $P(B_1) > 0, \ldots, P(B_m) > 0$ a $P(A) > 0$, pak pro každé $k = 1, \ldots, m$ je*

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{P(A)} = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^{m} P(A|B_i) \cdot P(B_i)}. \tag{14}$$

Důkaz. Dle definice a dále z (8) je

$$P(B_k|A) = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A|B_k) \cdot P(B_k)}{P(A)}.$$

Tím je dokázána první rovnost. Druhou rovnost obdržíme použitím věty o celkové pravděpodobnosti, tj. za $P(A)$ dosadíme dle (12). □

Poznámka 3.11. (a) Bayesova věta okamžitě plyne z definice podmíněné pravděpodobnosti a věty o celkové pravděpodobnosti. Má však mnoho praktických uplatnění, protože situace se má často tak, že neznámou pravděpodobnost $P(B_k|A)$ přímo určit nelze, ale pravděpodobnosti $P(A|B_i)$ a $P(B_i)$ všechny známé jsou.

(b) Tvrzení je připisováno Thomasu Bayesovi (1701–1761), který ho formuloval pro speciální případ, kdy všechny $P(B_i)$ jsou stejné. Obecný vzorec formuloval Laplace.

(c) Pro $m = 2$ dostaneme speciální případ,

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B})}. \tag{15}$$

(d) Bayesova věta je základem tzv. bayesovských metod. Pravděpodobnost $P(B_k|A)$ se nazývá a posteriorní pravděpodobnost (pravděpodobnost a posteriori). ROZVEST ZDE NEBO POZDEJI

Příklad 3.18. Vraťme se k příkladu 3.17. Jaká je praděpodobost, že příchozí hovor je od druhého zákazníka, pokud jej operátor musí předat k vyřízení technikovi?

Použijme stejné jevy jako k příkladu 3.17 a využijme toho, že jsme v příkladu 3.17 již spočítali $P(A) = 0.029$. Dle Bayesova vzorce pro hledanou pravděpodobnost $P(B_2|A)$ platí

$$P(B_2|A) = \frac{P(A|B_2) \cdot P(B_2)}{P(A)} = \frac{0.05 \cdot 0.35}{0.029} \approx 0.603.$$

Příklad 3.19. Po komunikačním kanálu jsou posílány znaky 0 a 1. Kanál je zatížen šumem a dopouští se chyb. Když je vyslána 0, je přijata 0 s pravděpodobností 0.94. Když je vyslána 1, je přijata 1 s pravděpodobností 0.91. Dále je známo, že pravděpodobnost vyslání 0 je 0.45. Je vyslán (neznámý) znak. Určete:

(a) pravděpodobnost, že bude přijata 1;

(b) pravděpodobnost, že bude přijata 0;

(c) pravděpodobnost, že byla vyslána 1, pokud bude přijata 1;

(d) pravděpodobnost, že byla vyslána 0, pokud bude přijata 0;

(e) pravděpodobnost, že při přenosu dojde k chybě.

Uvažujme následující jevy:

$$\begin{array}{lll} V_0 & \dots & \text{„vyslána 0“}, \\ V_1 & \dots & \text{„vyslána 1“}, \\ P_0 & \dots & \text{„přijata 0“}, \\ P_1 & \dots & \text{„přijata 1“}. \end{array}$$

Ze zadání je známo, že $P(P_0|V_0) = 0.94$, $P(P_1|V_1) = 0.91$ a $P(V_0) = 0.45$. Základní vlastnosti pravděpodobnostních měr a věta 3.3 okamžitě dávají: $P(P_1|V_0) = P(\overline{P_0}|V_0) = 1 - P(P_0|V_0) = 0.06$, $P(P_0|V_1) = P(\overline{P_1}|V_1) = 1 - P(P_1|V_1) = 0.09$, $P(V_1) = P(\overline{V_0}) = 1 - P(V_0) = 0.55$. To jsou tedy známé pravděpodobnosti.

Protože V_0 a V_1 tvoří úplný systém jevů, lze (a) určit z věty o celkové pravděpodobnosti:

$$P(P_1) = P(P_1|V_0)P(V_0) + P(P_1|V_1)P(V_1) = 0.06 \cdot 0.45 + 0.91 \cdot 0.55 = 0.5275.$$

(b) lze určit analogickým postupem,

$$P(P_0) = P(P_0|V_0)P(V_0) + P(P_0|V_1)P(V_1) = 0.94 \cdot 0.45 + 0.09 \cdot 0.55 = 0.4725$$

nebo s využitím (a) takto: $P(P_0) = P(\overline{P_1}) = 1 - P(P_1) = 1 - 0.5275 = 0.4725$.

(c), tj. pravděpodobnost $P(V_1|P_1)$, určíme z první rovnosti v (14):

$$P(V_1|P_1) = \frac{P(P_1|V_1)P(V_1)}{P(P_1)} = \frac{0.91 \cdot 0.55}{0.5275} \approx 0.949.$$

Podobně pro (d):

$$P(V_0|P_0) = \frac{P(P_0|V_0)P(V_0)}{P(P_0)} = \frac{0.94 \cdot 0.45}{0.4725} \approx 0.895.$$

Pravděpodobnost chyby je pravděpodobnost $P((V_1 \cap P_0) \cup (V_0 \cap P_1))$, tj. pravděpodobnost jevu „vyslána 1 a přijata 0, nebo vyslána 0 a přijata 1“. Protože $V_1 \cap P_0$ a $V_0 \cap P_1$ jsou disjunktní, dostáváme s využitím (8) pro pravděpodobnost chyby

$$\begin{aligned} P((V_1 \cap P_0) \cup (V_0 \cap P_1)) &= P(V_1 \cap P_0) + P(V_0 \cap P_1) \\ &= P(V_1|P_0)P(P_0) + P(V_0|P_1)P(P_1). \end{aligned}$$

Zbývá si uvědomit, že $P(V_1|P_0) = P(\overline{V_0}|P_0) = 1 - P(V_0|P_0) \approx 0.105$ a $P(V_0|P_1) = P(\overline{V_1}|P_1) = 1 - P(V_1|P_1) \approx 0.051$. Po dosazení dostaneme, že pravděpodobnost chyby je

$$P(V_1|P_0)P(P_0) + P(V_0|P_1)P(P_1) \approx 0.105 \cdot 0.4725 + 0.051 \cdot 0.5275 \approx 0.0765.$$

Zastavme se ještě u otázky, jakou konkrétní množinu elementárních jevů je možné pro tento příklad zvolit. Přirozeně se nabízí $\Omega = \{\langle v, p\rangle \mid v, p \in \{0, 1\}\}$, kde $\langle v, p\rangle$ reprezentuje výsledek, při kterém byl vyslán symbol v a přijat symbol p.

Podobná je následující úloha, která ve skutečnosti vznikne jen jinou interpretací daných podmínek.

Příklad 3.20. Používáme postup na testování správnosti výrobku. Přitom víme, že pravděpodobnost, že výrobek má vadu, je 0.02. Zařízení není zcela spolehlivé. Víme, že pokud je výrobek v pořádku, zařízení ho vyhodnotí jako správný s pravděpodobností 0.95; pokud není v pořádku, zařízení ho vyhodnotí jako nesprávný s pravděpodobostí 0.94. Pokud je výrobek vyhodnocen jako nesprávný, jaká je pravděpodobnost, že je skutečně nesprávný?

Označme B jev „výrobek je nesprávný“ a A jev „výrobek vyhodnocen jako nesprávný“. Pro výpočet hledané pravděpodobnosti $P(B|A)$ se nabízí použít Bayesův vzorec ve speciálním tvaru (15). Za zadání máme $P(A|B) = 0.94$, $P(B) = 0.02$, $P(A|\overline{B}) = 0.05$, protože ze zadání je $P(\overline{A}|\overline{B}) = 0.95$ a vzhledem k větě 3.3 je $P(A|\overline{B}) = 1 - P(\overline{A}|\overline{B})$, dále je $P(\overline{B}) = 0.98$. Dosazením do (15) dostaneme

$$P(B|A) = \frac{0.94 \cdot 0.02}{0.94 \cdot 0.02 + 0.05 \cdot 0.98} \approx 0.277.$$

Příklad 3.21. Máme dvě mince, jedna je normální a panna i orel na ní padají v pravděpodobností $1/2$, druhá je poškozená a padá na ní jen panna. Nejprve náhodně zvolíme minci a pak ji dvakrát hodíme. Jaká je pravděpodobnost, že jsme vybrali minci poškozenou, pokud dvakrát padla panna?

Zvolme $\Omega = \{\langle c, a, b\rangle \mid c \in \{n, p\}, a, b \in \{P, O\}\}$, přičemž $\langle c, a, b\rangle$ označuje výsledek, při kterém byla vybrána mince c (n – normální, p – poškozená), při prvním hodu padlo a a při druhé b (P – panna, O – orel). Označme A a B jevy popsané jako „dvakrát padla panna“ a „byla zvolena poškozená mince“. Pro výpočet hledané pravděpodobnosti $P(B|A)$

můžeme použít Bayesova vzorce, pokud se nám podaří rozdělit Ω na úplný systém jevů B_i, z nichž jeden je B, tak, že budeme schopni určit pravděpodobnosti $P(A|B_i)$ a $P(B_i)$. Nabízí se za takový systém vzít systém sestávající z B a $\overline{B}$. $P(B|A)$ je pak dána vzorcem (15). Snadno se zjistí, že $P(A|B) = 1$, $P(B) = 1/2$, $P(A|\overline{B}) = 1/4$, $P(\overline{B}) = 1/2$. Dostáváme tedy

$$P(B|A) = \frac{P(A|B) \cdot P(B)}{P(A|B) \cdot P(B) + P(A|\overline{B}) \cdot P(\overline{B})} = \frac{1 \cdot 1/2}{1 \cdot 1/2 + 1/4 \cdot 1/2} = \frac{4}{5}.$$

Příklad 3.22. Uveďme ještě příklad, který prezentovali již zmínění psychologivé Amos Tversky a Daniel Kahneman. Ve městě jezdí 85% zelených taxíků a 15% modrých. Podle svědka nehody nehodu zavinil modrý taxík, který pak ujel. Testy ukázaly, že tento svědek za podobných podmínek správně určí barvu taxíku v 80% případů a v 20% ji určí nesprávně. Jaká je pravděpodobnost, že nehodu skutečně zavinil modrý taxík?

Zaveďme tyto jevy:

M ... „taxík byl modrý“,
Z ... „taxík byl zelený“,
S ... „svědek říká, že taxík byl modrý“.

Za zadání víme, že $P(M) = 0.15$, $P(Z) = 0.85$, $P(S|M) = 0.8$, $P(S|Z) = 0.2$.

Hledáme $P(M|S)$. Použitím Bayesovy věty dostaneme

$$\begin{aligned} P(M|S) &= \frac{P(S|M)P(M)}{P(S)} = \frac{P(S|M)P(M)}{P(S|M)P(M) + P(S|Z)P(Z)} = \\ &\frac{0.8 \cdot 0.15}{0.8 \cdot 0.15 + 0.2 \cdot 0.85} \approx 0.414. \end{aligned}$$

Je tedy

$$P(Z|S) = P(\overline{M}|S) = 1 - P(M|S) \approx 0.586.$$

Docházíme tedy k zajímavému závěru, totiž přestože svědek tvrdí, že taxík byl modrý, je pravděpodobnější, že byl zelený.

Co by se stalo, kdyby svědek byl spolehlivější? Předpokládejme tedy, že je spolehlivější v tom smyslu, že $P(S|M) = 0.9$, $P(S|Z) = 0.1$. Pak analogický výpočet dá $P(M|S) \approx 0.614$ a $P(Z|S) \approx 0.386$, tedy pravděpodobnější je pak to, co tvrdí svědek.

Předpokládejme nyní, že svědek má původně danou spolehlivost, ale že k němu přibude druhý, stejně spolehlivý svědek. Označme nyní původní jev S symbolem S_1 a označme S_2 nový jev „druhý svědek říká, že taxík byl modrý". Zajímá nás pravděpodobnost, že taxík byl skutečně modrý, pokud oba svědkové tvrdí, že byl modrý, tj. pravděpodobnost $P(M|S_1 \cap S_2)$. Dle Bayesova vzorce je

$$\begin{aligned} P(M|S_1 \cap S_2) &= \frac{P(S_1 \cap S_2|M)P(M)}{P(S_1 \cap S_2)} = \\ &= \frac{P(S_1 \cap S_2|M)P(M)}{P(S_1 \cap S_2|M)P(M) + P(S_1 \cap S_2|Z)P(Z)}. \end{aligned}$$

Z popisu úlohy je rozumné předpokládat, že jevy S_1 a S_2 jsou tzv. podmíněně nezávislé za podmínky M a stejně tak za podmínky Z, tedy že platí $P(S_1 \cap S_2|M) = P(S_1|M) \cdot P(S_2|M) = 0.8 \cdot 0.8 = 0.64$ a $P(S_1 \cap S_2|Z) = P(S_1|Z) \cdot P(S_2|Z) = 0.2 \cdot 0.2 = 0.04$. Dosazením dostaneme

$$P(M|S_1 \cap S_2) = \cdots = \frac{0.64 \cdot 0.15}{0.64 \cdot 0.15 + 0.04 \cdot 0.85} \approx 0.738,$$

a tedy $P(Z|S_1 \cap S_2) = P(\overline{M}|S_1 \cap S_2) = 1 - P(M|S_1 \cap S_2) \approx 0.262$. Svědectví druhého svědka tedy změnilo situaci. Nyní je pravděpodobnější, dokonce výrazně, že taxík byl modrý. Zahrnování dalších informací, jako v tomto případě nového svědka, a aktualizace vytvořeného modelu je pro bayesovské modely typická.

POZDEJI POKRACOVANI

Příklad 3.23. lékařská diagnostika, později

Příklad 3.24 (filtr nevyžádané pošty – spam filter)**.** DOPLNIT

4 Náhodné veličiny

4.1 Pojem náhodné veličiny

Elementární jevy, tj. výsledky náhodného pokusu, jsme definovali jako prvky ω množiny Ω. Tyto prvky mohou, ale nemusí mít číselnou povahu. Často je ale užitečné charakterizovat výsledky pokusů číselně. Jsou-li možné výsledky lidé (výsledek = náhodně vybraný člověk), mohou nás zajímat charakteristiky jako například věk, hmotnost, počet dětí a podobně.

Další příklady: cena vybraného výrobku; doba bezporuchového chodu náhodně vybrané součásti; doba (počet kroků), kterou algoritmus potřebuje pro zpracování náhodně vybraného vstupu.

Elementárním jevům v těchto případech tedy přiřazujeme čísla. Takováto přiřazení, tj. zobrazení, množiny elementárních jevů do množiny reálných čísel, která splňují jisté podmínky, se nazývají náhodné veličiny (někdy také náhodné proměnné, místo „náhodné" se také používá „stochastická", angl. random nebo stochastic variable nebo quantity).

Význam takových přiřazení lze spatřovat v následujícím. Za prvé, s číselnými charakteristikami je často žádoucí provádět operace, které nám poskytnou zajímavé informace. Například se můžeme ptát, jaká je očekávaná hmotnost náhodně vybraného člověka. Za druhé, rozlišovat elementární jevy může být zbytečně detailní, například u posloupnosti pokusů nás nemusí vždy zajímat, které pokusy byly úspěšné, ale může nás zajímat jen počet úspěšných výsledků. Tedy: uvažujeme-li jen číslené charakteristiky, nerozlišujeme elementární jevy se stejnými číselnými charakteristikami, a tedy v tomto smyslu odhlížíme od nepodstatného detailu.

Nyní zavedeme obecný pojem náhodné veličiny.

Definice 4.1. Nechť $\mathcal{A}$ je σ-algebra jevů na množině Ω elementárních jevů, $\mathcal{B}$ nechť je borelovská σ-algebra na $\mathbb{R}$. Funkce $X : \Omega \to \mathbb{R}$ se nazývá *náhodná veličina*, pokud pro každou množinu $B \in \mathcal{B}$ platí $X^{-1}(B) \in \mathcal{A}$, kde

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}. \tag{16}$$

Poznámka 4.1. (a) $X^{-1}(B)$ je vzor množiny B, tj. množina všech elementárních jevů, pro které je jejich hodnota prvkem množiny B.

(b) Podmínka kladená v definici na zobrazení X tedy, volně řečeno, je: Vezměme libovolnou rozumnou podmnožinu B množiny $\mathbb{R}$ (tj. libovolnou $B \in \mathcal{B}$). Pak množina výsledků ω, jejichž hodnota leží v B (tj. množina $X^{-1}(B)$), je pozorovatelným jevem (tj. patří do $\mathcal{A}$). Množina $\mathcal{B}$ se obvykle nezmiňuje, často jen řekneme, že je dána σ-algebra $\mathcal{A}$ na Ω a náhodná veličina X, popř. že je dán pravděpodobnostní prostor $\langle \Omega, \mathcal{A}, P \rangle$ a náhodná veličina X.

(c) V teorii míry, kde se $\langle \Omega, \mathcal{A} \rangle$ a $\langle \mathbb{R}, \mathcal{B} \rangle$ nazývají měřitelné prostory, se funkce splňující podmínku kladenou na náhodnou veličinu nazývá měřitelná funkce.

(d) Podmínku (16) lze ekvivalentně vyjádřit takto: pro každé $x \in \mathbb{R}$ platí, že množina $B^{-1}(-\infty, x)$, tj. množina $\{\omega \in \Omega \mid X(\omega) \in (-\infty, x)\}$, patří do $\mathcal{A}$. Tedy vzory intervalů $(-\infty, x)$ patří do $\mathcal{A}$. Ekvivalentní je požadovat to samé pro vzory intervalů $(-\infty, x]$, to samé pro intervaly (x, ∞) nebo $[x, \infty)$. Tyto a další poznatky lze najít v teorii míry.

(e) Budeme rozlišovat dva speciální typy náhodných veličin, tzv. diskrétní a spojité náhodné veličiny. Ty budeme v dalším pojednávat samostatně. To má výhody (samotný pojem diskrétní náhodné veličiny a jeho studium je jednodušší než v obecném případě) i nevýhody (musíme postupovat dvakrát, jednou pro diskrétní a jednou pro spojitý případ). Obecný pojem uvádíme, aby čtenář viděl, že to tak lze.

Definice 4.2. Nechť $\langle \Omega, \mathcal{A}, P \rangle$ je praděpodobnostní prostor a $X : \Omega \to \mathbb{R}$ náhodná veličina. Funkce $F : \mathbb{R} \to [0, 1]$ definovaná

$$F(x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) \tag{17}$$

se nazývá *distribuční funkce* náhodné veličiny X. Někdy se bere $F(x) = P(\{\omega \in \Omega \mid X(\omega) < x\})$.

4.2 Diskrétní náhodné veličiny

4.2.1 Pravděpodobnostní a distribuční funkce diskrétní náhodné veličiny

Definice 4.3. Náhodná veličina se nazývá *diskrétní*, pokud nabývá jen konečně nebo spočetně mnoha hodnot.

Někdy se také říká, že X má rozdělení diskrétního typu.

Tedy náhodná veličina X je diskrétní, pokud je množina $X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$ konečná nebo spočetná. Všechny hodnoty veličiny X tedy tvoří konečnou nebo spočetnou posloupnost $x_1, x_2, \ldots$ Pokud použijeme $x_1, x_2, \ldots,$ budeme tím myslet právě tuto posloupnost.

Všimněme si, že systém množin $A_x = \{\omega \in \Omega \mid X(\omega) = x\}$, kde $x \in X(\Omega)$, tj. x označuje libovolnou hodnotu, které X může nabývat, tvoří úplný systém jevů. Můžeme uvažovat příslušné pravděpodobnosti $P(A_x)$, tj. pravděpodobnosti, že veličina X nabývá hodnoty x. Ty se také značí $p_X(x)$ nebo $P(X = x)$.

Definice 4.4. Funkce přiřazující hodnotám $x \in \mathbb{R}$ pravděpodobnosti $P(A_x)$ se značí p_X a nazývá se *pravěpodobnostní funkce náhodné veličiny* X (angl.

probability mass function (pmf), discrete density function). Pro $x \in \mathbb{R}$ tedy je

$$p_X(x) = P(X = x) = P(\{\omega \in \Omega \mid X(\omega) = x\}).$$

Poznámka 4.2. (a) V tom případě je distribuční funkce F veličiny X skokovitá a skok v bodě x_i má velikost $p_X(x_i)$. VIZ PREDNASKA.

(b) Základní vlastnosti p_X (snadno se vidí):

– pro každé $x \in \mathbb{R}$ je $0 \leq p_X(x) \leq 1$;

– množina $\{x \in \mathbb{R} \mid p_X(x) \neq 0\}$ je z definice diskrétní náhodné veličiny konečná nebo spočetná, její prvky tedy lze uspořádat do posloupnosti $x_1, x_2, \ldots$, pro kterou pak platí

$$\textstyle\sum_i p_X(x_i) = 1.$$

Poznámka 4.3. Ve výše použitém značení $P(X = x)$ lze výraz „$X = x$“ chápat jako označení jevu $\{\omega \in \Omega \mid X(\omega) = x\}$. Tento způsob označování jevů se často používá. Například $P(X \leq x)$ označuje pravděpodobnost jevu $\{\omega \in \Omega \mid X(\omega) \leq x\}$, $P(0 \leq X < x)$ pravděpodobnost jevu $\{\omega \in \Omega \mid 0 \leq X(\omega) < x\}$, pro jev $A \in \mathcal{A}$ je $P(X \in A)$ pravděpodobnost jevu $\{\omega \in \Omega \mid X(\omega) \in A\}$ a podobně.

Příklad 4.1. Uvažujme náhodný pokus, jehož realizace spočívá v tom, že třikrát provedeme experiment. Jedntotlivá provedení experimentu jsou nezávislá, experiment může skončit úspěchem s pravděpodobnosti p a neúspěchem s pravděpodobností $q = 1 - p$. Předpokládejme pro jednoduchost, že $p = 0.5$. V příkladu **??** uvidíme obecnější situaci. Je přirozené vzít $\Omega = \{\langle 0,0,0\rangle, \langle 0,0,1\rangle, \langle 0,1,0\rangle, \ldots, \langle 1,1,1\rangle\}$, $\mathcal{A} = 2^\Omega$ a P určenou vztahem $P(\{\omega\}) = {}^1/_{|\Omega|} = {}^1/_8$.

Uvažujme zobrazení $X : \Omega \to \mathbb{R}$ definované $X(\langle a,b,c\rangle) = a + b + c$ (počet úspěchů v sérii tří pokusů). Pak X je náhodná veličina. Je např. $X(\langle 0,0,0\rangle) = 0$, $X(\langle 0,1,0\rangle) = 1$, $X(\langle 1,0,1\rangle) = 2$ atd. Možné hodnoty X jsou 0, 1, 2 a 3. Pravděpodobnostní funkce je dána následovně:

$$p_X(0) = q^3 = 0.125, p_X(1) = 3 \cdot p \cdot q^2 = 0.375,$$
$$p_X(2) = 3 \cdot p^2 \cdot q = 0.375, p_X(3) = p^3 = 0.125.$$

Například jsou 3 elementární jevy, které mají 1 výskyt úspěchu, každý z nich má pravděpodobnost $p \cdot q^2 = 0.125$, celkem tedy $p_X(1) = 0.125 + 0.125 + 0.125 = 0.375$. $\Diamond$

Příklad 4.2. Uvažujme náhodný pokus a množinu Ω z příkladu 3.21. Definujme $X : \Omega \to \mathbb{R}$ takto: $X(\langle c, a, b\rangle)$ je počet padnutí panny při $\langle c, a, b\rangle$, tj. např. $X(\langle p, P, P\rangle) = 2$, $X(\langle n, P, O\rangle) = 1$ apod. X je náhodná veličina a je $X(\Omega) = \{0, 1, 2\}$, tj. možné hodnoty X jsou 0, 1 a 2. Popiště p_X. $\Diamond$

Definice 4.5. Funkce přiřazující hodnotám $x \in \mathbb{R}$ hodnoty $P(X \leq x)$ se značí F_X, popř. jen F, a nazývá se *distribuční funkce* náhodné veličiny X (angl. distribution function, cumulative distribution function (cdf)). Pro $x \in \mathbb{R}$ je tedy

$$F_X(x) = P(X \leq x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}).$$

Je tedy $F_X(x) = \sum_{x_i \leq x} P(X = x_i)$.

Poznámka 4.4. (a) Protože $(-\infty, a] \cup (a, b] = (-\infty, b]$, je $P(X \in (-\infty, b]) = P(X \in (-\infty, a] \cup (a, b]) = P(X \in (-\infty, a]) + P(X \in (a, b])$, a tedy

$$P(X \in (a, b]) = P(X \in (-\infty, b]) - P(X \in (-\infty, a]) = F_X(b) - F_X(a).$$

(b) Základní vlastnosti F_X (snadno se vidí):

- pro každé $x \in \mathbb{R}$ je $0 \leq F_X(x) \leq 1$;
- $x \leq y$ implikuje $F_X(x) \leq F_X(y)$, tj. F_X je neklesající;
- $\lim_{x\to\infty} F_X(x) = 1$ a $\lim_{x\to-\infty} F_X(x) = 0$. Pokud X nabývá jen konečně mnoha hodnot, pak existují x_0 a x_1 tak, že $F_X(x) = 0$ pro všechna $x \leq x_0$ a $F_X(x) = 1$ pro všechna $x \geq x_1$.
- F_X je zprava spojitá a má nejvýše spočetně mnoho bodů nespojitosti.
- Jsou-li $x_1 < x_2 < \cdots$ všechny hodnoty veličiny X, pak F_X má v každém bodě x_i skok velikosti $p_X(x_i)$ a je konstantní na intervalu $[x_i, x_{i+1})$ pro každé i.

Lze ukázat, že F_X je distribuční funkcí nějaké diskrétní náhodné veličiny, právě když má výše uvedené vlastnosti.

Poznámka 4.5. Je-li dán pravděpodobnostní prostor $\langle \Omega, \mathcal{A}, P\rangle$ a náhodná veličina $X : \Omega \to \mathbb{R}$, můžeme uvažovat dva druhy pravděpodobností. Za prvé, pravděpodobnosti jevů $A \in \mathcal{A}$ v původním prostoru. Ty jsou dány funkcí P pravděpodobnostního prostoru $\langle \Omega, \mathcal{A}, P\rangle$. Za druhé, pravděpodobnosti jevů popsaných například slovy „hodnota X je 60" nebo „hodnota X je v intervalu $[60, 80]$", tedy pravděpodobnosti určitých hodnot náhodné veličiny. Ty jsou dány distribuční funkcí F_X a v případě diskrétní náhodné veličiny také pravděpodobnostní funkcí p_X. Uvědomme si ale, že určování pravděpodobnosti hodnot nebo množin hodnot náhodné veličiny pomocí F_X nebo p_X vždy vede na určování pravděpodobnosti jevů pomocí P, tedy na určování pravděpodobnosti v původním pravděpodobnostním prostoru, protože F_X a p_X jsou definovány pomocí P. Zopakujme základní princip (kvůli němu je v definici náhodné veličiny ona podmínka): chce určit pravděpodobnost $P_X(B)$ borelovské množiny B hodnot veličiny X (borelovské proto, že borelovské množiny považujeme za rozumné); určíme množinu $X^{-1}(B)$ těch elementárních jevů ω, jimž X přiřazuje hodnotu, která patří do B, tj. $X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\}$; v původním prostoru $\langle \Omega, \mathcal{A}, P \rangle$ určíme pravděpodobnost $P(X^{-1}(B))$ této množiny (to lze, neboť z definice náhodné veličiny je $X^{-1}(B) \in \mathcal{A}$); prohlásíme, že $P(X^{-1}(B))$ je hledaná pravděpodobnost, tj. $P_X(B) = P(X^{-1}(B))$.

4.2.2 Charakteristiky diskrétních náhodných veličin

Pravděpodobnostní aspekty náhodné veličiny X popisuje její distribuční funkce $F(x)$; v případě diskrétních veličin pak také pravděpodobnostní funkce p_X (v případě spojité funkce pak hustota pravděpodobnosti $f(x)$). V mnoha situacích nepotřebujeme o náhodné veličině znát tuto úplnou informaci, stačí její popis pomocí jistých výstižných hodnot, kterým říkáme *charakteristiky náhodných veličin*. Ty základní teď uvedeme.

Budeme předpokládat, že X je diskrétní náhodná veličina, která nabývá hodnot $x_1, x_2, \ldots$

Střední hodnota

Střední hodnota diskrétní náhodné veličiny X se značí $E(X)$ a je definována vzorcem

$$E(X) = \sum_i x_i \cdot P(X = x_i). \tag{18}$$

Střední hodnota náhodné veličiny vyjadřuje, s přihlédnutím k rozdělení pravděpodobnosti, očekávanou hodnotu výsledku náhodného pokusu. Je to vážený součet hodnot náhodné veličiny, váhami jsou pravděpodobnosti těchto hodnot. Střední hodnota $E(X)$ nemusí být rovna žádné z hodnot, které X nabývá (stejně jako průměr několika čísel nemusí být roven ani jednomu z nich).

Příklad 4.3. Uvažujme $\Omega = \{\omega_1, \dots, \omega_4\}$ a rovnoměrně rozdělení pravděpodobnosti, tj. $P(\{\omega_i\}) = 1/4$ pro každé i. Náhodná veličina nechť je dána takto: $X(\omega_1) = 1$, $X(\omega_2) = 10$, $X(\omega_3) = 1$, $X(\omega_4) = 12$. Náhodná veličina má tři možné hodnoty, $x_1 = 1$, $x_2 = 10$ a $x_3 = 12$. Je $p_X(1) = P(X = 1) = P(\{\omega_1, \omega_3\}) = 1/2$, $p_X(10) = P(X = 10) = P(\{\omega_2\}) = 1/4$ a $p_X(12) = P(X = 12) = P(\{\omega_4\}) = 1/4$. Tedy

$$
\begin{aligned}
E(X) &= x_1 \cdot P(X = x_1) + x_2 \cdot P(X = x_2) + x_3 \cdot P(X = x_3) = \\
&= 1 \cdot P(X = 1) + 10 \cdot P(X = 10) + 12 \cdot P(X = 12) = \\
&= 1 \cdot 1/2 + 10 \cdot 1/4 + 12 \cdot 1/4 = 6.
\end{aligned}
$$

$\diamondsuit$

Příklad 4.4. Uvažujme náhodnou veličinu X z příkladu 4.2. Určete $E(X)$. $\diamondsuit$

Příklad 4.5. Nechť $\Omega = \{\omega_1, \dots, \omega_5\}$ je množina pěti žáků školního kroužku, nechť $P(\{\omega_i\}) = 1/5$ pro každý $\omega_i \in \Omega$. Uvažujme následující náhodnou veličinu, která žákům přiřazuje jejich hmotnosti v kg: $X(\omega_1) = 20$, $X(\omega_1) = 25$, $X(\omega_1) = 22$, $X(\omega_1) = 20$, $X(\omega_1) = 18$. Pak $E(X) = 21$ (ověřte).

Zadání lze upravit tak, že Ω je množina všech žáků 1. tříd všech základních škol v Olomouci, množina všech obyvatel České republiky a podobně. Střední hodnota pak vyjadřuje očekávanou hmotnost náhodně vybraného žáka 1. třídy v Olomouci, náhodně vybraného obyvatele České republiky a podobně.

Všimněte si, že je-li pravděpodobnost všech elementárních jevů ω stejná, pak je-li X prostá (tj. $\omega \neq \omega'$ implikuje $X(\omega) \neq X(\omega')$), je $E(X)$ rovna průměru všech hodnot, kterých X nabývá. $\diamondsuit$

Snadno lze dokázat, že jsou-li X a Y náhodné veličiny (na stejném pravděpodobnostním prostoru) a $c \in \mathbb{R}$, pak i funkce $X + Y$ a $c \cdot X$ jsou náhodné veličiny; tyto funkce jsou přitom definovány následovně: $[X + Y](\omega) = X(\omega) + Y(\omega)$. $[c \cdot X](\omega) = c \cdot X(\omega)$.

Věta 4.1 (linearita střední hodnoty)**.** *Pro náhodné veličiny X a Y a konstanty $a, b, c \in \mathbb{R}$ platí: $E(X + Y) = E(X) + E(Y)$; $E(cX) = cE(X)$; tedy $E(aX + bY) = aE(X) + bE(Y)$; dále pak $E(X + c) = E(X) + c$; tedy $E(aX + b) = aE(X) + b$.*

Důkaz. Snadno přímo z definice. $\square$

Momenty

S hodnotami náhodných veličin je často užitečné provádět úpravy. Je-li úprava realizována funkcí $h : \mathbb{R} \to \mathbb{R}$, máme pak často co do činění s funkcí $X \circ h : \Omega \to \mathbb{R}$, označovanou také $h(X)$, tj. funkcí přiřazující každému $\omega \in \Omega$ hodnotu $h(X(\omega))$.

Uvedeme dvě důležitá pomocná tvrzení.

Lemma 4.1. *Je-li $X : \Omega \to \mathbb{R}$ (diskrétní nebo spojitá) náhodná veličina a $h : \mathbb{R} \to \mathbb{R}$ spojitá funkce (obecněji: borelovsky měřitelná funkce[5]), pak funkce $h(X)$ je náhodná veličina, která je diskrétní, pokud X je diskrétní, a spojitá, pokud X je spojitá.*

Důkaz. Viz přednášky. □

Lemma 4.2. *Je-li $X : \Omega \to \mathbb{R}$ diskrétní náhodná veličina a $h : \mathbb{R} \to \mathbb{R}$ spojitá funkce (obecněji: borelovsky měřitelná funkce), pak pro náhodnou veličinu $Y = h(X)$ platí*

$$E(Y) = \sum_i h(x_i) \cdot P(X = x_i). \tag{19}$$

Důkaz. Označme $y_1, \dots, y_j, \dots$ posloupnost hodnot náhodné veličiny Y. Je zřejmé, že platí $\{\omega \in \Omega \mid Y(\omega) = y_j\} = \bigcup_{i,h(x_i)=y_j} \{\omega \in \Omega \mid X(\omega) = x_i\}$ a že uvedené množiny $\{\omega \in \Omega \mid X(\omega) = x_i\}$ jsou po dvou disjunktní. Je tedy $P(\{\omega \in \Omega \mid Y(\omega) = y_j\}) = P(\bigcup_{i,h(x_i)=y_j} \{\omega \in \Omega \mid X(\omega) = x_i\}) = \sum_{i,h(x_i)=y_j} P(\{\omega \in \Omega \mid X(\omega) = x_i\})$. Platí tedy

$$\begin{aligned}
E(Y) &= \sum_j y_j \cdot P(Y = y_j) = \sum_j y_j \cdot P(\{\omega \in \Omega \mid Y(\omega) = y_j\}) = \\
&= \sum_j y_j \cdot P(\bigcup_{i,h(x_i)=y_j} \{\omega \in \Omega \mid X(\omega) = x_i\}) = \\
&= \sum_j \sum_{i,h(x_i)=y_j} h(x_i) \cdot P(\{\omega \in \Omega \mid X(\omega) = x_i\}) = \\
&= \sum_i h(x_i) \cdot P(\{\omega \in \Omega \mid X(\omega) = x_i\}) = \sum_i h(x_i) \cdot P(X(\omega) = x_i).
\end{aligned}$$

□

[5] Tj. pro každé $a \in \mathbb{R}$ je množina $\{x \in \mathbb{R} \mid h(x) < a\}$ borelovská.

Důležité charakteristiky náhodných veličin jsou obecný a centrální moment. Nechť $r = 1, 2, \dots$ Pak r. *obecný moment* $\mu'_r(X)$ náhodné veličiny X je střední hodnota náhodné veličiny $h(X) = X^r$, tedy

$$\mu'_r(X) = E(X^r).$$

r. *centrální moment* $\mu_r(X)$ náhodné veličiny X je střední hodnota náhodné veličiny $h(X) = (X - E(X))^r$, tedy

$$\mu_r(X) = E((X - E(X))^r).$$

Uvědomme si, že ve výše uvedeném vzorci je $E(X)$ číslo. Označíme-li ho μ, pak je $\mu_r(X) = E(h(X))$, kde $h(X) = (X - \mu)^r$.

Snadno se ověří, že platí:

$$\begin{aligned} \mu_0(X) &= 1, \\ \mu_1(X) &= 0, \\ \mu_2(X) &= {\mu'}_2(X) - E(X)^2. \end{aligned}$$

Obecně pak platí POZDEJI PRO ZAJIMAVOST DOPLNIT

Rozptyl (také *variance*) $\mathrm{var}\, X$ náhodné veličiny X je druhý centrální moment X, tj.

$$\mathrm{var}\, X = \mu_2(X) = E((X - E(X))^2) = {\mu'}_2(X) - E(X)^2.$$

Směrodatná odchylka σ náhodné veličiny X je druhá odmocnina rozptylu, tj.

$$\sigma = \sqrt{\mathrm{var}\, X}.$$

Platí

$$\mathrm{var}(aX + b) = a^2 \mathrm{var}(X),$$

neboť

$$\begin{aligned} \mathrm{var}(aX + b) &= E([aX + b - E(aX + b)]^2) = E([aX + b - aE(X) - b]^2) = \\ &= E([aX - aE(X)]^2) = E(a^2[X - E(X)]^2) = \\ &= a^2 E([X - E(X)]^2) = a^2 \mathrm{var}(X). \end{aligned}$$

Rozptyl veličiny $aX + b$ tedy nezávisí na konstantě b; tedy je $\mathrm{var}(b) = 0$.

Zhruba řečeno, střední hodnota X je hodnota, kolem které veličina X kolísá; rozptyl vyjadřuje velikost tohoto kolísání. Ukazuje to následující příklad.

Příklad 4.6. (a) Uvažujme $\Omega = \{\omega_1, \ldots, \omega_4\}$, $P(\{\omega_i\}) = 1/4$, a náhodnou veličinu danou $X(\omega_1) = 0$, $X(\omega_2) = 2$, $X(\omega_3) = 4$, $X(\omega_4) = 6$. Je zřejmé, že $p_X(0) = 1/4$, $p_X(2) = 1/4$, $p_X(4) = 1/4$, $p_X(6) = 1/4$, a tedy

$$E(X) = 0 \cdot 1/4 + 2 \cdot 1/4 + 4 \cdot 1/4 + 6 \cdot 1/4 = 3.$$

Označíme-li $x_1 = 0, \ldots, x_4 = 6$, pak

$$\begin{aligned} \operatorname{var} X &= \sum_{i=1}^{4} (x_i - 3)^2 \cdot 1/4 = \\ &= [(0-3)^2 + (2-3)^2 + (4-3)^2 + (6-3)^2] \cdot 1/4 = \\ &= [9+1+1+9] \cdot 1/4 = 5. \end{aligned}$$

(b) Uvažujme veličinu Y danou $Y(\omega_1) = 1$, $Y(\omega_2) = 2$, $Y(\omega_3) = 4$, $Y(\omega_4) = 5$. Pak $p_Y(1) = 1/4$, $p_Y(2) = 1/4$, $p_Y(4) = 1/4$, $p_Y(5) = 1/4$. Pak $E(Y) = 3$. Označíme-li $y_1 = 1$, $y_2 = 2$, $y_3 = 4$, $y_4 = 5$, pak $\operatorname{var} Y = \sum_{i=1}^{4} (y_i - 3)^2 \cdot 1/4 = [(1-3)^2 + (2-3)^2 + (4-3)^2 + (5-3)^2] \cdot 1/4 = [4+1+1+4] \cdot 1/4 = 2.5$. $\Diamond$

Kvantily

Pro $q \in (0,1)$ je $100q\%$-*kvantil*, někdy q-*kvantil*, náhodné veličiny X je hodnota x_q, pro kterou je

$$F_X(x_q) \leq q \text{ a } \lim_{x \to x_q^+} F_X(x) \geq q.$$

Jak uvidíme dále, kvantily nejsou určeny jednoznačně (q-kvantilem může být interval hodnot).

Speciální názvy mají tyto kvantily:

medián … $x_{0.5}$ … 0.5-kvantil,

dolní kvartil … $x_{0.25}$ … 0.25-kvantil,

horní kvartil … $x_{0.75}$ … 0.75-kvantil,

k. *decil* … $x_{k/10}$ … $k/10$-kvantil,

k. *percentil* … $x_{k/100}$ … $k/100$-kvantil.

Modus náhodné veličiny je hodnota $\hat{x}$ taková, že pro každé $x \in \mathbb{R}$ je

$$P(X = \hat{x}) \geq P(X = x).$$

Je to tedy, zhruba řečeno, nejčastější hodnota náhodné veličiny.

Příklad 4.7. (a) Uvažujme příklad 4.6. Určete medián, dolní a horní kvartily, modus náhodné veličiny X.

(b) Uvažujme $\Omega = \{\omega_1, \ldots, \omega_4\}$, $P(\{\omega_i\}) = 1/4$, a náhodnou veličinu danou $X(\omega_1) = 2$, $X(\omega_2) = 2$, $X(\omega_3) = 4$, $X(\omega_4) = 6$. Pak $p_X(2) = 1/2$, $p_X(4) = 1/4$, $p_X(6) = 1/4$. Mediánem je každá hodnota z intervalu $[2, 4)$. Dolní kvartil neexistuje. Horní kvartil je každá hodnota z $[4, 6)$. Modus je jediný, je jím hodnota 2.

(c) Hod kostkou, medián, dolní a horní kvartily, modus příslušné náhodné veličiny. $\diamondsuit$

Charakteristiky polohy, variability, šikmosti a špičatosti

Charakteristika je funkce, která náhodným veličinám přiřazuje čísla. Číslo $\chi(X)$ přiřazené charakteristikou χ veličině X tuto charakteristiku v jistém smyslu charakterizuje. Charakteristiky jsou například momenty, které jsme zavedli výše. Určité charakteristiky, resp. jejich skupiny, jsou navrženy pro popis podobných typů vlastností. Takové charakteristiky mají společné vlastnosti. Dvě z nich teď uvedeme.

Charakteristika χ se nazývá *charakteristikou polohy*, pokud pro každou náhodnou veličinu X a konstantu c platí

$$\chi(X + c) = \chi(X) + c.$$

Charakteristikou polohy je tedy například střední hodnota, neboť $E(X + c) = E(X) + c$.

Za charakteristiky lze považovat i kvantily, pokud jsou dány jednoznačně (tak to je v jistém případě spojitých náhodných veličin, viz dále).

Charakteristika χ se nazývá *charakteristikou variability*, pokud pro každou náhodnou veličinu X a konstantu c platí

$$\chi(aX + b) = a^2\chi(X) \quad \text{nebo} \quad \chi(aX + b) = |a|\chi(X).$$

Charakteristikou polohy je tedy například rozptyl, neboť $\mathrm{var}(aX + b) = a^2\mathrm{var}(X)$, nebo směrodatná odchylka, neboť $\sqrt{\mathrm{var}(aX + b)} = |a|\sqrt{\mathrm{var}(X)}$.

V příkladu 4.7 vidíme, že pravděpodobnostní funkce (a distribuční funkce) je v případě (b) vychýlená vlevo (neboli protáhlá vpravo), zatímco v případě (a) je symetrická. Vlastnost, na kterou jsme právě narazili, popisují *charakteristiky šikmosti*, které mají splňovat následující vlastnosti (necháváme je zatím intuitivně chápané): pro veličinu X se symetrickým rozdělením pravděpodobnosti je $\chi(X) = 0$; pokud je rozdělení protáhlé vpravo (jako v příkladu 4.7 (b)), je $\chi(X) > 0$; pokud je rozdělení protáhlé vlevo, je $\chi(X) < 0$; pro $a \neq 0$ je $\chi(X + b) = a\chi(X)$.

Často používanou charakteristikou šikmosti je *koeficient šikmosti* definovaný vzrocem

$$\alpha_3(X) = \frac{\mu_3(X)}{\operatorname{var}(X)^{3/2}}.$$

Kromě šikmosti se charakterizuje také špičatost (plochost) rozdělení náhodné veličiny. K tomu se často používá *koeficient špičatosti* definovaný vzorcem

$$\alpha_4(X) = \frac{\mu_4(X)}{\operatorname{var}(X)^2} - 3.$$

Poznamenejme, že diskrétní náhodná veličina X je *symetrická podle bodu* μ (má *rozdělení symetrické podle bodu* μ), jestliže pro každý $x \in \mathbb{R}$ platí

$$p_X(x) = p_X(2\mu - x),$$

tj. $P(X = x) = P(X = 2\mu - x)$.

4.2.3 Vybrané diskrétní náhodné veličiny

Binomické rozdělení

Binomické rozdělení mají diskrétní náhodné veličiny, které přirozeně vznikají při opakovaném, nezávislém provádění pokusu se dvěma možnými výsledky.

Historická poznámka 4.1. Jacob Bernoulli (1655–1705)

Bernoulliho pokus (angl. Bernoulli trial): Pokus dvěma výsledky, úspěch s pravděpodobností p, neúspěch s pravděpodobností q (tedy $p + q = 1$, a tedy $q = 1 - p$, parametr je jeden).

Další interpretace úspěchu a neúspěchu: zásah a minutí, bezchybný a vadný, provede se větev za if a provede se větev za else, požadavek může být vyřízen hned a požadavek je přesunut do fronty, apod.

Bernoulliho schéma (posloupnost Bernoulliho pokusů): Posloupnost n nezávislých opakování Bernoulliho pokusů.

Jaká je pravděpodobnost, že nastane právě k úspěchů?

Zvolme $\Omega = \{0,1\}^n$ (n-tice z 0 a 1), úspěch 1, neúspěch 0, tedy elementární jevy jsou n-tice 0 a 1. Jde o pravděpodobnost $P(B_k)$ jevu B_k, kde B_k je množina všech n-tic obsahujících právě k jedniček. Každá taková n-tice má vzhledem k nezávislosti pravděpodobnost $p^k \cdot q^{n-k}$. B_k obsahuje $\binom{n}{k}$ takových n-tic (=počet výběrů k pozic z n pozic, tj. počet kombinací k z n). Je tedy

$$P(B_k) = \binom{n}{k} p^k q^{n-k}.$$

Všimněme si, že systém množin B_k tvoří rozklad množiny Ω, tj. tvoří tzv. úplný systém jevů dle (viz kapitola 3.5).

Jak rozumět zmíněné nezávislosti? Přesně takto: Elementární jev $\omega = \langle 1,\ldots,1,0,\ldots,0\rangle$ (k jedniček následovaných $n-k$ nulami) lze vyjádřit takto. Nechť A_i reprezentuje „úspěch při pokusu i“, tj. je to množina n-tic, které mají na pozici i (a možná i na dalších pozicích) hodnotu 1. Pak

$$\{\omega\} = A_1 \cap \cdots \cap A_k \cap \overline{A_{k+1}} \cap \cdots \cap \overline{A_n}.$$

Nezávislost (v zadání zmíněná neformálně) přesně znamená, že jevy $A_1,\ldots,A_n$ jsou vzájemně nezávislé dle definice 3.5. Tedy i jevy $A_1,\ldots,A_k,\overline{A_{k+1}},\ldots,\overline{A_n}$ jsou vzájemně nezávislé dle poznámky 3.9 (b). Protože $P(A_i) = p$ a $P(\overline{A_i}) = q$, platí Tedy platí

$$\begin{aligned} P(\{\omega\}) &= P(A_1 \cap \cdots \cap A_k \cap \overline{A_{k+1}} \cap \cdots \cap \overline{A_n}) = \\ &= P(A_1)\cdots P(A_k)\cdot P(\overline{A_{k+1}})\cdots P(\overline{A_n}) = p^k \cdot q^{n-k}. \end{aligned}$$

Pozorujeme-li počet úspešných pokusů, pak vlastně pozorujeme na množině $\Omega = \{0,1\}^n$ z příkladu **??** náhodnou veličinu X definovanou

$$X(\omega) = \omega_1 + \cdots + \omega_n,$$

tj. $X(\omega)$ je počet jedniček ve vektoru ω. Pro pravděpodobnostní funkci p_X této náhodné veličiny platí $p_X(k) = \binom{n}{k} p^k q^{n-k}$, neboť $p_X(k) = P(X = k) = P(B_k)$. Takovým náhodným veličinám se říká veličiny s binomickým rozdělením.

Obecně říkáme, že pravděpodobnostní funkce náhodné veličiny X je *binomická* (*s parametry* n *a* p), pokud pro $q = 1 - p$ platí

$$p_X(k) = P(X = k) = \begin{cases} \binom{n}{k} p^k q^{n-k} & \text{pro } k = 0, 1, 2, \dots, n \\ 0 & \text{jinak.} \end{cases}$$

V tom případě také říkáme, že náhodná veličina X má *binomické rozdělení* (*s parametry* n *a* p). Pro p_X se v tomto případě často používá $B(n, p)$ a píšeme

$$p_X(k) = B(k; n, p).$$

Příkladem je tedy výše uvedené náhodná veličina ma množině $\Omega = \{0, 1\}^n$ výsledků opakovaných Bernoulliho pokusů. (Podobně budeme hovořit o dalších speciálních typech náhodných veličin, tj. budeme říkat, že mají jisté rozdělení a budeme tím rozumět, že jejich pravděpodobnostní funkce má jistý tvar.)

Z binomické věty snadno plyne, že

$$\sum_{k=0}^{n} p_X(k) = \sum_{i=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = (p + (1-p))^n = 1,$$

kterýžto vztah plyne z toho, že množiny B_k (viz výše) tvoří úplný systém jevů. Uvedený vztah k binomické větě je důvodem pro název tohoto rozdělení.

Příslušná distribuční funkce je dána takto:

$$F_X(x) = \begin{cases} 0 & \text{pro } x < 0 \\ \sum_{k=0}^{\lfloor x \rfloor} \binom{n}{k} p^k q^{n-k} & \text{pro } 0 \le x \le n \\ 1 & \text{pro } x > n. \end{cases}$$

kde $\lfloor x \rfloor$ je dolní celá část čísla x.

Lze dokázat (více přednášky), že pro střední hodnoty, rozptyl, koeficient šikmosti a koeficient špičatosti platí:

$$\begin{aligned} E(X) &= n \cdot p, \\ \operatorname{var}(X) &= np(1-p), \\ \alpha_3(X) &= \frac{1-2p}{\sqrt{np(1-p)}}, \\ \alpha_4(X) &= \frac{1-6p(1-p)}{np(1-p)}. \end{aligned}$$

Poznamenejme, že speciálním případem pro $n = 1$ je rozdělení nazývané *alternativní* (někdy Bernoulliho). Uveďte vzorec pro distribuční funkci F náhodné veličiny s alternativním rozdělením a nakreslete graf F.

Typickou situací, ve které se vyskytuje Bernoulliho rozdělení, je právě série nezávisle provedených Bernoulliho pokusů. Jiné interpretace: Z množiny objektů, které jsou dvou druhů (bezvadné, kterých je $100p\%$, a vadné, kterých je $100(1-p)\%$), taháme n výrobků, po vytažení výrobek vždy vrátíme do množiny; číslo k je pak počet bezvadných výrobků, které byly vytaženy. Další ukazuje následující příklad.

Příklad 4.8. Po komunikačním kanálu posíláme n znaků, pravděpodobnost, že znak je přenesen bez chyby je p. Pravděpodobnost, že je přeneseno právě k správných znaků je tedy $B(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$. Pravděpodobnost, že dojde právě ke k chybám je pak $p_X(n-k) = \binom{n}{n-k} p^{n-k} (1-p)^k$. $\diamondsuit$

Poznámka 4.6. Uvedený model takových reálných situací se nazývá *binomický model*. Je to model, přijímáme určité předpoklady, které nemusí být zcela splněny, zejména předpoklady o parametrech n a p.

Máme-li danou reálnou situaci, můžeme přijmout předpoklady, vytvořit její model a studovat ji právě prostřednictvím jejího modelu. To je jedna z hlavních věcí, kterou nám teorie pravděpodobnosti a statistiky nabízí.

Příklad 4.9. Ve výrobě provádí kvalitu kontroly tak, že vyberou sadu 35 výrobků a zjistí, kolik z nich je vadných. Takových výběrů sady 35 výrobků proběhlo 800. Výsedky jsou shrnuty v tabulce 1. Nabízí se otázka, zda se počty vadných výrobků řídí nějakou zákonitostí. Počet vadných výrobků v sadě 35 výrobků lze nahlížet jako náhodnou veličinu X. Jako prostor elementárních jevů lze v tomto případě uvažovat $\Omega = \{0, 1\}^{35}$: výsledkem, tj. elementárním jevem, je vybraná sada; sady jsou reprezentovány vektory $\omega \in \Omega$; souřadnice i vektoru ω určuje, zda i. vybraný výrobek v sadě je bezvadný ($\omega_i = 0$), nebo vadný ($\omega_i = 1$). Veličina X přiřazuje počet vadných výrobků, tj. $X(\omega) = \omega_1 + \cdots + \omega_{35}$. Vzniká otázka, zda reálnou situaci, kterou pozorujeme, lze vysvětlit pomocí jejího vhodného pravděpodobnostního modelu. Zdá se rozumné předpokládat, že tato reálná situace je v souladu s binomickým modelem, tedy že výrobky jsou s jistou pravděpodobností p vadné a že tato pravděpodobnost pak určuje počty chyb ve vybraných sadách. Pokud to tak je, měly by pozorované pravděpodobnosti $\pi_X(k)$ dané poměry $s/800$ být přibližně rovny hodnotám pravděpodobnosti $p_X(k)$

k	s	$\frac{s}{800}$
0	11	0.01375
1	95	0.11875
2	139	0.17375
3	213	0.26625
4	143	0.17875
5	113	0.14125
6	49	0.06125
7	27	0.03375
8	6	0.00750
9	4	0.00500
10	0	0.00000
$\sum$	800	1.0

Tabulka 1: Vyhodnocení 800 výběrů sad výroků z příkladu 4.9. k je počet vadných výrobků v sadě, s je počet sad s daným počtem vadných výrobků.

podle binomického rozdělení $B(n, p)$, kde $n = 35$, tj. $\pi_X(k)$ by se mělo přibližně rovnat $B(k; 35, p)$ pro vhodnou hodnotu parametru p. Podívejme se, jak to dopadne pro $p = 0.1$. Výsledky ukazuje tabulka 2. Vidíme pozoruhodnou shodu, která nás inuitivně opravňuje k závěru, že model založený na binomické rozdělení dobře popisuje pozorovanou situaci. Jak parametr p určit a jak se vůbec postavit k tvrzení, zda pro daný parametr binomické rozdělení dobře popisuje pozorovanou situaci, popř. do jaké míry ji dobře popisuje, je otázka, kterou se zabývá statistika (takovými otázkami se budeme zabývat později). $\Diamond$

Zajímavost 4.1 (Simpsonův paradox). V kontextu porovnávání kvality je relevantní tzv. Simpsonův paradox[6] Předpokládejme, že dostaneme dvě dodávky výrobků od výrobce A a dvě od výrobce B. Od výrobce A je v první dodávce 100 výrobků, z nich 25 vadných, ve druhé pak 30 výrobků, z nich 5 vadných. Od výrobce B je v první dodávce 50 výrobků, z nich 14 vadných, ve druhé pak 150 výrobků, z nich 30 vadných. Který výrobce je spolehlivější?

Posuzujeme-li podle první dodávky, je spolehlivější výrobce A, neboť podíl vadných výrobků je $25/100 = 0.25$, zatímco u výrobce B je to $14/50 =$

[6] Edward H. Simpson, „The interpretation of interaction in contingency tables," J. Royal Statistical Society, Ser. B 13 (1951): 238–241.

k	$\pi_X(k)$	$p_X(k) = B(k; 35, 0.1)$
0	0.01375	0.0250
1	0.11875	0.0974
2	0.17375	0.1839
3	0.26625	0.2248
4	0.17875	0.1998
5	0.14125	0.1376
6	0.06125	0.0765
7	0.03375	0.0352
8	0.00750	0.0137
9	0.00500	0.0046
10	0.00000	0.0013

Tabulka 2: Porovnání pozorovaných pravděpodobností $\pi_X(k)$, že počet vadných výrobků v sadě je k, a teoretických hodnot $p_X(k) = B(k; 35, 0.1)$.

0.28. Posuzujeme-li podle druhé dodávky, je opět spolehlivější výrobce A, neboť podíl vadných výrobků je $5/30 = 0.167$, zatímco u výrobce B je to $30/150 = 0.2$. Posuzujeme-li však podle výrobků z obou dodávek, je úspěšnější výrobce B, protože podíl vadných výrobků je u něho $(14 + 30)/(50 + 150) \approx 0.22$, zatímco u výrobce A je to $(25 + 5)/(100 + 30) \approx 0.23$.

Kde je chyba, když intuitivně uvažujeme, že je-li A lepší z pohledu každé z obou dodávek, je lepší i z celkového pohledu? Výrobce B totiž dosáhl dobré spolehlivosti na velkém souboru druhé dodávky. Ve druhé dodávce dosáhl A sice ještě lepší spolehlivosti, ale tato dodávka byla malá, v jistém smyslu je to výjimka, o A více vypovídá spolehlivost z první dodávky. Pro B je naopak první dodávka malá a více vypovídající je dodávka druhá.

Může uvedený jev nastat i pokud celkový počet výrobků od A je stejný jako celkový počet od B?

Je-li n velké, není schůzdné hodnoty $B(k; n, p)$ počítat přímo dle vzorce. Lze je počítat dle tzv. Laplaceovy aproximace, podle které je

$$B(k; n, p) \approx \frac{1}{\sqrt{2\pi npq}} \cdot e^{\frac{-(k-np)^2}{2npq}}.$$

To je netriviální aproximace, my zůstaneme na této intuitivní úrovni (více přednášky nebo literatura). Jiný způsob aproximace (Poissonovým rozdělením) ukážeme dále.

Příklad 4.10. Nakreslete graf funkce $B(\cdot; n, p)$ pro různé hodnoty n a p. Co lze říci o šikmosti binomického rozdělení? $\Diamond$

Geometrické rozdělení

Předpokládejme, že je prováděna posloupnost Bernoulliho pokusů (úspěch s pravděpodobností p, neúspěch s pravděpodobností $1-p$). Zajímá nás, po kolika pokusech se dostaví první úspěšný pokus. Za elementární jevy lze považovat vektory $0^{k-1}1 = \langle 0, \dots, 0, 1\rangle$, tj. $k-1$ nul (může být $k-1 = 0$) následovaných jedničkou. Náhodnou veličinou je veličina definovaná takto: $X(0^{k-1}1) = k$.

Snadno se vidí, že pro $k = 1, 2, \dots$ platí

$$p_X(k) = P(X = k) = p \cdot (1-p)^{k-1}, \tag{20}$$

neboť $P(X = k) = P(\{\omega \in \Omega \mid X(\omega) = k\}) = P(\{0^{k-1}1\}) = (1-p)^{k-1} \cdot p$. Je to totiž pravděpodobnost jevu, ve kterém je nejdříve $k-1$ neúspěšných pokusů následovaných jedním úspěšným.

Snadno se ověří, že

$$\sum_{k=1}^{\infty} p_X(k) = 1.$$

O náhodné veličině, jejíž pravděpodobnostní funkce je dána (22), řekneme, že má *geometrické rozdělení* s parametrem p; příslušná pravděpodobnostní funkce p_X se pak nazývá *geometrická*. Název je odvozen od toho, že posloupnosti hodnot $p_X(1)$, $p_X(2)$, … tvoří geometrickou posloupnost první člen je $a_1 = p$, dále je $a_{n+1} = (1-p) \cdot a_n$.

Někdy se uvažuje $p_X(k) = P(X = k) = p\cdot(1-p)^k$, tj. $p_X(k)$ je pravděpodobnost, že úspěšný pokus nastane po k neúspěšných; příslušné rozdělení se nazývá XXX.

Příslušná distribuční funkce je pro $x \geq 0$ dána takto (použijte vzorec pro součet geometrické posloupnosti, tj. vzorec $1 + q + \cdots + q^{M-1} = \frac{1-q^M}{1-q}$, a aplikujte ho na $1 + (1-p) + \cdots + (1-p)^{\lfloor x\rfloor - 1}$):

$$F_X(x) = \sum_{k=1}^{\lfloor x\rfloor} p(1-p)^{k-1} = p\frac{(1-p)^{\lfloor x\rfloor} - 1}{(1-p) - 1} = 1 - (1-p)^{\lfloor x\rfloor},$$

kde $\lfloor x\rfloor$ je dolní celá část čísla x.

Lze dokázat (více přednášky), že pro střední hodnoty, rozptyl, koeficient šikmosti a koeficient špičatosti platí:

$$\begin{aligned} E(X) &= \frac{1}{p}, \\ \mathrm{var}(X) &= \frac{1-p}{p^2}, \\ \alpha_3(X) &= \frac{2-p}{\sqrt{1-p}}, \\ \alpha_4(X) &= 6 + \frac{p^2}{(1-p)}. \end{aligned}$$

Příklad 4.11. Předpokládejme, že zdrojový kód má obsahuje cyklus

$$\textbf{repeat}\ B \quad \textbf{until}\ C$$

a že pravděpodobnost splnění booleovské podmínky C je p. Pak počet průchodů tímto cyklem je náhodná veličina, která má geometrické rozdělení s parametrem p. Rozeberte podrobněji (co jsou elementární jevy? apod.). $\Diamond$

Negativní binomické rozdělení

K geometrickému rozdělení vedlo pozorování počtu pokusů po první úspěšný. Předpokládejme opět, že je prováděna posloupnost Bernoulliho pokusů (úspěch s pravděpodobností p, neúspěch s pravděpodobností $q = 1 - p$). Zajímá nás, po kolika pokusech se dostaví r. úspěšný pokus.

Za elementární jevy lze považovat vektory $\omega \in \bigcup_{n=1}^{\infty}\{0,1\}^n$ obsahující právě r jedniček a končící jedničkou. Příslušnou náhodnou veličinu označme X_r; je definována takto: $X_r(\omega) = |\omega|$ (počet souřadnic). Tato veličina nabývá hodnot $r, r+1, r+2, \ldots$ Pro výpočet pravděpodobnosti $P(X_r = k)$ uvažujme jevy $A \sim$ „v prvních $n-1$ pokusech nastalo právě $r-1$ úspěchů" a $B \sim$ „n. pokus je úspěšný". Je zřejmé, že $P(X_r = k) = P(A \cap B)$. Vzhledem k nezávislosti jednotlivých pokusů jsou A a B nezávislé, a tedy $P(A \cap B) = P(A)P(B)$. Přitom je

$$P(A) = \binom{k-1}{r-1} p^{r-1} q^{k-r} = \binom{k-1}{r-1} p^{r-1}(1-p)^{k-r} \quad \text{a} \quad P(B) = p.$$

Celkem tedy

$$p_{X_r}(k) = P(X_r = k) = \binom{k-1}{r-1} p^r (1-p)^{k-r}. \tag{21}$$

O náhodné veličině, jejíž pravděpodobnostní funkce je dána (21), řekneme, že má *negativní binomické rozdělení* s parametrem p. Lze ukázat, že

$$p_{X_r}(k) = (-1)^{k-r}\binom{-r}{k-r}p^r(1-p)^{k-r}.$$

Odtud pochází název tohoto rozdělení (porovnejte se vzorcem pro binomické rozdělení).

Lze ověřit, že

$$\sum_{k=1}^{\infty} p_{X_r}(k) = 1.$$

Poznámka 4.7. Často se používá následující alternativní definice, příslušné rozdělení se pak nazývá *modifikované negativní binomické rozdělení*. Příslušná náhodná veličina Y_r pak přiřazuje počet neúspěšných pokusů, než nastane r. úspěšný. Lze snadno ověřit, že

$$P(Y_r = k) = \binom{k+r-1}{r-1}p^r(1-p)^k$$

Pro $r = 1$ je to právě modifikované geometrické rozdělení.

Poznámka 4.8. Existuje také rozšíření výše uvedeného rozdělení pro reálné hodnoty r. Pak samozřejmě nelze použít výše uvedenou interpretaci, ve které r je počet úspěchů. Více viz literatura.

Poissonovo rozdělení

Předpokládejme, že pozorujeme počet úloh, které přijdou v časovém intervalu $(0, t]$ ke zpracování. Předpokládejme dále, že pravděpodobnost, že pro dostatečně malý časový úsek, Δt, je pravděpodobnost, že se v tomto časovém úseku objeví nově příchozí úloha, je $\alpha \cdot \Delta t$ a že pravděpodobnost, že se v tomto úseku objeví dvě nebo více úloh je nulová (nebo je tak malá, že ji za nulovou můžeme považovat). Chceme zjistit pravděpodobnost toho, že se v intervalu $(0, t]$ nebo v jiném intervalu délky t objeví k nových úloh. Rozdělme interval délky t na n intervalů délky t/n. Předpokládejme, že to, zda se nová úloha objeví v každém takovém intervalu délky t/n je nezávislé na tom, zda se nová úloha objeví v nějakém jiném takovém intervalu. Pro dostatečně velké n lze na situaci pohlížet jako na posloupnost n Bernoulliho pokusů s pravděpodobností úspěchu (úloha přijde) $p = \alpha \cdot t/n$.

Pravděpodobnost, že během těchto n intervalů délky t/n přijde k nových úloh je tedy pro $k = 0, 1, 2, \dots$

$$B(k; n, {}^{\alpha t}\!/\!{}_{n}) = \binom{n}{k} ({}^{\alpha t}\!/\!{}_{n})^k (1 - {}^{\alpha t}\!/\!{}_{n})^{n-k}$$

Kvůli výše uvedeným předpokladům nás zajímá chování pro velmi malé intervaly, tj. velká n. Vyšetřujme tedy limitu uvedeného výrazu pro n jdoucí k ∞. Je

$$\begin{aligned} B(k; n, {}^{\alpha t}\!/\!{}_{n}) &= \frac{n(n-1)\cdot(n-k+1)}{k!n^k} (\alpha t)^k (1 - {}^{\alpha t}\!/\!{}_{n})^{n-k} = \\ &= \frac{n}{n}\frac{n-1}{n} \cdots \frac{n-k+1}{n} \frac{(\alpha t)^k}{k!} (1 - {}^{\alpha t}\!/\!{}_{n})^{-k} (1 - {}^{\alpha t}\!/\!{}_{n})^{n}. \end{aligned}$$

Při $n \to \infty$ jsou limity jednotlivých členů následující: Limita každého z prvních k činitelů je 1; limita dalšího činitele je $\frac{(\alpha t)^k}{k!}$ (nezávisí na n); limita předposledního činitele je 1; pro limitu posledního činitele platí

$$\lim_{n\to\infty} (1 - {}^{\alpha t}\!/\!{}_{n})^n = \lim_{n\to\infty} ([1 - {}^{\alpha t}\!/\!{}_{n}]^{-n/(\alpha t)})^{-\alpha t} = \lim_{h\to 0} ([1+h]^{1/h})^{-\alpha t} = e^{-\alpha t},$$

protože $\lim_{h\to 0}[1+h]^{1/h} = e$ je definice čísla e. Je tedy

$$\lim_{n\to\infty} B(k; n, {}^{\alpha t}\!/\!{}_{n}) = e^{-\alpha t} \frac{(\alpha t)^k}{k!}$$

a provedeme-li substituci λ za αt, dostaneme, že hledaná limita je rovna

$$e^{-\lambda} \frac{\lambda^k}{k!}.$$

Tak jsme přirozeně došli k pravděpodobnostní funkci náhodné veličiny s tzv. Poissonovým rozdělením.

Náhodná veličina X má *Poissonovo rozdělení* s parametrem λ, pokud pro její pravděpodobnostní funkci platí

$$p_X(k) = P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!} \tag{22}$$

pro $k = 0, 1, 2, \dots$ a $p_X(k) = 0$ jinak. Jak jsme viděli, takové rozdělení má náhodná veličina udávající počet výskytů nějakých událostí (např. příchozích úloh) za danou časovou jednotku.

Approximace binomického rozdělení Poissonovým: pro velká n a malá p (např. se doporučuje $n > 30$ a $p < 0.1$) platí pro $\lambda = np$, že

$$\binom{n}{k} p^k (1-p)^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!}.$$

Lze ověřit, že

$$\sum_{k=1}^{\infty} p_X(k) = 1.$$

Lze dokázat (více přednášky), že pro střední hodnoty, rozptyl, koeficient šikmosti a koeficient špičatosti platí:

$$\begin{aligned} E(X) &= \lambda, \\ \operatorname{var}(X) &= \lambda, \\ \alpha_3(X) &= {}^1\!/\!\sqrt{\lambda}, \\ \alpha_4(X) &= {}^1\!/\lambda. \end{aligned}$$

Příklad 4.12. DOPLNIT ♢

Hypergeometrické rozdělení

Předpokládejme, že z N objektů, mezi kterými je d objektů s defektem, vybereme (bez opakování, tj. vybraný objekt nevracíme zpět) n objektů. Jaká je pravděpodobnost, že mezi vybranými objekty je právě k objektů s defektem. Tato pravděpodobnost se značí $h(k; n, d, N)$ a snadno se vidí, že pro $\max\{0, d+n-N\} \leq k \leq \min\{d, n\}$ platí

$$h(k; n, d, N) = \frac{\binom{d}{k} \cdot \binom{N-d}{n-k}}{\binom{N}{n}}.$$

Pokud pro pravděpodobnostní funkci p_X náhodné veličiny X je $p_X(k) = h(k; n, d, N)$, říkáme, že X má *hypergeometrické rozdělení*.

Lze dokázat (více přednášky), že pro střední hodnoty, rozptyl, koeficient šikmosti a koeficient špičatosti platí:

$$\begin{aligned} E(X) &= n\frac{d}{N}, \\ \operatorname{var}(X) &= n\frac{d}{N}(1-\frac{d}{N})\frac{N-n}{N-1}. \end{aligned}$$

Příklad 4.13. ♢

Rovnoměrné (diskrétní) rozdělení

Náhodná veličina s hodnotami $\{x_1, \ldots, x_n\}$, taková, že pravděpodobnostní funkce p_X nabývá stejných hodnot p, tj. $p_X(x_k) = p$ pro každé $k = 1, \ldots, n$. Pak je zřejmě

$$p_X(x_k) = \frac{1}{n}$$

pro $k = 1, \ldots, n$ a $p_X(x) = 0$ jinak.

Předpokládejme pro konkrétnost, že $\{x_1, \ldots, x_n\} = \{1, \ldots, n\}$. Pro distrubuční funkci pak platí

$$F_X(x) = \sum_{k=1}^{\lfloor x \rfloor} p_X(k) = \frac{\lfloor x \rfloor}{n}$$

pro $1 \leq x \leq n$, $F_X(x) = 0$ pro $x < 0$ a $F_X(x) = 1$ pro $x > n$.

Snadno se dokáže, že pro střední hodnotu, rozptyl, koeficient šikmosti a koeficient špičatosti platí:

$$\begin{aligned} E(X) &= \frac{n+1}{2}, \\ \mathrm{var}(X) &= \frac{n^2-1}{12}. \end{aligned}$$

Konstantní náhodná veličina a indikátorová náhodná veličina

Konstantní náhodná veličina přiřazuje každému $\omega \in \Omega$ jednu danou hodnotu, c. Tj. je $X(\omega) = c$ pro každé $\omega \in \Omega$. Platí pak

$$p_X(x) = P(X = x) = \begin{cases} 1 & \text{pro } x = c, \\ 0 & \text{jinak} \end{cases}$$

a

$$F_X(x) = \begin{cases} 0 & \text{pro } x < c, \\ 1 & \text{pro } x \geq c. \end{cases}$$

Nechť $A \subseteq \Omega$ je jev. Pak náhodná veličina I_A definovaná

$$I_A(\omega) = \begin{cases} 1 & \text{pro } \omega \in A, \\ 0 & \text{pro } \omega \notin A \end{cases}$$

se nazývá *indikátor* jevu A. Pro její pravděpodobnostní funkci zřejmě platí $p_X(0) = 1 - P(A)$ a $p_X(1) = P(A)$. Jde o užitečný technický pojem.

4.3 Spojité náhodné veličiny

4.3.1 Hustota pravděpodobnosti a distribuční funkce spojité náhodné veličiny

Náhodná veličina může nabývat i nespočetně mnoha hodnot. Distribuční funkce pak může být spojitá, tj. ne skokovitá jako v případě diskrétní náhodné veličiny.

Definice 4.6. Náhodná veličina se nazývá *spojitá*, pokud existuje nezáporná reálná funkce f_X tak, že pro každé $x \in \mathbb{R}$ je

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t. \tag{23}$$

V takovém případě se f_X nazývá *hustota pravděpodobnosti* (někdy jen hustota, angl. probability density function, pdf) náhodné veličiny X.

Někdy se také říká, že X *má rozdělení spojitého typu*. Často také píšeme jen f místo f_X.

Poznámka 4.9. (a) Stejně jako v případě diskrétní veličiny lze dokázat, že

$$P(X \in (a,b]) = P(X \in (-\infty,b]) - P(X \in (-\infty,a]) = F_X(b) - F_X(a)$$

a vzhledem k $F_X(b) - F_X(a) = \int_a^b f(t)\mathrm{d}t$ je

$$P(X \in (a,b]) = P(X \in (-\infty,b]) - P(X \in (-\infty,a]) = \int_a^b f(t)\mathrm{d}t.$$

(b) Hustota náhodné veličiny nese informaci o distribuční funkci spojité náhodné veličiny. V případě diskrétní veličiny podobnou informaci nese pravděpodobnostní funkce náhodné veličiny definovaná vztahem $p_X(x) = P(X = x)$. Tímto způsobem však o spojité veličině informaci nést nemůžeme. Pravděpodobnost, že spojitá náhodná veličina nabývá hodnoty x, je totiž pro každé x nulová: $P(X = x) = 0$. Je totiž $P(X = x) = P(X \leq x) - P(X < x) = F_X(x) - \lim_{t\to x^-} F_X(t) = F_X(x) - F_X(x) = 0$. Z toho vyplývá, že

$$P(X \in (a,b]) = P(X \in [a,b)) = P(X \in (a,b)) = P(X \in [a,b]) = F_X(b) - F_X(a).$$

(c) Základní vlastnosti F_X (snadno se vidí, podobně jako u diskrétních):

– pro každé $x \in \mathbb{R}$ je $0 \leq F_X(x) \leq 1$;

– $x \leq y$ implikuje $F_X(x) \leq F_X(y)$, tj. F_X je neklesající;

– $\lim_{x\to\infty} F_X(x) = 1$ a $\lim_{x\to-\infty} F_X(x) = 0$.

– F je spojitá v každém bodě; f nemusí být spojitá. V každém bodě, kde má F derivaci, platí

$$\frac{\mathrm{d}F}{\mathrm{d}x} = f(x).$$

OBRAZEK ILUSTRUJICI FUNKCI F, f a $P(a < X \leq b)$.

Příklad 4.14. Uvažujme pro $h > 0$ interval $(\mu - h, \mu + h)$ a funkci

$$f(x) = \begin{cases} 1/2h & \text{pro } x \in (\mu - h, \mu + h), \\ 0 & \text{jinak.} \end{cases}$$

Pak $F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t$ je dána následovně:

$$F(x) = \begin{cases} 0 & \text{pro } x \leq \mu - h, \\ 1/2h(x - \mu + h) & \text{pro } x \in (\mu - h, \mu + h), \\ 1 & \text{pro } x \geq \mu - h. \end{cases}$$

Náhodná veličina s takovou hustotou pravděpodobnosti se nazává veličinou s *rovnoměrným rozdělením* na intervalu $(\mu - h, \mu + h)$. ♢

4.3.2 Charakteristiky spojitých náhodných veličin

Nyní se budeme zabývat charkateristikami spojitých náhodných veličin. Jejich význam je analogický charakteristikám diskrétních veličin.

Budeme předpokládat, že X je spojitá náhodná veličina, F_X její distribuční funkce a f_X její hustota pravděpodobnosti.

Střední hodnota

Střední hodnota spojité náhodné veličiny X se značí $E(X)$ a je definována vzorcem

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x. \tag{24}$$

Příklad 4.15. Pro spojitou veličinu s rovnoměrným rozdělením (příklad 4.15) platí

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\mathrm{d}x = \int_{\mu-h}^{\mu+h} \frac{x}{2h}\mathrm{d}x = \frac{(\mu+h)^2 - (\mu-h)^2}{4h} = \mu.$$

$\diamondsuit$

Střední hodnota vyjadřuje, s přihlédnutím k rozdělení pravděpodobnosti, očekávanou hodnotu výsledku náhodného pokusu.

Jako v diskrétním případě platí, že jsou-li X a Y náhodné veličiny (na stejném pravděpodobnostním prostoru) a $c \in \mathbb{R}$, pak i funkce $X+Y$ a $c \cdot X$ jsou náhodné veličiny ($[X+Y](\omega) = X(\omega) + Y(\omega)$. $[c \cdot X](\omega) = c \cdot X(\omega)$).

Věta 4.2 (linearita střední hodnoty)**.** *Pro náhodné veličiny X a Y a konstanty $a, b, c \in \mathbb{R}$ platí: $E(X+Y) = E(X) + E(Y)$; $E(cX) = cE(X)$; tedy $E(aX + bY) = aE(X) + bE(Y)$; dále pak $E(X+c) = E(X) + c$; tedy $E(aX+b) = aE(X) + b$.*

Důkaz. Viz přednášky. $\square$

Momenty

Připoměňme lemma 4.1 (složením spojité funkce h a náhodné veličiny X vznikne náhodná veličina).

Lemma 4.3. *Je-li $X : \Omega \to \mathbb{R}$ spojitá náhodná veličina a $h : \mathbb{R} \to \mathbb{R}$ spojitá funkce (obecněji: borelovsky měřitelná funkce), pak pro náhodnou veličinu $Y = h(X)$ platí*

$$E(Y) = \int_{-\infty}^{\infty} h(x) f_X(x)\mathrm{d}x. \tag{25}$$

Momenty se definují stejně jako v případě diskrétních náhodných veličin; mají analogické interpretace. Platí pro ně stejné vztahy jako ty, které jsme uvedli pro diskrétní případ. Tedy: Nechť $r = 1, 2, \ldots$ Pak r. *obecný moment* $\mu'_r(X)$ náhodné veličiny X je střední hodnota náhodné veličiny $h(X) = X^r$, tedy

$$\mu'_r(X) = E(X^r).$$

r. *centrální moment* $\mu_r(X)$ náhodné veličiny X je střední hodnota náhodné veličiny $h(X) = (X - E(X))^r$, tedy

$$\mu_r(X) = E((X - E(X))^r).$$

Snadno se ověří, že platí:

$$\begin{aligned}\mu_0(X) &= 1,\\ \mu_1(X) &= 0,\\ \mu_2(X) &= {\mu'}_2(X) - E(X)^2.\end{aligned}$$

Rozptyl (také *variance*) $\operatorname{var} X$ náhodné veličiny X je druhý centrální moment X, tj.

$$\operatorname{var} X = \mu_2(X) = E((E - E(X))^2) = {\mu'}_2(X) - E(X)^2.$$

Směrodatná odchylka σ náhodné veličiny X je druhá odmocnina rozptylu, tj.

$$\sigma = \sqrt{\operatorname{var} X}.$$

Platí

$$\operatorname{var}(aX + b) = a^2 \operatorname{var}(X).$$

Příklad 4.16. Uvažujme spojitou veličinu s rovnoměrným rozdělením (příklad 4.15). Platí

$$\mu_2'(X) = E(X^2) = \int_{\mu-h}^{\mu+h} \frac{x^2}{2h} \mathrm{d}x = \frac{(\mu+h)^3 - (\mu-h)^3}{6h} = \mu^2 + \frac{h^2}{3},$$

a tedy (viz příklad 4.15)

$$\operatorname{var} X = \mu_2'(X) - (E(X))^2 = \mu^2 + \frac{h^2}{3} - \mu^2 = \frac{h^2}{3}.$$

$\Diamond$

Kvantily

Zavádí se jako u diskrétních náhodných veličin (tj. definice jsou stejné, jen v nich X je spojitou nádodnou veličinou).

Charakteristiky polohy, variability, šikmosti a špičatosti

Zavádí se jako u diskrétních náhodných veličin (tj. definice jsou stejné, jen v nich X je spojitou nádodnou veličinou).

Modus spojité náhodné veličiny je hodnota $\hat{x}$, pro kterou $f(\hat{x}) \geq f(x)$ pro každé $x \in \mathbb{R}$.

4.3.3 Vybrané spojité náhodné veličiny

Normální rozdělení

Normální rozdělení (někdy též Gaussovo nebo Laplaceovo-Gaussovo) je jedním z nejdůležitějších rozdělení spojitých náhodných veličin. Tímto rozdělením lze modelovat veličiny jako je hmotnost nebo výška lidí i řadu jiných veličin, které mají společné to, že jsou v populaci rozmístěny symetricky kolem očekávané hodnoty μ, a to tak, že s rostoucí vzdáleností od očekávané hodnoty klesá pravděpodobnost výskytu jedinců s danou hodnotou.

Spojitá náhodná veličina má *normální rozdělení* s parametry $\mu \in \mathbb{R}$ a $\sigma^2 > 0$, jestliže její hustota pravděpodobnosti je definována pro každé $x \in \mathbb{R}$ následovně:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{26}$$

Říkáme také, že jde o rozdělení $N(\mu, \sigma^2)$. Předpokládá se, že $\sigma > 0$; proč se za parametr považuje σ^2, uvidíme za chvíli.

Normální rozdělení $N(0,1)$ se nazývá *normované* (někdy také *standardizované*). Jeho hustota, φ, je tedy dána vztahem

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}. \tag{27}$$

Snadno se vidí, že

$$f(x) = \frac{1}{\sigma}\varphi(\frac{x-\mu}{\sigma}).$$

Platí důležité tvrzení (důkaz viz příklad 4.26 na str. 90): Má-li náhodná veličina X rozdělení $N(\mu, \sigma^2)$, má veličina

$$U = \frac{X-\mu}{\sigma}$$

rozdělení $N(0,1)$.

Veličina s rozdělením $N(\mu, \sigma^2)$ má dle definice distribuční funkci

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \mathrm{d}t.$$

Veličina s normovaným rozdělením $N(0,1)$ má distribuční funkci

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \mathrm{d}t.$$

GRAFY FUNKCÍ f a F, VIZ PŘEDNÁŠKA

Protože $U = (X-\mu)/\sigma$ má rozdělení $N(0,1)$, je $\Phi(u) = P(U \leq u)$. Platí tedy následující důležitý vztah:

$$F(x) = P(X \leq x) = P(\frac{X-\mu}{\sigma} \leq \frac{x-\mu}{\sigma}) = P(U \leq \frac{x-\mu}{\sigma}) = \Phi(\frac{x-\mu}{\sigma}).$$

Lze dokázat (více přednášky), že pro střední hodnoty, rozptyl, koeficient šikmosti a koeficient špičatosti náhodné veličiny s rozdělením $N(\mu, \sigma^2)$ platí:

$$\begin{aligned} E(X) &= \mu, \\ \mathrm{var}(X) &= \sigma^2, \\ \alpha_3(X) &= 0, \\ \alpha_4(X) &= 0. \end{aligned}$$

Platí také následující důležité tvrzení (bez důkazu): Jsou-li $X_1, \ldots, X_n$ vzájemně nezávislé náhodné veličiny (definice viz dále) s rozděleními $N(\mu_i, \sigma_i^2)$ a jsou-li $a_1, \ldots, a_n, b \in \mathbb{R}$, kde aspoň jedno a_i je nenulové, pak veličina $Y = \sum_{i=1}^n a_i X_i + b$ má normální rozdělení s parametry $E(Y) = \sum_{i=1}^n a_i \mu_i + b$ (střední hodnota) a $\mathrm{var}(Y) = \sum_{i=1}^n a_i^2 \sigma_j^2$.

Logaritmicko-normální rozdělení

Je alternativou k normálnímu rozdělení. Tímto rozdělením lze modelovat např. následující veličiny: čas, který čtenáři stráví čtením online článků; délka komentářů v diskuzích na internetu; příjem populaci kromě příjmu v nejvyšší kategorii zahrnující asi 3% populace; v teorii spolehlivosti čas potřebný k opravě systému.

Spojitá náhodná veličina má *logaritmicko-normální rozdělení* (někdy jen lognormální) s parametry $\mu \in \mathbb{R}$ a $\sigma^2 > 0$, jestliže její hustota pravděpodobnosti je definována pro každé $x \in \mathbb{R}$ následovně:

$$f(x) = \begin{cases} \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases} \tag{28}$$

Říkáme také, že jde o rozdělení $LN(\mu, \sigma^2)$. Předpokládá se, že $\sigma > 0$; proč se za parametr považuje σ^2, uvidíme za chvíli.

Platí (viz příklad 4.27 na str. 90), že má-li X rozdělení $LN(\mu, \sigma^2)$, má veličina $Y = \ln X$ rozdělení $N(\mu, \sigma^2)$. Naopak, má-li Y rozdělení $N(\mu, \sigma^2)$, má $X = e^Y$ rozdělení $LN(\mu, \sigma^2)$.

Protože má $\ln X$ rozdělení $N(\mu, \sigma^2)$, má X distribuční funkci F, pro kterou platí (viz vztahy o distribuční funkci Φ normalizovaného normálního rozdělení)

$$F(x) = P(\ln X \le \ln x) = \Phi(\frac{\ln x - u}{\sigma}) \text{ pro } x > 0 \text{ a } = 0 \text{ pro } x \le 0.$$

Lze dokázat (více přednášky), že pro střední hodnoty, rozptyl, koeficient šikmosti a koeficient špičatosti náhodné veličiny s rozdělením $LN(\mu, \sigma^2)$ platí:

$$\begin{aligned} E(X) &= e^{\mu + \frac{\sigma^2}{2}}, \\ \mathrm{var}(X) &= e^{2\mu + \sigma^2}(e^{\sigma^2} - 1), \end{aligned}$$

Exponenciální rozdělení

Exponenciální rozdělení mají náhodné veličiny, které vyjadřují životnost výrobků, např. součástek. (Obecně potom veličiny udávající dobu, po kterou má systém nějaký stav, než dojde k jeho změně, např. čas než přijde telefonní hovor.) Tj. hodnota veličiny X je doba do poruchy dané součástky. Přitom povaha chyby je taková, že nezáleží na tom, jak dlouho součástka pracovala bez chyby (chyba tedy nastává z vnějších přícin, nikoli kvůli opotřebení).

Spojitá náhodná veličina má *exponenciální rozdělení* s parametry $A \in \mathbb{R}$ a $\delta > 0$, jestliže její hustota pravděpodobnosti je definována pro každé $x \in \mathbb{R}$ následovně:

$$f(x) = \begin{cases} \frac{1}{\delta} e^{-\frac{x-A}{\delta}} & \text{pro } x > A, \\ 0 & \text{pro } x \le A. \end{cases} \tag{29}$$

Říkáme také, že jde o rozdělení $E(A, \delta)$.

Pro distribuční funkci platí (viz přednášky)

$$F(x) = \begin{cases} \int_A^x f(t)\mathrm{d}t = 1 - e^{-\frac{x-A}{\delta}} & \text{pro } x > A, \\ 0 & \text{pro } x \le A. \end{cases} \tag{30}$$

Rozdělení $E(0,1)$ se nazývá *normované* (někdy *standardizované*). Snadno se dokáže, že má-li X rozdělení $E(A,\delta)$, má veličina

$$V = \frac{X-A}{\delta}$$

rozdělení $E(0,1)$. Hustota veličiny s normovaným exponenciálním rozdělením je

$$\varphi(v) = \begin{cases} e^{-v} & \text{pro } x > 0, \\ 0 & \text{pro } x \le 0. \end{cases}$$

GRAFY FUNKCÍ f a F, VIZ PŘEDNÁŠKA

Lze dokázat (více přednášky), že pro střední hodnoty, rozptyl, koeficient šikmosti a koeficient špičatosti náhodné veličiny s rozdělením $E(A,\delta)$ platí:

$$\begin{aligned} E(X) &= A + \delta, \\ \mathrm{var}(X) &= \delta^2, \\ \alpha_3(X) &= 2, \\ \alpha_4(X) &= 6. \end{aligned}$$

Weibullovo rozdělení

Weibullovo rozdělení mají náhodné veličiny, které vyjadřují životnost výrobků, např. součástek, u které je významným faktorem opotřebení. Tj. hodnota veličiny X je doba do poruchy dané součástky, přitom povaha chyby je taková, že nastává kvůli opotřebení.

Spojitá náhodná veličina má *Weibullovo rozdělení* s parametry $\delta > 0$ a $c > 0$, jestliže její hustota pravděpodobnosti je definována pro každé $x \in \mathbb{R}$ následovně:

$$f(x) = \begin{cases} \frac{cx^{c-1}}{\delta^c} e^{-(\frac{x}{\delta})^c} & \text{pro } x > 0, \\ 0 & \text{pro } x \le 0. \end{cases} \tag{31}$$

Říkáme také, že jde o rozdělení $W(\delta, c)$.

Všimněme si, že pro $c = 1$ jde o exponencilní rozdělení $E(0,\delta)$.

Platí, že náhodná veličina

$$V = (\frac{X}{\delta})^c$$

má exponenciální rozdělení $E(0,1)$.

Pro distribuční funkci Weibullova rozdělení tedy platí

$$F(x) = \begin{cases} 1 - e^{-(\frac{x}{\delta})^c} & \text{pro } x > 0, \\ 0 & \text{pro } x \leq 0. \end{cases} \tag{32}$$

Lze dokázat (více přednášky), že pro střední hodnoty, rozptyl, koeficient šikmosti a koeficient špičatosti náhodné veličiny s rozdělením $W(\delta, c)$ platí: NEBUDEME UVADET, ZAJEMCE VIZ PREDNASKY; elegantne se vyjadri pomoci tzv. funkce gamma, např.:

$$E(X) = \Gamma(\frac{1}{c} + 1)\delta.$$

4.4 Náhodné vektory

4.4.1 Pojem náhodného vektoru

Náhodnou veličinu chápeme jako číselnou charakteristiku výsledku náhodného pokusu. Často je třeba takových charakteristik sledovat více. Pokud například Ω reprezentuje množinu obyvatel, může nás zajímat náhodná veličina X_1 přiřazující obyvatelům $\omega \in \Omega$ jejich věk $X_1(\omega)$, veličina X_2 přiřazující hmotnost, popřípadě další náhodné veličiny (měsíční příjem, počet dětí a podobně).

Definice 4.7. Jsou-li $X_1, \ldots, X_n$ náhodné veličiny (pro stejný pravděpodobnostní prostor), pak $n$tice $\langle X_1, \ldots, X_n \rangle$ se nazývá *náhodný vektor* (někdy také *n-rozměrná náhodná veličina*).

Distribuční funkce náhodného vektoru $X = \langle X_1, \ldots, X_n \rangle$ (někdy také *sdružená distribuční funkce* náhodných veličin $X_1, \ldots, X_n$) je funkce $F_X : \mathbb{R}^n \to \mathbb{R}$ definovaná

$$F_X(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

Místo F_X, tj. místo $F_{\langle X_1,\ldots,X_n \rangle}$, píšeme také $F_{X_1,\ldots,X_n}$ nebo jen F.

Základní vlastnosti F_X (snadno se vidí, podobně jako u náhodných veličin):

- pro každé $x_1, \ldots, x_n \in \mathbb{R}$ je $0 \leq F_X(x_1, \ldots, x_n) \leq 1$;
- $x_i \leq y_i$ implikuje $F_X(x_1, \ldots, x_n) \leq F_X(y_1, \ldots, y_n)$, tj. F_X je neklesající v každé proměnné;

– pro každé $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n$ je $\lim_{x \to -\infty} F_X(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n) = 0$ a $\lim_{x_1, \dots, x_n \to \infty} F_X(x_1, \dots, x_n) = 1$.

– F je zprava spojitá v každé proměnné.

Analogií vztahu $P(a < X \leq b) = F(b) - F(a)$ pro (jednorozměrné) náhodné veličiny je v případě dvourozměrné náhodné veličiny vztah

$$P(a_1 < X_1 \leq b_1, a_2 < X_1 \leq b_2) = F(b_1, b_2) - F(b_1, a_2) - F(a_1, b_2) + F(a_1, a_2),$$

který se snadno odvodí.

4.4.2 Diskrétní náhodné vektory

Náhodný vektor $X = \langle X_1, \dots, X_n \rangle$ se nazývá *diskrétní* (nebo: X má *rozdělení diskrétního typu*), pokud nabývá jen konečně mnoha hodnot (hodnotami náhodného vektoru jsou $n$tice $\langle x_1, \dots, x_n \rangle$ reálných čísel).

Hodnoty diskrétního náhodného vektoru X, tj. $n$tice $x_i = \langle x_{i1}, \dots, x_{in} \rangle$, lze uspořádat do posloupnosti $x_1, x_2, \dots$.[7]

Platí tedy

$$\sum_{x_1 \in \mathbb{R}} \cdots \sum_{x_n \in \mathbb{R}} P(X_1 = x_1, \dots, X_n = x_n) = 1.$$

Funkce $p_X : \mathbb{R}^n \to \mathbb{R}$ definovaná předpisem

$$p_X(x_1, \dots, x_n) = P(X_1 = x_1, \dots, X_n = x_n)$$

se nazývá *pravděpodobnostní funkce náhodného vektoru* $X = \langle X_1, \dots, X_n \rangle$. Někdy se také nazývá *sdružená pravděpodobnostní funkce náhodných veličin* $X_1, \dots, X_n$.

Pro distribuční funkci diskrétního náhodného vektoru zřejmě platí

$$F_X(x_1, \dots, x_n) = \sum_{t_1 \leq x_1} \cdots \sum_{t_n \leq x_n} P(X_1 = t_1, \dots, X_n = t_n). \qquad (33)$$

Příklad 4.17. Pravděpodobnostní funkci náhodného vektoru $X = \langle X_1, X_2 \rangle$ lze uspořádat do tabulky. Například tabulka

[7] Symbol x_i podle kontextu označuje reálné číslo nebo vektor reálných čísel. Ve druhém případě označuje x_{ij} $j$tou souřadnici vektoru x_i.

$x_1 \backslash x_2$	0	1	2
0	0.35	0.25	0.06
1	0.18	0.12	0.04

popisuje vektor, který nabývá hodnot $\langle 0,0\rangle, \ldots \langle 1,2\rangle$ tj. X_1 nabývá hodnot 0 a 1 a X_2 hodnot 0, 1 a 2. Pro p_X platí $P(X_1 = 0, X_2 = 0) = 0.35$, $P(X_1 = 0, X_2 = 1) = 0.25$, $P(X_1 = 0, X_2 = 2) = 0.06$, $P(X_1 = 1, X_2 = 0) = 0.18$, $P(X_1 = 1, X_2 = 1) = 0.12$, $P(X_1 = 1, X_2 = 2) = 0.04$. Z této pravděpodobnostní funkce lze zkonstruovat distribuřní funkci F_X dle (33). Platí tedy například

$$\begin{aligned} F_X(1,1) &= P(X_1 = 0, X_2 = 0) + P(X_1 = 0, X_2 = 1) \\ &\quad + P(X_1 = 1, X_2 = 0) + P(X_1 = 1, X_2 = 1) = \\ &= 0.35 + 0.25 + 0.18 + 0.12 = 0.9. \end{aligned}$$

Výše uvedenou pravděpodobnostní funkci má například náhodný vektor sestávající z náhodných veličin X_1 a X_2 definovaných na pravděpodobnostním prostoru, ve kterém $\Omega = \{\omega_1, \ldots, \omega_{100}\}$, pro $i = 1, \ldots, 35$ je $X_1(\omega_i) = 0$ a $X_2(\omega_i) = 0$, pro $i = 36, \ldots, 60$ je $X_1(\omega_i) = 0$ a $X_2(\omega_i) = 1$, pro $i = 61, \ldots, 66$ je $X_1(\omega_i) = 0$ a $X_2(\omega_i) = 2$, pro $i = 67, \ldots, 84$ je $X_1(\omega_i) = 1$ a $X_2(\omega_i) = 0$, pro $i = 85, \ldots, 96$ je $X_1(\omega_i) = 1$ a $X_2(\omega_i) = 1$ a pro $i = 97, \ldots, 100$ je $X_1(\omega_i) = 1$ a $X_2(\omega_i) = 2$.

4.4.3 Spojité náhodné vektory

Náhodný vektor $X = \langle X_1, \ldots, X_n\rangle$ se nazývá *spojitý* (nebo: X má *rozdělení spojitého typu*), pokud existuje nezáporná reálná funkce $f : \mathbb{R}^n \to \mathbb{R}$ tak, že platí

$$F_X(x_1, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \ldots, t_n) \mathrm{d}t_1 \cdots \mathrm{d}t_n.$$

Funkce f se pak nazývá *hustota pravděpodobnosti* (nebo jen *hustota*) *náhodného vektoru* X, popřípadě *sdružená hustota pravděpodobnosti náhodných veličin* $X_1, \ldots, X_n$.

Platí analogie vlastností spojitých náhodných veličin:

$$P(a_1 < X_1 \leq b_1, \ldots, a_n < X_n \leq b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(t_1, \ldots, t_n) \mathrm{d}t_1 \cdots \mathrm{d}t_n.$$

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(t_1, \ldots, t_n) \mathrm{d}t_1 \cdots \mathrm{d}t_n = 1.$$

Příklad 4.18. Uvažujme spojitý náhodný vektor $X = \langle X_1, X_2 \rangle$, jehož hustota je dána vztahem

$$f(x_1, x_2) = \begin{cases} e^{-x_1-x_2} & \text{pro } x_1 > 0 \text{ a } x_2 > 0, \\ 0 & \text{jinak.} \end{cases}$$

Distribuční funkce je tedy dána vztahem

$$F(x_1, x_2) = \begin{cases} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} e^{-t_1-t_2} \mathrm{d}t_1 \mathrm{d}t_2 = (1 - e^{-x_1})(1 - e^{-x_2}) & \text{pro } x_1 > 0 \text{ a } x_2 > 0, \\ 0 & \text{jinak.} \end{cases}$$

4.4.4 Marginální rozdělení

Při práci s náhodnými veličinami se často vyskytuje situace, kdy známe distribuční funkci F (popř. pravděpodobnostní funkci p_X nebo hustotu pravděpodobnosti f_X) náhodného vektoru X a neznáme přímo distribuční funkci (popř. pravděpodobnostní funkci nebo hustotu) jednotlivých náhodných veličin nebo skupin náhodných veličin, ze kterých vektor X sestává. Je-li např. $X = \langle X_1, X_2, X_3 \rangle$ a známe-li pravděpodobnostní funkci p_X náhodného vektoru X, může nás zajímat pravděpodobnostní funkce p_{X_1} náhodné veličiny X_1 nebo sdružená pravděpodobnostní funkce $p_{\langle X_1, X_3 \rangle}$ náhodných veličin X_1 a X_3.

Začněme případem náhodného vektoru $X = \langle X_1, X_2 \rangle$ s distribuční funkcí $F_X(x_1, x_2)$. Pak *marginální distribuční funkce veličiny* X_1 je funkce $F_1 : \mathbb{R} \to \mathbb{R}$ definovaná předpisem

$$F_1(x_1) = \lim_{x_2 \to \infty} F_X(x_1, x_2).$$

Protože platí

$$\lim_{x_2 \to \infty} F_X(x_1, x_2) = \lim_{x_2 \to \infty} P(X_1 \le x_1, X_2 \le x_2) = P(X_1 \le x_1) = F_{X_1}(x_1), \tag{34}$$

je

$$F_1(x_1) = F_{X_1}(x_1),$$

tj. pro daný dáhodný vektor $X = \langle X_1, X_2 \rangle$ je marginální distribuční funkce veličiny X_1 je rovna distribuční funkci (samostatné) veličiny X_1. Proto lze F_1 a F_{X_1} navzájem zaměňovat.

Zdůdovněme (34). Označíme-li pro systém jevů $\{A_x \mid x \in \mathbb{R}\}$, který splňuje, že $x \leq y$ implikuje $A_x \subseteq A_y$, symbolem $\lim_{x\to\infty} A_x$ jev $\bigcup_{x\in\mathbb{R}} A_x$ a dále symbolem $[X_1 \leq x_1]$ jev $\{\omega \in \Omega \mid X_1(\omega) \leq x_1\}$ (apod. pro další jevy), platí

$$\begin{aligned}
& \lim_{x_2\to\infty} F_X(x_1, x_2) = \lim_{x_2\to\infty} P(X_1 \leq x_1, X_2 \leq x_2) = \\
= \; & \lim_{x_2\to\infty} P([X_1 \leq x_1] \cap [X_2 \leq x_2]) == P(\lim_{x_2\to\infty}([X_1 \leq x_1] \cap [X_2 \leq x_2])) = \\
= \; & P(\bigcup_{x_2\in\mathbb{R}}([X_1 \leq x_1] \cap [X_2 \leq x_2])) = P([X_1 \leq x_1] \cap \bigcup_{x_2\in\mathbb{R}}[X_2 \leq x_2]) = \\
= \; & P([X_1 \leq x_1] \cap \Omega) = P(X_1 \leq x_1) = F_{X_1}(x_1).
\end{aligned}$$

Analogické vlastnosti má marginální distribuční funkce F_2 veličiny X_2. Ta je definována vztahem $F_2(x_2) = \lim_{x_1\to\infty} F_X(x_1, x_2)$ a platí pro ni $F_2 = F_{X_2}$.

Je-li $X = \langle X_1, X_2 \rangle$ diskrétní, platí

$$P(X_1 = x_1) = \sum_{x_2} P(X_1 = x_1, X_2 = x_2) \text{ neboli } p_{X_1}(x_1) = \sum_{x_2} p_X(x_1, x_2),$$

kde součet probíhá přes x_2, pro které $P(X_1 = x_1, X_2 = x_2) > 0$ (takových x_2 je konečně mnoho). To plyne z faktu, že jev $[X_1 = x_1] = \bigcup_{x_2}[X_1 = x_1, X_2 = x_2]$ a že jevy $[X_1 = x_1, X_2 = x_2]$ jsou vzájemně disjunktní. Podobně platí $P(X_2 = x_2) = \sum_{x_1} P(X_1 = x_1, X_2 = x_2)$.

Příklad 4.19. Uvažujme náhodný vektor z příkladu 4.17. Tabulku popisující jeho pravděpodobnostní funkci můžeme rozšířit o řádek a sloupec popisující pravděpodobnostní funkce marginálních veličin X_1 a X_2, tj. funkce $p_{X_1}(x_1) = P(X_1 = x_1)$ a $p_{X_2}(x_2) = P(X_2 = x_2)$:

$x_1 \backslash x_2$	0	1	2	$P(X_1 = x_1)$
0	0.35	0.25	0.06	0.66
1	0.18	0.12	0.04	0.34
$P(X_2 = x_2)$	0.53	0.37	0.1	1

Platí totiž např. $P(X_1 = 0) = P(X_1 = 0, X_2 = 0) + P(X_1 = 0, X_2 = 1) + P(X_1 = 0, X_2 = 2)$, $P(X_2 = 1) = P(X_1 = 0, X_2 = 1) + P(X_1 = 1, X_2 = 1)$ apod.

Příklad 4.19 ilustruje, proč se veličiny X_1 a X_2 náhodného vektoru $X = \langle X_1, X_2 \rangle$ nazývají marginální: při grafické znázornění informací o pravděpodobnostním rozdělení vektoru X (např. informace o pravděpodobnostní funkci) vzniká informace o X_1 a X_2 na okraji (angl. margin).

Podobné vztahy platí pro marginální rozdělení spojitého náhodného vektoru $X = \langle X_1, X_2 \rangle$. Je-li $f_X(x_1, x_2)$ jeho hustota pravděpodobnosti, platí pro hustoty f_1 a f_2 marginálních veličin X_1 a X_2 vztahy:

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} f(x_1, x_2)\mathrm{d}x_2, \\ f_2(x_2) &= \int_{-\infty}^{\infty} f(x_1, x_2)\mathrm{d}x_1, \end{aligned}$$

pro každé $x_1 \in \mathbb{R}$ a $x_2 \in \mathbb{R}$.

Příklad 4.20. Uvažujme spojitý náhodný vektor $X = \langle X_1, X_2 \rangle$ z příkladu 4.18. Pro marginální distribuční funkce veličin X_1 a X_2 platí

$$\begin{aligned} F_1(x_1) &= \lim_{x_2 \to \infty} F(x_1, x_2) = \begin{cases} 1 - e^{-x_1} & \text{pro } x_1 > 0, \\ 0 & \text{pro } x_1 \leq 0, \end{cases} \\ F_2(x_2) &= \lim_{x_1 \to \infty} F(x_1, x_2) = \begin{cases} 1 - e^{-x_2} & \text{pro } x_2 > 0, \\ 0 & \text{pro } x_2 \leq 0. \end{cases} \end{aligned}$$

Pro marginální hustoty pak platí

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} f(x_1, x_2)\mathrm{d}x_2 = \begin{cases} \int_0^{\infty} f(x_1, x_2)\mathrm{d}x_2 = e^{-x_1} & \text{pro } x_1 > 0, \\ 0 & \text{pro } x_1 \leq 0, \end{cases} \\ f_2(x_2) &= \int_{-\infty}^{\infty} f(x_1, x_2)\mathrm{d}x_1 = \begin{cases} \int_0^{\infty} f(x_1, x_2)\mathrm{d}x_1 = e^{-x_2} & \text{pro } x_2 > 0, \\ 0 & \text{pro } x_2 \leq 0. \end{cases} \end{aligned}$$

Sestává-li náhodný vektor z více než dvou veličin, tj. obecně $X = \langle X_1, \ldots, X_n \rangle$, zavádí se přirozeným způsobem, který zobecňuje výše uvedené vztahy, marginální rozdělení (marginální distribuční funkce, marginální pravděpodobnostní funkce, marginální hustota) pro jakoukoli podmnožinu $\{j_1, \ldots, j_k\} \subseteq \{1, \ldots, n\}$. Uvažujme dva příklady. V prvním je $k = 1$ a $j_1 = 1$, tj. zajímá nás marginální rozdělení veličiny X_1; ve druhém je $k = 2$, $j_1 = 1$ a $j_2 = n$, tj. zajímá nás marginální rozdělení dvojice $\langle X_1, X_n \rangle$ (tato dvojice je sama náhodným vektorem). Definujeme pak:

$$\begin{aligned} F_1(x_1) &= \lim_{x_2 \to \infty, \ldots, x_n \to \infty} F(x_1, x_2, \ldots, x_n), \text{ tj. } = F(x_1, \infty, \ldots, \infty), \\ F_{1,n}(x_1, x_n) &= \lim_{x_2 \to \infty, \ldots, x_{n-1} \to \infty} F(x_1, x_2, \ldots, x_n), \text{ tj. } = F(x_1, \infty, \ldots, \infty, x_n). \end{aligned}$$

Je spojitém případě pro marginální hustotu f_1 a marginální sdruženou hustotu $f_{1,n}$ je

$$\begin{aligned} f_1(x_1) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) \mathrm{d}x_2 \cdots \mathrm{d}x_n, \\ f_{1,n}(x_1, x_n) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n) \mathrm{d}x_2 \cdots \mathrm{d}x_{n-1} \end{aligned}$$

V diskrétním případě pro marginální pravděpodobnostní funkci $P(X_1 = x_1)$ a sdruženou pravděpodobnostní funkci $P(X_1 = x_1, X_n = x_n)$ platí

$$\begin{aligned} P(X_1 = x_1) &= \sum_{x_2} \cdots \sum_{x_n} P(X_1 = x_1, \ldots, X_n = x_n), \\ P(X_1 = x_1, X_n = x_n) &= \sum_{x_2} \cdots \sum_{x_{n-1}} P(X_1 = x_1, \ldots, X_n = x_n). \end{aligned}$$

Obecné vztahy pro libovolnou podmnožinu $\{j_1, \ldots, j_k\} \subseteq \{1, \ldots, n\}$ si každý snadno odvodí sám.

4.4.5 Charakteristiky náhodných vektorů

Vektor středních hodnot $\langle E(X_1), \ldots, E(X_n) \rangle$ se nazývá *střední hodnota náhodného vektoru* X (popř. vektor středních hodnot náhodného vektoru X).

Nechť $X = \langle X_1, \ldots, X_n \rangle$ je náhodný vektor a $h : \mathbb{R}^n \to \mathbb{R}$ spojitá (obecněji: borelovsky měřitelná) funkce. Pak, jako v podobném případě náhodné veličiny (str. 51) je složená funkce $h \circ \langle X_1, \ldots, X_n \rangle$ náhodnou veličinou. Tato veličina, značená také $h(X_1, \ldots, X_n)$, je definovaná takto: $[h(X_1, \ldots, X_n)](\omega) = h(X_1(\omega), \ldots, X_n(\omega))$.

Pro střední hodnotu této náhodné veličiny platí, podobně jako v jednorozměrném případě, že pro diskrétní a spojitý náhodný vektor je

$$\begin{aligned} E(h(X_1, \ldots, X_n)) &= \sum_{x_1} \cdots \sum_{x_n} h(x_1, \ldots, x_n) P(X_1 = x_1, \ldots, X_n = x_n), \\ E(h(X_1, \ldots, X_n)) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(x_1, \ldots, x_n) f(x_1, \ldots, x_n) \mathrm{d}x_1 \cdots \mathrm{d}x_n. \end{aligned}$$

Můžeme tedy určovat i další charakteristiky náhodné veličiny $h(X_1, \ldots, X_n)$. Tento případ zahrnuje i dříve uváděné charakteristiky náhodných veličin, a to v následujícím smyslu. Např. pro $h(X_1, \ldots, X_n) = X_i^r$ lze snadno odvodit, že $E(h(X_1, \ldots, X_n)) = E(X_i^r)$ je právě r. obecný moment náhodné veličiny X_i.

Kromě charakteristik jednotlivých veličin X_i náhodného vektoru X jsou významné charakteristiky dvojic $\langle X_i, X_j \rangle$.

Kovariance náhodných veličin X_i a X_j, je definována vztahem

$$\operatorname{cov}(X_i, X_j) = E([X_i - E(X_i)][X_j - E(X_j)]).$$

Jde tedy o střední hodnotu náhodné veličiny $h(X_1, \ldots, X_n) = [X_i - E(X_i)][X_j - E(X_j)]$. Místo $\operatorname{cov}(X_i, X_j)$ píšeme také jen σ_{ij}. Matice

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$$

se nazývá *kovarianční matice náhodného vektoru* $X = \langle X_1, \ldots, X_n \rangle$ (někdy varianční matice, angl. covariance matrix, dispersion matrix nebo variance–covariance matrix).

Je zřejmé, že kovarianční matice je symetrická, tj. $\sigma_{ij} = \sigma_{ji}$. Dále platí

$$\sigma_{ii} = \operatorname{cov}(X_i, X_i) = E([X_i - E(X_i)]^2) = \operatorname{var}(X_i).$$

na diagonále kovarianční matice jsou tedy rozptyly jednotlivých náhodných veličin.

Příklad 4.21. Sestrojte kovarianční matici náhodného vektoru z příkladu 4.19.

Musíme určit $\sigma_{ij} = \operatorname{cov}(X_i, X_j)$ pro $i, j = 1, 2$, přitom víme, že $\sigma_{ij} = \sigma_{ji}$ a $\sigma_{ii} = \operatorname{var}(X_i)$.

Označme $\mu_i = E(X_i)$. Z marginálních pravděpodobnostních funkcí $p_{X_1}(x_1) = P(X_1 = x_1)$ a $p_{X_1}(x_1) = P(X_1 = x_1)$, jejichž hodnoty jsme spočítali v tabulce v příkladu 4.19 dostáváme

$$\begin{aligned} \mu_1 &= E(X_1) = 0 \cdot 0.66 + 1 \cdot 0.34 = 0.34, \\ \mu_2 &= E(X_2) = 0 \cdot 0.53 + 1 \cdot 0.37 + 2 \cdot 0.1 = 0.57. \end{aligned}$$

Proto je

$$\begin{aligned}
\sigma_{11} &= \mathrm{var}(X_1) = E([X_1-\mu_1]^2) = \sum_{x_1}(x_1-\mu_1)^2\cdot p_{X_1}(x_1) = \\
&= (0-0.34)^2\cdot 0.66 + (1-0.34)^2\cdot 0.34 = 0.2244 \\
\sigma_{22} &= \mathrm{var}(X_2) = E([X_2-\mu_2]^2) = \sum_{x_2}(x_2-\mu_2)^2\cdot p_{X_1}(x_1) = \\
&= (0-0.57)^2\cdot 0.53 + (1-0.57)^2\cdot 0.37 + (2-0.57)^2\cdot 0.1 = 0.4451.
\end{aligned}$$

Konečně

$$\begin{aligned}
\sigma_{12} &= E([X_1-\mu_1][X_2-\mu_2]) = \sum_{x_1,x_2}(x_1-\mu_1)\cdot(x_2-\mu_2)\cdot p_X(x_1,x_2) = \\
&= (0-0.34)(0-0.57)0.35 + (0-0.34)(1-0.57)0.25 + \\
&\quad (0-0.34)(2-0.57)0.06 + (1-0.34)(0-0.57)0.18 + \\
&\quad (1-0.34)(1-0.57)0.12 + (1-0.34)(2-0.57)0.04 = 0.0062.
\end{aligned}$$

Kovarianční matice je tedy

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} 0.2244 & 0.0062 \\ 0.0062 & 0.4451 \end{pmatrix}$$

Pokud $\mathrm{var}(X_i) > 0$ a $\mathrm{var}(X_j) > 0$, nazývá se číslo

$$\rho(X_i, X_j) = \frac{\mathrm{cov}(X_i, X_j)}{\sqrt{\mathrm{var}(X_i)\mathrm{var}(X_j)}}$$

korelační koeficient náhodných veličin X_i a X_j. Někdy se označuje $\mathrm{corr}(X_i, X_j)$.

Lze dokázat, že

$$-1 \leq \rho(X_i, X_j) \leq 1.$$

Korelační koeficient vyjadřuje míru lineární závislosti náhodných veličin. Předpokládejme totiž, že $X_j = aX_i + b$, kde $0 \neq a, b \in \mathbb{R}$. Pak

$$\mathrm{cov}(X_i, X_j) = E([X_i - E(X_i)][aX_i + b - E(aX_i + b)] = \cdots = a\mathrm{var}(X_i),$$

a tedy

$$\rho(X_i, X_j) = \frac{a\mathrm{var}(X_i)}{\sqrt{a^2\mathrm{var}(X_i)^2}} = \begin{cases} 1 & \text{pro } a > 0, \\ -1 & \text{pro } a < 0. \end{cases}$$

Matice $\mathrm{corr}(X)$ sestavená z korelačních koeficientů, tj.

$$\mathrm{corr}(X)_{ij} = \rho(X_i, X_j),$$

se nazývá *korelační matice náhodného vektoru* X. Protože $\rho(X_i, X_i) = 1$, má tato matice na diagonále jedničky.

Příklad 4.22. Sestrojte korelační matici náhodného vektoru z příkladu 4.19.

Vzhledem k příkladu 4.21 je

$$\rho(X_1, X_2) = \frac{\mathrm{cov}(X_1, X_2)}{\sqrt{\mathrm{var}(X_1)\mathrm{var}(X_2)}} = \frac{0.0062}{\sqrt{0.2244 \cdot 0.4451}} \approx 0.0196.$$

Vztah (korelace) mezi X_1 a X_2 je tedy velmi slabý, což je vidět i z tabulky popisující sdruženou pravděpodobnostní funkci veličin X_1 a X_2. Korelační matice je tedy

$$\mathrm{corr}(X) = \begin{pmatrix} \rho(X_1, X_1) & \rho(X_1, X_2) \\ \rho(X_2, X_1) & \rho(X_2, X_2) \end{pmatrix} = \begin{pmatrix} 1 & 0.0196 \\ 0.0196 & 1 \end{pmatrix}.$$

4.4.6 Podmíněné rozdělení a nezávislost náhodných veličin

Podmíněné rozdělení

Uvažujme náhodný vektor $X = \langle X_1, X_2 \rangle$.

Je-li X diskrétní a má-li sdruženou pravděpodobnostní funkci $p_X(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$, je *podmíněná pravděpodobnostní funkce náhodné veličiny X_1 za podmínky, že veličina X_2 má hodnotu x_2*, funkce $P(X_1 = x_1 | X_2 = x_2)$, označovaná také $p_X(x_1|x_2)$ (je to funkce jedné proměnné, totiž x_1; x_2 je pevná hodnota), která je pro $P(X_2 = x_2) > 0$ definována vztahem

$$P(X_1 = x_1 | X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)}$$

Zde je $P(X_2 = x_2)$ marginální pravděpodobnostní funkce veličiny X_2.

Uvědomme si, že $P(X_1 = x_1 | X_2 = x_2)$ je právě podmíněná pravděpodobnost $P(A|B)$, kde A je jev $[X_1 = x_1] = \{\omega \mid X_1(\omega) = x_1\}$ a B je jev $[X_2 = x_2] = \{\omega \mid X_2(\omega) = x_2\}$.

Definiční vztah lze přepsat (podobně jako u podmíněné pravděpodobnosti): pokud $P(X_2 = x_2) > 0$, pak $P(X_1 = x_1, X_2 = x_2) = P(X_2 = x_2)P(X_1 =$

$x_1|X_2 = x_2)$; pokud $P(X_1 = x_1) > 0$, pak $P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2|X_1 = x_1)$.

Pokud $P(X_1 = x_1) > 0$ a $P(X_2 = x_2) > 0$, je tedy

$$P(X_2 = x_2|X_1 = x_1) = \frac{P(X_1 = x_1|X_2 = x_2)P(X_2 = x_2)}{P(X_1 = x_1)}, \tag{35}$$

přičemž

$$P(X_1 = x_1) = \sum_{y:P(X_2=y)>0} P(X_1 = x_1|X_2 = y)P(X_2 = y).$$

Vzorec (35) se nazývá *Bayesův vzorec pro diskrétní náhodné veličiny.*

Příklad 4.23. Určete hodnoty podmíněné pravděpodobnostní funkce $p_X(x_1|x_2) = P(X_1 = x_1|X_2 = y)$ náhodného vektoru z příkladu 4.19.

Určeme hodnoty $P(X_1 = x|X_2 = y)$ pro $y = 0$, 1 a 2. Dle vzorce je

$$P(X_1 = x_1|X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_2 = x_2)} \text{ neboli } p_X(x_1|x_2) = \frac{p_X(x_1, x_2)}{p_{X_2}(x_2)}.$$

Hodnoty $p_X(x_1, x_2) = P(X_1 = x_1, X_2 = x_2)$ sdružené pravděpodobnostní funkce jsou přímo dány tabulkou; hodnoty $p_{X_2}(x_2) = P(X_2 = x_2)$ marginální pravděpodobnostní funkce veličiny X_2 jsme spočítali v příkladu 4.19.

$y = 0$:

$$p_X(0|0) = \frac{p_X(0,0)}{p_{X_2}(0)} = \frac{0.35}{0.53}, \quad p_X(1|0) = \frac{p_X(1,0)}{p_{X_2}(0)} = \frac{0.18}{0.53};$$

$y = 1$:

$$p_X(0|1) = \frac{p_X(0,1)}{p_{X_2}(1)} = \frac{0.25}{0.37}, \quad p_X(1|1) = \frac{p_X(1,1)}{p_{X_2}(1)} = \frac{0.12}{0.37};$$

$y = 2$:

$$p_X(0|2) = \frac{p_X(0,2)}{p_{X_2}(2)} = \frac{0.06}{0.1}, \quad p_X(1|2) = \frac{p_X(1,2)}{p_{X_2}(2)} = \frac{0.04}{0.1}.$$

Všimněme si, že pro každou hodnotu y je funkce $p_X(x|y)$ funkce jedné proměnné, x, a že tato funkce je pravděpodobnostní funkcí na množině hodnot veličiny X_1, tj. na $\{0, 1\}$.

Je-li vektor $X = \langle X_1, X_2 \rangle$ spojitý a má-li sdruženou hustotu $f(x_1, x_2)$, je *podmíněná hustota pravděpodobnosti náhodné veličiny X_1 za podmínky, že X_2 má hodnotu x_2*, funkce $f(x_1|x_2)$ (je to opět funkce jedné proměnné, x_1; x_2 je pevná hodnota), která je pro $f_2(x_2) > 0$ definována vztahem

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)}.$$

Zde je $f_2(x_2)$ marginální hustota pravděpodobnosti veličiny X_2.

Definiční vztah lze opět přepsat: pokud $f_2(x_2) > 0$, pak $f(x_1, x_2) = f_2(x_2)f(x_1|x_2)$; pokud $f_1(x_1) > 0$, pak $f(x_2, x_1) = f_1(x_1)f(x_2|x_1)$.

Pokud $f_1(x_1) > 0$ a $f_2(x_2) > 0$, je tedy

$$f(x_2|x_1) = \frac{f(x_1|x_2)f_2(x_2)}{f_1(x_1)}, \tag{36}$$

přičemž

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\mathrm{d}x_2 = \int_{x_2 : f_2(x_2) > 0} f(x_1|x_2)f_2(x_2)\mathrm{d}x_2.$$

Vzorec (36) se nazývá *Bayesův vzorec pro spojité náhodné veličiny.*

Příklad 4.24. Uvažujme spojitý náhodný vektor $X = \langle X_1, X_2 \rangle$ z příkladu 4.20. V uvedeném příkladu jsme spočítali, že

$$f_2(x_2) = \begin{cases} \int_0^\infty f(x_1, x_2)\mathrm{d}x_1 = e^{-x_2} & \text{pro } x_2 > 0, \\ 0 & \text{pro } x_2 \leq 0. \end{cases}$$

Podmíněná hustota $f(x_1|x_2)$ je tedy definovaná pro $x_2 > 0$ a platí

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{e^{-x_1 - x_2}}{e^{-x_2}} = e^{-x_1}$$

pro $x_1 > 0$ a $f(x_1|x_2) = 0$ pro $x_1 \leq 0$.

Podmíněné charakteristiky

Charakteristiky, jako je střední hodnota nebo rozptyl, podmíněných rozdělení (tj. podmíněné pravděpodobnostní funkce a podmíněné hustoty) se nazývají podmíněné charakteristiky. *Podmíněná střední hodnota* náhodné veličiny X_1

za podmínky, že X_2 má hodnotu x_2 je v diskrétním případě tedy definována vztahem

$$E(X_1|X_2 = x_2) = \sum_{x_1} x_1 P(X_1 = x_1|X_2 = x_2)$$

a ve spojitém případě vztahem

$$E(X_1|X_2 = x_2) = \int_{-\infty}^{\infty} x_1 f(x_1|x_2)\mathrm{d}x_1.$$

Podmíněnou střední hodnotu lze nahlížet jako funkci proměnné x_2.

Příklad 4.25. Určete střední hodnoty $E(X_1|X_2 = x_2)$ pro náhodné veličiny z příkladu 4.19.

Je

$$E(X_1|X_2 = x_2) = 0 \cdot p_X(0|x_2) + 1 \cdot p_X(1|x_2),$$

a tedy s využitím hodnot $p_X(x_1|x_2)$ z příkladu 4.23 je

$$\begin{aligned}
E(X_1|X_2 = 0) &= 0 \cdot p_X(0|0) + 1 \cdot p_X(1|0) = 0 \cdot \frac{0.35}{0.53} + 1 \cdot \frac{0.18}{0.53} = 0.3396, \\
E(X_1|X_2 = 1) &= 0 \cdot p_X(0|1) + 1 \cdot p_X(1|1) = 0 \cdot \frac{0.25}{0.37} + 1 \cdot \frac{0.12}{0.37} = 0.3243, \\
E(X_1|X_2 = 2) &= 0 \cdot p_X(0|2) + 1 \cdot p_X(1|2) = 0 \cdot \frac{0.06}{0.1} + 1 \cdot \frac{0.04}{0.1} = 0.4.
\end{aligned}$$

Podmíněný rozptyl náhodné veličiny X_1 za podmínky, že X_2 má hodnotu x_2 je definován vztahem

$$\mathrm{var}(X_1|X_2 = x_2) = E([X_1 - E(X_1|X_2 = x_2)]^2|X_2 = x_2)$$

a platí pro něj

$$\mathrm{var}(X_1|X_2 = x_2) = E(X_1^2|X_2 = x_2) - E^2(X_1|X_2 = x_2),$$

tj. vztah analogický vztahu pro nepodmíněný rozptyl.

Nezávislost náhodných veličin

Uvažujme opět dvě náhodné veličiny, tj. náhodný vektor $X = \langle X_1, X_2 \rangle$. Nechť má sdruženou distribuční funkci $F(x_1, x_2)$ a marginální distribuční funkce $F_1(x_1)$ a $F_2(x_2)$. Náhodné veličiny X_1 a X_2 jsou *nezávislé*, pokud

$$F(x_1, x_2) = F_1(x_1)F_2(x_2) \tag{37}$$

pro každá $x_1, x_2 \in \mathbf{R}$. Uvědomme si, že uvedený vztah říká, že

$$P(X_1 \leq x_1, X_2 \leq x_2) = P(X_1 \leq x_1)P(X_2 \leq x_2),$$

tedy, že jevy $[X_1 \leq x_1]$ a $[X_2 \leq x_2]$ jsou nezávislé dle definice nezávislosti jevů. Lze ukázat, že pro nezávislé náhodné veličiny platí také

$$P(a_1 < X_1 \leq b_1, a_2 < X_2 \leq b_2) = P(a_1 < X_1 \leq b_1)P(a_2 < X_1 \leq b_2),$$

a že, obecněji, pro libovolné borelovské množiny $A_1, A_2 \subseteq \mathbb{R}$ je

$$P(X_1 \in A_1, X_2 \in A_2) = P(X_1 \in A_1)P(X_2 \in A_2).$$

V diskrétním případě je podmínka nezávislosti (37) ekvivalentní tomu, že pro pravděpodobnostní funkce platí

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2),$$

pro každá $x_1, x_2 \in \mathbf{R}$. Z výše uvedených vztahů vyplývá, že pokud $P(X_2 = x_2) > 0$, pro nezávislé veličiny X_1 a X_2 platí

$$P(X_1 = x_1 | X_2 = x_2) = P(X_1 = x_1);$$

pokud $P(X_1 = x_1) > 0$, pro nezávislé veličiny X_1 a X_2 platí

$$P(X_2 = x_2 | X_1 = x_1) = P(X_2 = x_2).$$

Ve spojitém případě je nezávislost náhodných veličin ekvivalentní tomu, že pro hustoty platí

$$f(x_1, x_2) = f_1(x_1)f_2(x_2)$$

pro každá $x_1, x_2 \in \mathbf{R}$. Pokud $f_2(x_2) > 0$, pro nezávislé veličiny X_1 a X_2 platí

$$f(x_1|x_2) = f_1(x_1);$$

pokud $f_1(x_1) > 0$, pro nezávislé veličiny X_1 a X_2 platí

$$f(x_2|x_1) = f_2(x_2).$$

4.5 Transformace náhodných veličin

Na str. 51, 69 a 81 jsme se setkali s funkcemi náhodných veličin. Víme, že za jistých předpokladů vznikne složením náhodné veličiny X a funkce h náhodná veličina, $h(X)$. Toho jsme s využili při úvahách o charakteristikách náhodných veličin. Například r. obecný moment veličiny X je roven střední hodnotě transformované veličiny $Y = h(X)$ pro funkci $h(x) = x^r$. Transformace náhodných veličin jsou významné i z jiných než výše uvedených důvodů. Představme si, že daný systém (např. daný algoritmus nebo daný technický systém) realizuje vstupně-výstupní chování, které je popsáno funkcí h, tj. vstupní hodnotu x převede na výstupní hodnotu $y = h(x)$. Známe-li pravděpodobnostní rozdělení vstupů, tj. rozdělení náhodné veličiny X, přirozeně nás zajímá rozdělení výstupů, tj. rozdělení náhodné veličiny Y. Jiný případ: Předpokládejme, že umíme generovat náhodná čísla, tj. např. čísla z intervalu $[0, 1]$ s rovnoměrným rozdělením, a že chceme generovat čísla z jiného intervalu s jiným rozdělením. Otázka je, zda lze využít generátor, který máme, zvolit vhodnou funkci h a generovaná náhodná čísla x převádět na $h(x)$ tak, aby čísla $h(x)$ měla požadované rozdělení.

Podíváme se nyní, jak lze v případě spojitých náhodných veličin ze znalosti hustoty veličiny X odvodit hustotu veličiny Y.

Věta 4.3. *Nechť X je spojitá náhodná veličina s hustotou f_X, která je nenulová na množině $I \subseteq \mathbb{R}$, tj. $f_X(x) > 0$, právě když $x \in I$; nechť $h : I \to \mathbb{R}$ je ryze monotónní (tj. rostoucí, nebo klesající) diferencovatelná funkce. Pak $Y = h(X)$ je spojitá náhodná veličina, jejíž hustota je dána vztahem*

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \cdot |(h^{-1})'(y)| & \text{pro } y \in h(I), \\ 0 & \text{jinak}, \end{cases} \tag{38}$$

kde h^{-1} je inverzní funkce k funkci h, $(h^{-1})'$ je její derivace a $|\cdot|$ je absolutní hodnota.

Důkaz. Že Y je spojitá náhodná veličina, plyne z lemma 4.1 (z toho, že h je diferencovatelná, tedy že á v kažém bodě z I derivaci, plyne, že je spojitá). Předpokládejme, že h je rostoucí. Pro distribuční funkci F_Y veličiny Y platí (třetí rovnost platí proto, že h je rostoucí)

$$F_Y(y) \;=\; P(Y \leq y) = P(h(X) \leq y) = P(X \leq h^{-1}(y)) = F_X(h^{-1}(y)).$$

Protože hustota je vždy derivací distribuční funkce, je $f_Y = F_Y'$, a tedy:

$f_Y(y) = F_Y'(y) = (F_X(h^{-1}))'(y) = F_X'(h^{-1}(y)) \cdot (h^{-1})'(y) = f_X(h^{-1}(y)) \cdot (h^{-1})'(y)$.

Pro klesající h je $F_Y(y) = P(h(X) \leq y) = 1 - P(h(X) \geq y) = 1 - P(X \leq h^{-1}(y)) = 1 - F_X(h^{-1}(y))$, a tedy $f_Y(y) = -F_X'(h^{-1}(y)) \cdot (h^{-1})'(y) = -f_X(h^{-1}(y)) \cdot (h^{-1})'(y)$. $\square$

Častým případem věty 4.3 je situace, kdy funkce h le lineární, tj. $h(x) = ax + b$ pro $a \neq 0$, tedy náhodná veličina Y je tvaru $aX + b$. Pak je

$$f_Y(y) = \begin{cases} \frac{1}{|a|} \cdot f_X(\frac{y-b}{a}) & \text{pro } y \in aI + b, \\ 0 & \text{jinak}, \end{cases} \tag{39}$$

neboť $h^{-1}(y) = \frac{y-b}{a}$ a $|(h^{-1})'(y)| = \frac{1}{|a|}$.

Příklad 4.26. Nechť X normální rozdělení s parametry μ a σ^2, tj. $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Pak náhodná veličina $Y = \frac{X-\mu}{\sigma}$ má normované normální rozdělení, tj. rozdělení $N(0,1)$ s hustotou $f_Y(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}$. Je totiž $Y = \frac{1}{\sigma} X + \frac{-\mu}{\sigma}$, tedy Y je lineární funkcí X a $a = \frac{1}{\sigma}$ a $b = -\frac{\mu}{\sigma}$. Hustota f_Y veličiny Y je tedy pro $y \in aI + b = (-\infty, \infty)$ dle (39) rovna

$$f_Y(y) = \sigma \cdot f_X(\sigma(y + \frac{\mu}{\sigma})) = \sigma \cdot \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(\sigma(y+\frac{\mu}{\sigma})-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}},$$

což je fakt, který jsme uvedli bez důkazu na str. 71.

Příklad 4.27. Nechť X normální rozdělení s parametry μ a σ^2, tj. $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$. Uvažujme náhodnou veličinu $Y = e^X$, tj.transformující funkce je $h(x) = e^x$. Je tedy $h^{-1}(y) = \ln(y)$ a $(h^{-1})'(y) = \frac{1}{y}$. Dle věty 4.3 má Y hustotu

$$f_Y(y) = \frac{f_X(\ln(y))}{y} = \begin{cases} \frac{1}{\sigma y\sqrt{2\pi}} e^{-\frac{(\ln y-\mu)^2}{2\sigma^2}} & \text{pro } y > 0, \\ 0 & \text{pro } x \leq 0. \end{cases}$$

Tedy Y má logaritmicko-normální rozdělení (viz str. 72).

Podívejme se nyní na důležitou transformační funkci.

Příklad 4.28. Nechť X má hustotu f_X a distribuční funkci F_X. Zvolme $h = F_X$, tj. $Y = F_X(X)$. Pak Y nabývá hodnot v intervalu $(0,1)$, popř. i 0 a 1, a tedy pro $y \in (0,1)$ je (srovnej s odvozením v důkazu Věty 4.3)

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(F_X(X) \le y) = P(F_X^{-1}(F_X(X)) \le F_X^{-1}(y)) = \\ &= P(X \le F_X^{-1}(y)) = F_X(F_X^{-1}(y)) = y, \end{aligned}$$

dále pak $F_Y(y) = 0$ pro $y < 0$ a $F_Y(y) = 1$ pro $y > 1$. Pro hustotu veličiny Y tedy platí

$$f_Y(y) = F_Y'(y) = \begin{cases} 1 & \text{pro } y \in (0,1), \\ 0 & \text{jinak,} \end{cases}$$

a tedy Y je náhodná proměnná s rovnoměrným rozdělením v intervalu $(0,1)$. Platí tedy, že pro libovolnou spojitou náhodnou proměnnou X má náhodná proměnná $Y = F_X(X)$ rovnoměrné rozdělení v intervalu $(0,1)$.

Úvahy v příkladu 4.28 lze použít pro generování náhodných čísel s požadovaným rozdělením pravděpodobnosti za předpokladu, že máme k dispozici generátor náhodných čísel s rovnoměrným rozdělením. O generátorech náhodných čísel pojednáme později (jsou k tomu třeba statistické pojmy a úvahy); zatím chápeme tento pojem intuitivně: je to zařízení, které v jednotlivých krocích $i = 1, 2, \dots$ generuje čísla $x_i \in \mathbb{R}$ tak, že množina vygenerovaných čísel má dané rozdělení. Předpokládejme, že máme generátor náhodných čísel y_i s rovnoměrným rozdělením v intervalu $(0,1)$. Chceme-li generovat náhodná čísla s požadovanou hustotou f_X a distribuční funkcí F_X, stačí dle příkladu 4.28 transformovat generovaná čísla y_i funkcí F_X^{-1}. Odpovídající hodnoty $x_i = F_X^{-1}(y_i)$ pak mají požadované rozdělení s hustotou f_X a distribuční funkcí F_X. Otázkou tedy zůstává, zda lze efektivně počítat hodnoty funkce F_X^{-1}. V některých případech, jako je exponenciální rozdělení, to lze, v jiných, jako je normální rozdělení, je třeba postupovat jinak.

Příklad 4.29. Chceme-li generovat náhodná čísla s exponenciálním rozdělením s parametry δ a $A = 0$, postupujeme podle právě uvedeného postupu takto. Generujeme čísla y_i s rovnoměrným rozdělením v intervalu $(0,1)$ a transformujeme je funkcí F^{-1}. Připomeňme, že distribuční funkce exponenciálního rozdělení je $F(x) = 1 - \frac{1}{\delta}e^{-\frac{x}{\delta}}$, která je rostoucí na intervalu $(0,\infty)$ a jejím oborem hodnot na tomto intervalu je $(0,1)$. Snadno odvodíme, že $F^{-1}(y) = -\delta \ln(1-y)$. Proto $x_i = F^{-1}(y_i)$ budou náhodná čísla s exponenciálním rozdělením.

5 Úvod do pravděpodobnostní analýzy algoritmů

5.1 Úvodní úvahy

Předmětem pravděpodobnostní analýzy algoritmů je analýza jejich chování, zejména jejich výpočetní složitosti. Typickou úlohou je určit pro daný algoritmus jeho časovou složitost v průměrném případě, tedy nikoli v nejhorším případě, jak se obvykle činí v úvodních kurzech o algoritmech.[8]

Časová složitost algoritmu v nejhorším případě je užitečnou informací. Víme-li např., že taková časová složitost je $5n^2 + 10$, víme, že pro žádný vstup velikosti n nebude výpočet trvat více než $5n^2+10$ výpočetních kroků. Mnohdy nás ale více zajímá, jak dlouho výpočet potrvá v průměrném případě, tj. očekávaná doba trvání, a nikoli největší možná doba trvání.

Intuitivně je zřejmé, že ta závisí na tom, jak pravděpodobné jsou jednotlivé vstupy algoritmu. Jsou-li velmi pravděpodobné vstupy, pro které je výpočet rychlý, může být přijatelný i algoritmus, jehož časová složitost v nejhoším případě je vysoká (vstup, pro který nejhorší případ nastává, totiž může mít malou pravděpodobnost).

Koncepční uchopení situace, které se při pravděpodobnostní analýze algoritmů používá, je následující.[9] Předpokládáme, že na množině vstupů velikosti n daného algoritmu je dáno pravděpodobnostní rozdělení. Přesněji, že množina vstupů velikosti n je množinou Ω_n elementrárních jevů nějakého pravděpodobnostního prostoru $\langle \Omega_n, \mathcal{B}_n, P_n \rangle$. Pak $P_n(\{\omega\})$ je pravděpodobnost, že vstupem algoritmu je ω. Počet instrukcí, které algoritmus vykoná pro vstup délky n lze chápat jako náhodnout veličinu X_n, a to v tom smyslu, že pro vstup $\omega \in \Omega_n$ je $X_n(\omega)$ počet kroků, které algoritmus pro vstup ω vykoná. Složitost algoritmu v průměrném případě lze pak chápat jako soubor těchto náhodných veličin X_n, $n = 0, 1, \dots$ a jejich středních hodnost $E(X_n)$, resp. přesněji jako funkci T, která velikosti n vstupu přiřadí střední hodnotu náhodné veličiny X_n, tj.

$$T(n) = E(X_n).$$

[8]Předpokládáme, že čtenář zná základní pojmy jako je časová (paměťová) složitost algoritmu.

[9]Poznamenejme také, že tzv. pravděpodobnostní algoritmy (randomized algorithms) představují něco jiného. Pravděpodobnostní algoritmy jsou algoritmy, které obsahují instrukce prováděné s jistou pravděpodobností. Např. v **if** C **then** B_1 **else** B_2 může podmínka C znít „náhodně vygenerované číslo je větší než 0.8“. V případě, který uvažujeme my, jsou algoritmy deterministické.

Max($A[0..n-1]$)
1 $j \leftarrow n-1$; $k \leftarrow n-2$; $max \leftarrow A[n-1]$
2 **while** $k \geq 0$ **do**
3 **if** $A[k] > max$ **then**
4 $j \leftarrow k$; $max \leftarrow A[k]$
5 $k \leftarrow k-1$
6 **return**(j)

Obrázek 1: Algoritmus MAX

5.2 Pravděpodobnostní analýza jednoduchého programu

Uvažujme následující jednoduchý problém. Vstupem je pole $A[0..n-1]$ navzájem různých čísel množiny $\{1, 2, \ldots, n\}$, tj.

$$\{A[0], A[1], \ldots, A[n-1]\} = \{1, 2, \ldots, n\}$$

Problém je zjistit index j největšího prvku. Daný problém řeší algoritmus MAX z obrázku 1 (viz [4]). Kolik instrukcí algoritmus MAX vykoná při zpracování vstupního pole A velikosti n? Bez ohledu na to, jaký vstup $A[0..n-1]$ (tj. pole délky n) je zpracováván, vykoná se:

3 instrukce přiřazení na řádku 1,

n instrukcí porovnání na řádku 2,

$n-1$ instrukcí porovnání na řádku 3,

$n-1$ instrukcí přiřazení na řádku 5,

1 instrukce návratu 6.

K tomu se ale vykoná $2 \cdot y_n$ instrukcí na řádku 4, kde y_n je počet kladných vyhodnocení podmínky $A[k] > max$. Tento počet závisí na vstupu $\omega = A[0..n-1]$. Lze ho tedy chápat jako hodnotu náhodné veličiny Y_n, tj. $y_n = Y_n(\omega)$. Celkem se tedy vykoná $3 + n + n - 1 + n - 1 + 1 + 2Y_n(\omega)$, tj.

$$X_n(\omega) = 3n + 2 + 2Y_n(\omega)$$

instrukcí. X_n je tedy náhodná veličina, jejíž hodnota $X_n(\omega)$ pro vstup ω je počet vykonaných instrukcí pro n-prvkové vstupní pole ω. Vstupy ω

velikosti n jsou právě všechna výše popsaná pole $A[0..n-1]$, tj. vlastně permutace prvků $1, 2, \dots n$. Jejich množinu označme Ω_n. Hodnota náhodné X_n veličiny závisí na pravděpodobnostním prostoru $\langle \Omega_n, \mathcal{B}_n, P_n \rangle$, tedy na hodnotách pravděpodobnosti $P_n(\{\omega\})$ jednotlivých vstupů $\omega = A \in \Omega_n$.

Je zřejmé, že $Y_n(A) \in \{0, 1, \dots, n-1\}$. Např. $Y_n(A) = 0$, právě když největším prvkem pole A je $A[n-1]$; $Y_n(A) = n-1$, právě když pole A je uspořádané sestupně, tj. $A[0] > \cdots > A[n-1]$. Vzhledem k předpokladům je jasné, že možných vstupů velikosti n je $n!$ (počet permutací n prvků). Předpokládejme, že pravděpodobnost každého vstupního pole A je stejná, tj. $P(\{A\}) = {}^1\!/_{n!}$. Pro pravděpodobnostní funkci p_{Y_n} veličiny Y_n platí

$$\begin{aligned} p_{Y_n}(k) &= P(Y_n = k) = P(\{A \in \Omega \mid Y_n(A) = k\}) = \\ &= \frac{\text{počet vstupních polí } A[0..n-1], \text{ pro které } Y_n(A) = k}{n!} \end{aligned}$$

pro každou $k \in \{0, \dots, n-1\}$.

Uvažujme vstup $A = \langle A[0], \dots A[n-1] \rangle$ a dva jevy:

$$B = \{A \in \Omega_n \mid A[0] = n\} \text{ a } \overline{B} = \{A \in \Omega_n \mid A[0] \neq n\}.$$

Za předpokladu, že nastane jev B, se po porovnání na řádku 3 s hodnotou $A[0]$ řádek 4 provede. Tedy hodnota Y_n v tomto případě, tj. hodnota $Y_n(\langle A[0], \dots A[n-1] \rangle)$ bude o 1 vyšší než počet kladných vyhodnocení podmínky $A[k] > max$ při zpracování posloupnosti $\langle A[1], \dots, A[n-1] \rangle$. Protože je $\langle A[1], \dots, A[n-1] \rangle$ pole velikosti $n-1$, je počet těchto kladných vyhodnocení roven $Y_{n-1}(\langle A[1], \dots, A[n-1] \rangle)$. Z toho, co jsme řekli, plyne, že

$$P(Y_n = k | B) = P(Y_{n-1} = k-1) = p_{Y_{n-1}}(k-1).$$

Za předpokladu, že nastane jev $\overline{B}$, se po porovnání na řádku 3 s hodnotou $A[0]$ řádek 4 neprovede (největší hodnota pole A je již v proměnné max). Tedy hodnota Y_n v tomto případě, tj. hodnota $Y_n(\langle A[0], \dots A[n-1] \rangle)$ bude rovna počtu kladných vyhodnocení podmínky $A[k] > max$ při zpracování posloupnosti $\langle A[1], \dots, A[n-1] \rangle$, tj.

$$P(Y_n = k | \overline{B}) = P(Y_{n-1} = k) = p_{Y_{n-1}}(k).$$

Snadno se vidí, že $P(B) = \frac{1}{n}$ a $P(\overline{B}) = \frac{n-1}{n}$. Podle věty o úplné pravděpodobnosti proto platí

$$\begin{aligned} p_{Y_n}(k) &= P(Y_n = k) = P(Y_n = k|B)P(B) + P(Y_n = k|\overline{B})P(\overline{B}) = \\ &= \frac{1}{n} p_{Y_{n-1}}(k-1) + \frac{n-1}{n} p_{Y_{n-1}}(k). \end{aligned} \tag{40}$$

Tato rekurentní rovnice nám umožňuje vyčíslit jednotlivé pravděpodobnosti $p_{Y_n}(k) = P(Y_n = k)$, pomocí kterých můžeme snadno spočítat střední hodnotu veličiny Y_n, totiž $E(Y_n) = \sum_{k=0}^{n-1} k \cdot P(Y_n = k)$. Nakonec pak můžeme spočítat střední hodnotu veličiny X_n, tedy očekávaný počet vykonaných instrukcí potřebných pro zpracování pole velikosti n:

$$E(X_n) = E(3n + 2 + 2Y_n) = 3n + 2 + 2E(Y_n),$$

neboť n je konstanta a, jak víme, pro libovolnou náhodnou veličinu platí $E(aZ + b) = aE(Z) + b$.

Příklad 5.1. Určete $E(X_n)$ pro $n = 2, 3, 4$.

Pro výpočet podle rekurentní rovnice (40) je třeba určit hodnoty p_{Y_1}. Jedinou možnou hodnotou veličiny Y_1 je 0, je tedy $p_{Y_1}(0) = 1$ a $p_{Y_1}(k) = 0$ pro $k \neq 0$.

$n = 2$: Možné hodnoty veličiny Y_2 jsou 0 a 1 a platí

$$\begin{aligned} p_{Y_2}(0) &= \frac{1}{n} p_{Y_1}(-1) + \frac{n-1}{n} p_{Y_1}(0) = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}, \\ p_{Y_2}(1) &= \frac{1}{n} p_{Y_1}(0) + \frac{n-1}{n} p_{Y_1}(1) = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 0 = \frac{1}{2}. \end{aligned}$$

Tedy $E(Y_2) = 0 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = \frac{1}{2}$ a

$$E(X_2) = 3n + 2 + 2E(Y_n) = 3 \cdot 2 + 2 + 2 \cdot \frac{1}{2} = 9.$$

$n = 3$: Možné hodnoty veličiny Y_3 jsou 0, 1 a 2 a platí

$$\begin{aligned} p_{Y_3}(0) &= \frac{1}{n} p_{Y_2}(-1) + \frac{n-1}{n} p_{Y_2}(0) = \frac{1}{3} \cdot 0 + \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}, \\ p_{Y_3}(1) &= \frac{1}{n} p_{Y_2}(0) + \frac{n-1}{n} p_{Y_2}(1) = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{2}, \\ p_{Y_3}(2) &= \frac{1}{n} p_{Y_2}(1) + \frac{n-1}{n} p_{Y_2}(2) = \frac{1}{3} \cdot \frac{1}{2} + \frac{2}{3} \cdot 0 = \frac{1}{6}. \end{aligned}$$

Tedy $E(Y_3) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{6} = \frac{5}{6}$ a konečně

$$E(X_3) = 3n + 2 + 2E(Y_3) = 3 \cdot 2 + 2 + 2 \cdot \frac{5}{6} \approx 9.67.$$

$n = 4$: Dopočítejte.

5.3 Pravděpodobnostní analýza algoritmu Quicksort

Viz podklady na přednášce:

- Cormen-probabilistic-analysis-of-algorithms.pdf,
- Cormen-randomized-quicksort-analysis.pdf,
- Cormen-birthday-paradox-analysis.pdf (pro zajímavost).

6 Limitní věty

Limitní věty teorie pravděpodobnosti představují důležité poznatky ohledně situací, které lze nahlížet tak, že v nich jde o posloupnost $X_1, X_2, \ldots$ náhodných veličin X_i. Představme si například, že provedeme n nezávislých opakování náhodného pokusu, ve kterém sledujeme hodnotu nějaké veličiny. Situaci můžeme nahlížet tak, že sledujeme hodnoty n nezávislých náhodných veličin, $X_1, \ldots, X_n$, které mají stejné rozdělení. Intuitivně cítíme, že čím více opakování provedeme, tím více informace (například o střední hodnotě daného rozdělení) získáme. Zvětšování počtu opakování vede na otázku o limitním (pro n jdoucím k nekonečnu) chování různých charakteristik, které lze z hodnot veličin $X_1, \ldots, X_n$ odvodit.

6.1 Zákon velkých čísel

6.1.1 Neformální úvahy

Zákony velkých čísel (angl. laws of large numbers) říkají,[10] že při velkém počtu nezávislých opakovaní daného náhodného pokusu je průměr hodnot sledované veličiny blízko střední hodnotě této veličiny a že s rostoucím počtem opakování je tento průměr střední hodnotě více a více blíží. Zákon velkých čísel v jistém smyslu zaručuje, že průměr sledovaných hodnot je z dlouhodobého hediska stabilní. Například zatímco o hodnotě veličiny v jednom (příště provedeném) pokusu nelze obvykle s rozumnou jistotou nic říct, zákon velkých čísel říká, že z dlouhodobého hlediska se bude průměr pozorovaných hodnot blížit střední hodnotě. Při hazardní hře tedy v příštím kole může hráč proti kasinu vyhrát (s malou

[10] Existuje několik verzí zákona velkých čísel.

pravděpodobností, ale může); zákon velkých čísel ale říká, že při opakovaném počtu sázek se jeho průměrná výhra bude blížit očekávané, a bude tedy záporná (protože tak jsou nastaveny parametry hry), tj. hráč bude tratit, kasino bude vyhrávat.

Historická poznámka 6.1. V 16. století vyslovil tvrzení, že přesnost statistických charakteristik s rostoucím počtem opakování pokusů roste, matematik Gerolamo Cardano (1501–1576). Toto tvrzení lze považovat za předchůdce později formulovaných a dokázaných matematických tvrzení, která jsou známa jako zákony velkých čísel. Prvním, kdo takové tvrzení přesně formuloval a dokázal, byl Jacob Bernoulli (1655–1705) ve své práci *Ars Conjectandi* z roku 1713. Název „zákon velkých čísel" poprvé použil Siméon Denis Poisson (1781–1840) v roce 1837.

Předpokládejme, že je dána posloupnost $X_1, X_n, \ldots$, nezávislých náhodných veličin se stejným rozdělením pravděpodobnosti a že pro střední hodnoty a rozptyl těchto veličin platí $E(X_i) = \mu$ a $\mathrm{var}(X_i) = \sigma^2$ pro $i = 1, 2 \ldots$ Zákon velkých čísel pak v tomto případě říká (existují však jeho formulace pro obecnější předpoklady), že veličina $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_n$ konverguje k dané střední hodnotě μ; symbolicky $\overline{X}_n \to \mu$ pro $n \to \infty$. Existují však dvě pojetí konvergence: konvergence podle pravděpodobnosti (angl. convergence in probability), pro kterou se uvedené tvrzení nazývá *slabý zákon velkých čísel*, a konvergence skoro jistě (angl. convergence almost surely), pro kterou se tvrzení nazývá *silný zákon velkých čísel*. Konvergence skoro jistě implikuje konvergenci podle pravděpodobnosti, naopak ne.

6.1.2 Matematická formulace

Omezíme se na slabý zákon velkých čísel. Uveďme nejprve potřebný pojem konvergence.

Posloupnost náhodných veličin $X_1, X_2, \ldots$ *konverguje podle pravděpodobnosti* k hodnotě a, jestliže ro každé $\varepsilon > 0$ platí

$$\lim_{n\to\infty} P(|X_n - a| \geq \varepsilon) = 0. \tag{41}$$

Zde "$|X_n - a| \geq \varepsilon$" označuje jev $\{\omega \in \Omega \mid |X_n(\omega) - a| \geq \varepsilon\}$. Daná konvergence tedy říká, že pravděpodobnost libovolné absolutní hodnoty odchylky X_n od a konverguje k 0.

Příklad 6.1. Ověřme, že posloupnost X_n náhodných veličin s normálním rozdělením $N(\mu, \sigma^2/n)$ konverguje podle pravděpodobnosti k μ.

(Nakreslete is obrázek distribuční funkce F veličiny X_n.) Platí, že $P(|X_n-\mu| \geq \varepsilon) = P([X_n \leq \mu-\varepsilon] \cup [X_n \geq \varepsilon+\mu]) = P(X_n \leq \mu-\varepsilon) + P(X_n \geq \varepsilon+\mu) = 2P(X_n \leq \mu - \varepsilon)$. Poslední rovnost platí vzhledem k symetrii patrné např. z grafu F.

Je $P(X_n \leq \mu - \varepsilon) = F(\mu - \varepsilon)$. Protože pro distribuční funkci Φ normovaného normálního rozdělení platí $F(x) = \Phi(\frac{x-\mu}{s})$, kde μ a s jsou parametry normálního rozdělení s distribuční funkcí F, je

$$F(x) = \Phi(\frac{x-\mu}{\frac{\sigma}{\sqrt{n}}}) = \Phi(\frac{\sqrt{n}(x-\mu)}{\sigma}),$$

a tedy

$$F(\mu-\varepsilon) = \Phi(\frac{\sqrt{n}(-\varepsilon)}{\sigma}) = 1 - \Phi(\frac{\sqrt{n}\varepsilon}{\sigma}),$$

neboť $\Phi(-u) = 1 - \Phi(u)$. Protože $\lim_{n\to\infty} \Phi(\frac{\sqrt{n}\varepsilon}{\sigma}) = 1$, je pro $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n-\mu| \geq \varepsilon) = \lim_{n\to\infty} 2(1-\Phi(\frac{\sqrt{n}\varepsilon}{\sigma})) = 2(1-\lim_{n\to\infty} \Phi(\frac{\sqrt{n}\varepsilon}{\sigma})) = 2(1-1) = 0.$$

Označme pro posloupnost X_1, X_2, … náhodných veličin

$$Z_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - E(X_i)). \tag{42}$$

Tj. Z_n je nová náhodná veličina vyjadřující průměr odchylek veličin X_1, …, X_n od jejich středních hodnot. Posloupnost X_1, X_2, … náhodných veličin *splňuje slabý zákon velkých čísel*, jestliže postloupnost Z_1, Z_2, … konverguje podle pravděpodobnosti k 0, tj. pro libovolné $\varepsilon > 0$ je

$$\lim_{n\to\infty} P(|Z_n| \geq \varepsilon) = 0$$

neboli

$$\lim_{n\to\infty} P\left(|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| \geq \varepsilon\right) = 0.$$

Uvedeme nyní několik podmínek, které zaručují, že daná posloupnost slabý zákon velkých čísel splňuje.

Věta 6.1 (Čebyševova nerovnost)**.** *Má-li náhodná veličina X střední hodnotu $E(X)$ a rozptyl $\mathrm{var}(X)$, pak pro každé $\varepsilon > 0$ platí*

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\mathrm{var}(X)}{\varepsilon^2}. \tag{43}$$

Důkaz. Přednášky. □

Věta 6.2 (Čebyševova)**.** *Nechť pro posloupnost X_1, X_2, … náhodných veličin existuje $K > 0$ tak, že $\mathrm{var}(X_i) \leq K$ pro $i = 1, 2, \ldots$ a $\mathrm{cov}(X_i, X_j) = 0$ pro $i \neq j$, i,j=1,2,…. Pak pro X_1, X_2, … platí slabý zákon velkých čísel.*

Důkaz. Uvažujme veličiny Z_n definované v (42). Snadno se ukáže, že $E(Z_n) = 0$, dále pak

$$\mathrm{var}(Z_n) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{var}(X_i) \leq \frac{1}{n^2} \sum_{i=1}^{n} K = \frac{K}{n}.$$

Podle Čebyševovy nerovnosti je

$$P(|Z_n| \geq \varepsilon) = P(|Z_n - 0| \geq \varepsilon) \leq \frac{\mathrm{var}(Z_n)}{\varepsilon^2} \leq \frac{K}{n\varepsilon^2}$$

a vzhledem k $\lim_{n\to\infty} \frac{K}{n\varepsilon^2} = 0$ je tedy $\lim_{n\to\infty} P(|Z_n| \geq \varepsilon) = 0$. □

Čebyševova věta má důležité praktické důsledky. Uvedeme jeden obecnější, jeden konkrétnější.

Věta 6.3. *Nechť je dána posloupnost $X_1, X_2, \ldots$ vzájemně nezávislých náhodných veličin, které mají stejnou střední hodnotu $E(X_i) = \mu$ a stejný rozptyl $\mathrm{var}(E_i) = \sigma^2$. Tato posloupnost splňuje slabý zákon velkých čísel a posloupnost $\overline{X}_1, \overline{X}_2, \ldots,$ kde $\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$, konverguje podle pravděpodobnosti k μ.*

Důkaz. Přímo z Čebyševovy věty plyne, že Z_1, Z_n, … konverguje podle pravděpodobnosti k 0. Ze vzájemné nezávislosti X_i a X_j totiž plyne $\mathrm{cov}(X_i, X_j) = 0$. Protože $\frac{1}{n} \sum_{i=1}^{n} E(X_i) = \mu$, znamená to, že pro každé $\varepsilon > 0$ je

$$\lim_{n\to\infty} P\left(|\frac{1}{n} \sum_{i=1}^{n} X_i - \mu| \geq \varepsilon\right) = 0,$$

což zbývalo dokázat. □

Teď ten konkrétnější:

Věta 6.4 (Bernoulliho)**.** *Nechť je dána posloupnost X_1, X_2, ... vzájemně nezávislých náhodných veličin s alternativním rozdělením s parametrem p (tj. s binomickým rozdělením s parametry $n = 1$ a p). Pak posloupnost $\overline{X}_1$, $\overline{X}_2$, ..., kde $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, konverguje podle pravděpodobnosti k p.*

Důkaz. Přímo z věty 6.3 s využitím skutečnosti, že $E(X_i) = p$. □

6.2 Centrální limitní věty

Centrální limitní věty[11] jsou tvrzení následujícího typu: Provádíme opakovaně náhodný pokus a tato opakovaná provádění jsou reprezentována posloupností X_1, X_2, ... náhodných veličin. Pak za určitých předpokladů (např. že veličiny jsou vzájemně nezávislé a mají stejná rozdělení) platí, že aritmetický průměr $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ má v limitě přibližně normální rozdělení, ať je rozdělení původních náhodných veličin jakékoli. Provádíme-li tedy například opakovaně a nezávisle náhodný pokus, při kterém sledujeme hodnoty náhodné veličiny (třeba házíme mincí a sledujeme, zda padne panna nebo orel, které odpovídají hodnotám 1 a 0), pak podle centrální limitní věty mají průměrné hodnoty (průměrný počet padnutí panny) přibližně normální rozdělení a čím vice opakování provedeme, tím více se bude blížit normálnímu rozdělení. Přitom má-li daná náhodná veličina střední hodnotu μ, má i ono normální rozdělení střední hodnotu μ (při zmíněném hodu mincí má tedy ono normální rozdělení střední hodnotu 0.5).

Posloupnost náhodných veličin X_1, X_2, ... s distribučními funkcemi F_1, F_2, ... *konverguje v distribuci* k náhodné veličině X s distribuční funkcí F, jestliže vkaždém bodě x spojitosti funkce F platí

$$\lim_{n\to\infty} F_n(x) = F(x).$$

Pokud veličina X má normální rozdělení $N(\mu, \sigma^2)$, což vzhledem ke vztahu $F(x) = \Phi(\frac{x-\mu}{\sigma})$ (Φ je distribuční funkce normovaného normálního rozdělení $N(0,1)$) znamená, že pro $x \in \mathbb{R}$ je $\lim_{n\to\infty} F_n(x) = \Phi(\frac{x-\mu}{\sigma})$, posloupnost X_1, X_2, ... konverguje v distribuci k rozdělení $N(\mu, \sigma^2)$ nebo že X_n má asymptoticky rozdělení $N(\mu, \sigma^2)$.

[11] Existuje několik verzí centrální limitní věty.

Věta 6.5 (Lindenbergova-Lévyho)**.** *Nechť je dána posloupnost $X_1, X_2, \dots$ vzájemně nezávislých náhodných veličin, které mají stejnou střední hodnotu $E(X_i) = \mu$ a stejný rozptyl $\mathrm{var}(E_i) = \sigma^2$. Pak posloupnost $Y_1, Y_2, \dots,$ kde*

$$Y_n = \frac{\sqrt{n}}{\sigma}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$$

konverguje v distribuci k rozdělení $N(0,1)$.

Důkaz. Přednášky. □

Poznámka 6.1. (a) Y_n lze nahlížet jako normované průměry $\overline{X}_n$ prvních n veličin, $X_1, \dots, X_n$ (normování: centrování odečtením μ, normalizace rozptylu dělením σ) násobené $\sqrt{n}$. Definice veličin Y_n lze ekvivalentně popsat takto (vytknutím n):

$$Y_n = \frac{1}{\sigma\sqrt{n}}\left(\sum_{i=1}^{n} X_i - n\mu\right)$$

(b) Všimněte si, že věta 6.5 má stejné předpoklady jako věta 6.3. Každá z nich říká o průměru $\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ něco jiného. Věta 6.3 hovoří o konvergeci hodnot průměru; věta 6.5 o rozptylu normovaného průměru.

(c) Větu 6.5 lze ekvivalentně formulovat tak, že za uvedených předpokladů posloupnost veličin

$$Y_n = \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right)$$

konverguje v distribuci k rozdělení $N(0,\sigma^2)$. (Ověřte.)

Speciálním případem Lindenbergovy-Lévyho věty je následující věta.

Věta 6.6 (Moivreova-Laplaceova)**.** *Nechť je dána posloupnost $X_1, X_2, \dots$ vzájemně nezávislých náhodných veličin s alternativním rozdělením s parametrem p (tj. s binomickým rozdělením s parametry $n = 1$ a p). Pak posloupnost $Y_1, Y_2, \dots,$ kde*

$$Y_n = \frac{1}{\sqrt{np(1-p)}}\left(\sum_{i=1}^{n} X_i - np\right)$$

konverguje v distribuci k rozdělení $N(0,1)$.

Důkaz. Plyne z věty 6.5 a z toho, že $E(X_i) = p$ a $\mathrm{var}(X_i) = p(1-p)$. □

Obecnější případy jsou předmětem dalších centrálních limitních vět. Jako příklad uveďme následující.

Věta 6.7 (Ljapunovova)**.** *Nechť je dána posloupnost* $X_1, X_2, \ldots$ *vzájemně nezávislých náhodných veličin, pro které* $E(X_i) = \mu_i$ *a* $\operatorname{var}(X_i) = \sigma_i^2$ *a existuje třetí absolutní centrální moment* $E(|X_i - \mu_i|^3)$*. Pokud*

$$\lim_{n\to\infty} \frac{\left(\sum_{i=1}^n E(|X_i - \mu_i|^3)\right)^{\frac{1}{3}}}{\sqrt{\sum_{i=1}^n \sigma_i^2}} = 0,$$

pak posloupnost $Y_1, Y_2, \ldots$*, kde*

$$Y_n = \frac{\left(\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i\right)}{\sqrt{\sum_{i=1}^n \sigma_i^2}}$$

konverguje v distribuci k rozdělení $N(0,1)$*.*

Důkaz. Přednášky. □

Reference

[1] Anděl, J. 2003. *Matematika náhody.* Praha: Matfyzpress.

[2] Likeš, J., Machek, J. 1987. *Počet pravděpodobnosti.* Praha: Státní nakladatelství technické literatury.

[3] Rényi, A. 2007. *Probability Theory.* Mineola, NY: Dover. Původně vyšlo v roce 1970 v North-Holland a Akadémiai Kiadó.

[4] Trivedi, K. S. 2001. *Probability and Statistics with Reliability, Queueing, and Computer Science Applications*, 2nd ed. Hoboken, NJ: Wiley.

Další:

Reference

[1] Fine, T. L. 1973. *Theories of Probability: An Examination of Foundations.* New York: Academic Press.

[2] Hacking I. 2001. *An Introduction to Probability nad Inductive Logic.* Cambridge, MA: Cambridge Univ. Press.

[3] Kolmogorov A. N. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung* [německy]. Berlin: Springer. Angl. překlad: Foundations of the Theory of Probability (2nd ed.). New York: Chelsea, 1956.