Úvod do teorie pravděpodobnosti v předmětu Diskrétní matematika

Jiří Matoušek (KAM MFF UK)

Verze: 31/X/2009

Úvod

Tento podrobný sylabus obsahuje definice pojmů, formulace tvrzení a stručné komentáře. Nenahrazuje učebnici a nenajdete v něm důkazy ani řešené příklady. Část látky je podrobněji sepsána v knize J. Matoušek, J. Nešetřil: *Kapitoly z diskrétní matematiky*, Nakladatelství Karolinum, Praha, 2007 (kap. 10). Učebnic teorie pravděpodobnosti existuje mnoho, několik i v češtině (např. K. Zvára, J. Štěpán: *Pravděpodobnost a matematická statistika*, Matfyzpress, Praha, 2002 nebo A. Rényi: *Teorie pravděpodobnosti*, NČSAV, Academia, Praha, 1972), ty jsou ovšem mnohem obsažnější než náš minimalistický úvod.

1. Symbolem náhody je hrací kostka. Pravděpodobnost, že při jednom hodu padne šestka, je $\frac{1}{6}$. Pravděpodobnost, že při dvou hodech po sobě padnou šestky, je $\frac{1}{36}$. Ještě jednodušší je *hod spravedlivou mincí*: rub i líc mají pravděpodobnost $\frac{1}{2}$.

2. Co myslíme pravděpodobností? Pro „skutečnou“ pravděpodobnost je to filozofický problém, různé odpovědi, žádná dokonalá. (Hody kostkou můžeme opakovat, „empirická“ pravděpodobnost jako limita pro obrovský počet hodů – ale má smysl mluvit o pravděpodobnosti jedinečných jevů, třeba že gorily – nebo lidé – přežijí rok 2010?) Kde se bere náhodnost? (Možná v kvantové teorii? Nebo v teorii chaosu – co se nám nepoučeným zdá náhodné, je možná ve skutečnosti předurčeno?)

3. Matematická teorie pravděpodobnosti se těmito filozofickými otázkami nezabývá. Sestrojí matematický *model* pravděpodobnosti. Ten funguje skvěle, zeptejte se pojišťoven (které zrovna nekrachují)! Ale na reálné problémy se musí aplikovat uvážlivě (zeptejte se pojišťoven, které právě krachují). Subjektivní odhady pravděpodobností často selhávají, zejména u hodně malých nebo hodně velkých pravděpodobností. (Experiment: jen 45 % studentů dokončilo diplomku do termínu, o kterém si mysleli, že ji budou mít hotovou na 99 %.)

4. Pravděpodobnost v informatice: pravděpodobnostní algoritmy (rychlejší a jednodušší než deterministické), statistické testy a zpracování dat, matematické důkazy.

Pravděpodobnostní prostory, hlavně diskrétní

5. Provedeme náhodný pokus, množinu všech možných výsledků označíme Ω. Příklady:

   - Hod mincí: $\Omega = \{L, R\}$ (líc, rub).
   - Hod hrací kostkou: $\Omega = \{1, 2, 3, 4, 5, 6\}$.
   - Tři hody mincí po sobě: $\Omega = \{LLL, LLR, LRL, LRR, RLL, RLR, RRL, RRR\}$.
   - První kapka deště na čtvercový zahradní stolek: $\Omega =$ všechny body čtverce.

   Prvky Ω se nazývají **elementární jevy**.

6. **Jev** je podmnožina $A \subseteq \Omega$. Příklady:

   - hod mincí – můžeme uvažovat 4 jevy: $A_1 = \{L\}$ (padl líc), $A_2 = \{R\}$ (padl rub), $A_3 = \{L, R\}$ (padl líc *nebo* rub, to je **jev jistý**, nastane vždy) a $A_4 = \emptyset$ (nepadlo nic, to je **jev nemožný**, nenastane nikdy).
   - hod hrací kostkou, $\Omega = \{1, 2, 3, 4, 5, 6\}$: „padlo sudé číslo“ $A_1 = \{2, 4, 6\}$, „padla šestka“ $A_2 = \{6\}$, „nepadlo nic“ $A_3 = \emptyset$.
   - první kapka deště: „padne do levé poloviny stolku“, „padne nejvýš 10 cm od okraje“ – všelijaké geometrické obrazce ve čtverci.

7. Každému jevu A je přiřazeno reálné číslo $\mathrm{P}[A] \in [0, 1]$, zvané **pravděpodobnost jevu A**. Příklady:

   - Pro hod spravedlivou mincí máme $\mathrm{P}[\{L\}] = \mathrm{P}[\{R\}] = \frac{1}{2}$, $\mathrm{P}[\{L, R\}] = 1$, $\mathrm{P}[\emptyset] = 0$.
   - Pro hod „spravedlivou“ kostkou máme $\mathrm{P}[\{1\}] = \mathrm{P}[\{2\}] = \cdots = \mathrm{P}[\{6\}] = \frac{1}{6}$, a například také $\mathrm{P}[\{1, 3, 5\}] = \frac{1}{2}$.
   - Můžeme také uvažovat kostku falešného hráče, pro niž platí třeba $\mathrm{P}[\{6\}] = \frac{1}{5}$, $\mathrm{P}[\{1\}] = \mathrm{P}[\{2\}] = \cdots = \mathrm{P}[\{5\}] = \frac{4}{25}$. Tedy ne všechny jednoprvkové jevy musí mít vždycky stejnou pravděpodobnost!

8. *Pravděpodobnostní prostor* je formálně uspořádaná trojice $(\Omega, \mathcal{F}, \mathrm{P})$, kde

   - Ω říká, jakou uvažujeme množinu elementárních jevů,
   - $\mathcal{F} \subseteq 2^{\Omega}$ říká, které podmnožiny Ω připouštíme jako jevy,
   - $\mathrm{P}\colon \mathcal{F} \to [0, 1]$ je funkce, která každému jevu přiřazuje jeho pravděpodobnost.

9. Každý pravděpodobnostní prostor musí splňovat určité axiomy. Zde je nebudeme uvádět v plné obecnosti, protože na to zatím nemáme matematické prostředky. Přesně zavedeme jen *diskrétní* pravděpodobnostní prostory, pro něž je množina Ω konečná nebo spočetná a *každá* její podmnožina je jevem.

   Ponecháváme stranou například „geometrickou“ pravděpodobnost (jaká je pravděpodobnost, že první kapka deště dopadne nejvýš 10 cm od okraje stolku, apod.) i pravděpodobnosti, týkající se libovolně dlouhých posloupností. (Například: házíme opakovaně mincí, padá líc/rub. Co je pravděpodobnější, že se dřív objeví sekvence LLR, nebo LRR? „Reálná“ otázka – vsaďte se! Překvapivá odpověď: LLR vyhraje dvakrát častěji!)

10. **Diskrétní pravděpodobnostní prostor** je trojice $(\Omega, \mathcal{F}, \mathrm{P})$, kde Ω je konečné nebo spočetné, $\mathcal{F} = 2^\Omega$ (tj. každá podmnožina je jev), pravděpodobnost každého jevu $A \subseteq \Omega$ splňuje

$$\mathrm{P}[A] = \sum_{\omega \in A} \mathrm{P}[\{\omega\}]$$

   a $\mathrm{P}[\Omega] = 1$.

   To znamená, že diskrétní pravděpodobnostní prostor je plně určen pravděpodobnostmi všech jednoprvkových jevů. Pravděpodobnosti jednoprvkových jevů přitom můžeme zvolit jako libovolná nezáporná čísla, jejichž součet přes celé Ω je roven 1 (pro Ω nekonečné je to součet nekonečné řady).

11. Zde budeme pracovat hlavně s *konečnými* diskrétními pravděpodostními prostory, kde Ω je konečná množina. (V takovém případě budeme slovo „diskrétní“ zpravidla vynechávat.)

12. Základním příkladem konečného pravděpodobnostního prostoru je *klasický pravděpodobnostní prostor*, kde $\mathrm{P}[A] = \frac{|A|}{|\Omega|}$; zde mají všechny jednoprvkové jevy stejnou pravděpodobnost. (Tohle se pro spočetné Ω udělat nedá!)

13. Konkrétní příklady zvlášť důležité pro diskrétní matematiku:

    - Náhodná posloupnost n hodů spravedlivou mincí: $\Omega = \{L, R\}^n$, každá možná posloupnost výsledků má pravděpodobnost $1/2^n$.
    - Náhodná permutace množiny $\{1, 2, \ldots, n\}$: $\Omega = S_n$ (množina všech permutací), každá permutace má pravděpodobnost $1/n!$.

14. Pro daný matematický nebo praktický problém se vhodný pravděpodobnostní prostor dá často zvolit více způsoby.

15. Na co jsou potřeba pravděpodobnostní prostory? Bezpečný základ, mohou pomoct vyjasnit zapeklité otázky. Příklad: *Bertrandův paradox s krabicemi*, zde v karetní verzi. V klobouku jsou tři karty: jedna z nich je z obou stran červená, druhá z obou stran černá, a třetí má jednu stranu červenou a druhou černou. Vytáhnete naslepo jednu kartu a položíte ji na stůl. Pak se podíváte a vidíte, že horní strana je červená. Jaká je pravděpodobnost, že dolní strana je také červená? Intuitivní odpověď je $\frac{1}{2}$, ale ve skutečnosti je to $\frac{2}{3}$. (Viz též modernizovanou verzi – *Monty Hallův problém*).

Podmíněná pravděpodobnost, nezávislé jevy

16. **Podmíněná pravděpodobnost** jevu A za předpokladu, že nastal jev B, se definuje jako

$$\mathrm{P}[A \mid B] = \frac{\mathrm{P}[A \cap B]}{\mathrm{P}[B]}$$

   (definováno pouze pro $\mathrm{P}[B] > 0$).

17. Často je užitečné počítat pravděpodobnost jevu A „rozlišením případů“: Jsou-li $B_1, \ldots, B_n$ *disjunktní* jevy, jejichž sjednocení je celé Ω, pak

$$\mathrm{P}[A] = \mathrm{P}[A \mid B_1]\,\mathrm{P}[B_1] + \mathrm{P}[A \mid B_2]\,\mathrm{P}[B_2] + \cdots + \mathrm{P}[A \mid B_n]\,\mathrm{P}[B_n]\,.$$

   Toto jednoduché tvrzení se nazývá **věta o úplné pravděpodobnosti**.

18. Příklad (poučný). Řekněme, že podíl nakažených HIV v populaci je 0,1 %. Představte si, že podstoupíte test na HIV, o kterém se ví, že u *skutečně nakažených* dá pozitivní výsledek v 95 % případů a u *nenakažených* dá (falešný) pozitivní výsledek v 5 % případů. Nemáte žádný zvláštní důvod předpokládat, že byste byli HIV pozitivní, ale test vám vyjde pozitivně. Jaká je pravděpodobnost, že jste skutečně nakaženi? Pravděpodobnostní prostor = populace (vy z ní náhodně vybraní), jev H = HIV pozitivní, jev T = pozitivní test, předpoklady: $\mathrm{P}[H] = 0{,}001$, $\mathrm{P}[T \mid H] = 0{,}95$, $\mathrm{P}[T \mid \text{ne } H] = 0{,}05$, zajímá nás: $\mathrm{P}[H \mid T]$, vyjde méně než 2 %.

   Formulujeme-li řešení tohoto příkladu obecně, dostaneme **Bayesovu větu**:

$$\mathrm{P}[B_i \mid A] = \frac{\mathrm{P}[A \mid B_i]\,\mathrm{P}[B_i]}{\sum_{j=1}^{n} \mathrm{P}[A \mid B_j]\,\mathrm{P}[B_j]},$$

   kde $B_1, \ldots, B_n$ jsou jevy jako v předchozím bodu (disjunktní a pokrývají Ω).

19. Jevy A a B se nazývají **nezávislé**, pokud $\mathrm{P}[A \cap B] = \mathrm{P}[A]\,\mathrm{P}[B]$. Ekvivalentně, A a B jsou nezávislé, pokud $\mathrm{P}[B] = 0$ nebo $\mathrm{P}[A \mid B] = \mathrm{P}[A]$.

(Intuitivně: dozvíme-li se, zda nastal B, nezískáme žádnou novou informaci o pravděpodobnosti A.) Obecněji: jevy $A_1, A_2, \ldots, A_n$ jsou nezávislé, pokud pro každou podmnožinu $I \subseteq \{1, 2, \ldots, n\}$ platí $\mathrm{P}\left[\bigcap_{i\in I} A_i\right] = \prod_{i\in I} \mathrm{P}[A_i]$.

20. Příklad: Náhodná posloupnost 10 nul a jedniček,
    $A_1 = \{$v prvních 5 hodech padly samé 1$\}$,
    $A_2 = \{$v 6. hodu padla 0$\}$,
    $A_3 = \{$v posledních 5 hodech padl lichý počet 1$\}$.
    Jevy A_1 a A_2 jsou „zjevně“ nezávislé. Jevy A_1, A_2, A_3 jsou též nezávislé, ale to se musí pečlivě ověřit podle definice.

    Jiný příklad: náhodná permutace π čísel $1, 2, \ldots, 32$ (zamíchané karty), $A = \{\pi(1) = 1\}$, $B = \{\pi(2) = 2\}$ – *nejsou* nezávislé!

21. Častý omyl v intuitivní interpretaci nezávislých jevů („Už tak dlouho mi nepadla šestka, tak teď to konečně musí vyjít!“ – ve skutečnosti hrací kostka nemá paměť a pravděpodobnost nezávisí na tom, co padlo dřív).

22. Příklad: $\Omega = \{$modrooká blondýna, hnědooká bruneta, modrooký brunet, hnědooký blondýn$\}$, klasická pravděpodobnost, $A_1 =$ modré oči, $A_2 =$ světlé vlasy, $A_3 =$ žena (nebo, máte-li radši čísla, $\Omega = \{000,\ 011,\ 110,\ 101\}$). Každé dva z těchto jevů jsou nezávislé, ale všechny 3 nezávislé nejsou.

23. **Součin diskrétních pravděpodobnostních prostorů:**
    $(\Omega_1, 2^{\Omega_1}, \mathrm{P}_1) \times (\Omega_2, 2^{\Omega_2}, \mathrm{P}_2) = (\Omega, 2^{\Omega}, \mathrm{P})$,
    kde $\Omega = \Omega_1 \times \Omega_2$ a $\mathrm{P}[A] = \sum_{(\omega_1,\omega_2)\in A} \mathrm{P}_1[\{\omega_1\}]\mathrm{P}_2[\{\omega_2\}]$.

24. Pokud jev A_1 v součinu pravděpodobnostních prostorů „závisí“ jen na první složce a A_2 „závisí“ jen na druhé složce, potom A_1 a A_2 jsou nezávislé. Příklad „náhodná posloupnost n hodů spravedlivou mincí“ můžeme zkonstruovat jako součinový pravděpodobnostní prostor, součin n prostorů, každý pro hod jednou mincí. Zajímavější příklad: $\Omega = \{0, 1\}$, $\mathrm{P}[\{1\}] = p$ (pokus s pravděpodobností úspěchu p), součin n exemplářů modeluje posloupnost n nezávislých pokusů po sobě. Je-li $\omega \in \{0, 1\}^n$, máme $\mathrm{P}[\{\omega\}] = p^j(1-p)^{n-j}$, kde j je počet jedniček v ω.

Náhodné veličiny, střední hodnota

25. **(Reálná) náhodná veličina** na diskrétním pravděpodobnostním prostoru $(\Omega, 2^{\Omega}, \mathrm{P})$ je libovolná funkce $X\colon \Omega \to \mathbf{R}$.

    Příklad: n-krát po sobě hodíme spravedlivou mincí (množina elementárních jevů Ω je zde $\{L, R\}^n$, všechny n-členné posloupnosti L a R). Příklady náhodných veličin: kolik padlo líců; o kolik padlo více líců než rubů; sinus počtu líců, atd. (Poznámka: Náhodné veličiny mohou být též komplexní, nebo jejich hodnoty mohou být body roviny a pod., zde však budeme uvažovat pouze reálné.) Pozor, rozlišovat mezi *jevem* (pouze dva možné výsledky, nastal/nenastal, formálně je to množina) a *náhodnou veličinou* (číselný výsledek, formálně je to funkce).

26. Buď $X\colon \Omega \to \mathbf{R}$ náhodná veličina na diskrétním pravděpodobnostním prostoru $(\Omega, 2^{\Omega}, \mathrm{P})$. **Střední hodnota $\boldsymbol{X}$** je definována jako
    $$\mathbf{E}[X] := \sum_{\omega\in\Omega} \mathrm{P}[\{\omega\}]\, X(\omega).$$

    (Pro nekonečné Ω nemusí nekonečná řada definující $\mathbf{E}[X]$ konvergovat, a potom střední hodnota neexistuje.) Představa: $\mathbf{E}[X]$ je průměrná hodnota X při velkém počtu pokusů.

27. Jiný pojem, který se v praxi někdy plete se střední hodnotou: **medián** náhodné veličiny X je číslo m takové, že $\mathrm{P}[X < m] \le \frac{1}{2}$ a $\mathrm{P}[X > m] \le \frac{1}{2}$. (Zde $\mathrm{P}[X < m]$ je obvyklý zkrácený zápis pro pravděpodobnost jevu $\{\omega \in \Omega : X(\omega) \le m\}$.) Např. polovina lidí má menší plat než medián platu náhodně vybraného člověka a polovina větší. Ale zpravidla velká většina lidí má menší plat než průměrný (a všichni chtějí aspoň plat průměrný – což jsou mnozí politici ochotni před volbami slíbit).

28. *Linearita střední hodnoty*: $\mathbf{E}[\alpha X] = \alpha\mathbf{E}[X]$, $\mathbf{E}[X+Y] = \mathbf{E}[X] + \mathbf{E}[Y]$ pro zcela libovolné náhodné veličiny X a Y a $\alpha \in \mathbf{R}$ (za předpokladu, že střední hodnoty existují). Důkaz okamžitě z definice. Pozor, obecně $\mathbf{E}[XY] \ne \mathbf{E}[X]\,\mathbf{E}[Y]$.

29. **Indikátor** jevu A je náhodná veličina $I_A\colon \Omega \to \{0, 1\}$, daná předpisem
    $$I_A(\omega) = \begin{cases} 1 & \text{pro } \omega \in A \\ 0 & \text{pro } \omega \notin A. \end{cases}$$

    Platí $\mathbf{E}[I_A] = \mathrm{P}[A]$.

30. Příklady (výpočet střední hodnoty):

    - $X =$ počet líců v posloupnosti n hodů spravedlivou mincí. Výpočet $\mathbf{E}[X]$ podle definice. Metoda indikátorů: $A_i :=$ i-tý hod je líc; $X = I_{A_1} + I_{A_2} + \cdots + I_{A_n}$, $\mathbf{E}[I_{A_i}] = \frac{1}{2}$, $\mathbf{E}[X] = \frac{n}{2}$.

- Y = počet levých maxim náhodné permutace π množiny $\{1, 2, \dots, n\}$, tj. počet indexů i takových, že $\pi(j) < \pi(i)$ pro všechna $j < i$. Jev $A_i :=$ index i je levé maximum. Spočítáme $\mathbf{E}[Y] = 1 + \frac{1}{2} + \dots + \frac{1}{n} \approx \ln n$.

31. Každý graf $G = (V, E)$ má bipartitní podgraf s aspoň $\frac{1}{2}|E|$ hranami. Důkaz: Náhodné rozdělení množiny vrcholů na dvě části, X = počet hran jdoucích napříč, $\mathbf{E}[X] = \frac{1}{2}|E|$.

32. Náhodné veličiny X, Y (na témže pravděpodobnostním prostoru) se nazývají **nezávislé**, pokud pro každé $a, b \in \mathbf{R}$ jsou jevy „$X \leq a$" a „$Y \leq b$" nezávislé. (Nezávislost se nepozná z distribučních funkcí – musíme vědět, jaký mají vztah X a Y k sobě navzájem.) Podobně pro více než 2 náhodné veličiny.

33. Pro nezávislé náhodné veličiny platí $\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]$.

Rozptyl, Čebyševova nerovnost

34. **Rozptyl** náhodné veličiny X je definován jako

$$\mathrm{Var}[X] := \mathbf{E}\left[(X - \mathbf{E}[X])^2\right].$$

Veličiny s malým rozptylem se spíše drží kolem střední hodnoty, veličiny s velkým rozptylem se často hodně odchylují.

35. Pojištění = snižování rozptylu. Dům má cenu 10 miliónů, pravděpodobnost požáru 0,01 % za rok, X = ztráta z požáru za rok, $X = 10^7$ s pravděpodobností 10^{-4}, $X = 0$ jinak, $\mathbf{E}[X] = 1000$. „Dokonale spravedlivé" pojištění 1000 za rok, místo X konstantní náhodná veličina se stejnou střední hodnotou. Proč se lidé pojišťují? Užitek z dvakrát většího majetku není dvakrát větší, pojištění zvyšuje střední hodnotu „užitku z majetku", lidé dokonce platí za pojištění víc než střední hodnotu budoucí škody. Proč je to výhodné pro pojišťovnu? Má hodně peněz, její funkce užitku je v uvažovaném rozmezí víceméně lineární.

36. **Markovova nerovnost:** Pro *nezápornou* náhodnou veličinu X platí

$$\mathrm{P}[X \geq t\mathbf{E}[X]] \leq \frac{1}{t}$$

pro všechna $t \geq 1$. Tedy: pravděpodobnost, že nezáporná náhodná veličina mnohokrát překročí svou střední hodnotu, je malá. Důkaz: $\mathbf{E}[X] \geq a\mathrm{P}[X \geq a]$, $a := t\mathbf{E}[X]$. (Obráceně to platit nemusí: nezáporná náhodná veličina X může být skoro vždycky mnohem menší než $\mathbf{E}[X]$; příklad: škoda na vašem domě z požárů za rok.)

37. **Čebyševova nerovnost** říká kvantitativně, jak se veličina s malým rozptylem drží blízko střední hodnoty:

$$\mathrm{P}[\,|X - \mathbf{E}[X]\,| \geq a] \leq \frac{\mathrm{Var}[X]}{a^2}$$

(pro každé $a \geq \sqrt{\mathrm{Var}[X]}$ a pro každou náhodnou veličinu X takovou, že $\mathbf{E}[X]$ a $\mathrm{Var}[X]$ jsou definovány). Důkaz: Markovova nerovnost pro náhodnou veličinu $Y := (X - \mathbf{E}[X])^2$.

Základní pravděpodobnostní rozdělení

38. **Rozdělení** náhodné veličiny X popisuje, s jakou pravděpodobností nabývá X jednotlivých hodnot. Nejčastěji se rozdělení popisuje pomocí **distribuční funkce** náhodné veličiny X, což je funkce $F\colon \mathbf{R} \to [0, 1]$ definovaná předpisem

$$F(z) := \mathrm{P}[\{\omega \in \Omega : X(\omega) \leq z\}]$$

(obvyklý zkrácený zápis pro pravou stranu: $\mathrm{P}[X \leq z]$).

39. Náhodná veličina na konečném pravděpodobnostním prostoru nabývá jen konečně mnoha hodnot. Její distribuční funkce je „schodovitá". Rozdělení můžeme také popsat jako množinu uspořádaných dvojic $\{(a_1, p_1), (a_2, p_2), \dots, (a_n, p_n)\}$ s významem „X nabývá hodnoty a_1 s pravděpodobností p_1, hodnoty a_2 s pravděpodobností $p_2, \dots$". Z rozdělení se dají spočítat parametry jako střední hodnota, medián nebo rozptyl.

40. Náhodné veličiny definované na různých pravděpodobnostních prostorech mohou mít stejné rozdělení.

41. Dvě náhodné veličiny se stejnou střední hodnotou a rozptylem mohou mít velice odlišná rozdělení. Pokud je to možné, měli bychom pro náhodnou veličinu udávat celé rozdělení (např. ve statistice).

42. Důležitá rozdělení náhodných veličin: *alternativní* (též *Bernoulliho*); *rovnoměrné*; *binomické*; *Poissonovo*. (Toto jsou *diskrétní* rozdělení. Velmi důležitá jsou též spojitá rozdělení, například *normální rozdělení*, ale ta se týkají nespočetných pravděpodobnostních prostorů.)