9 Některá význačná diskrétní rozdělení pravděpodobnosti

Břetislav Fajmon, UMAT FEKT, VUT Brno

Této přednášce odpovídá část kapitoly 11 ze skript [1] a vše, co se nachází v kapitole 5 sbírky úloh [2] – tuto kapitolu 5 sbírky úloh je vhodné vzít jako cvičení tohoto tématu.

Diskrétních náhodných veličin je celá řada – a všechny musí splňovat vlastnosti uvedené v otázce číslo 3 k této části předmětu:

1.

$$P(X \in \langle a; b\rangle) = \sum_{k \in \langle a;b\rangle} p(k);$$

2. $0 \le p(k) \le 1$ **pro každé** $k \in \Omega$**;**

3.

$$\sum_{k \in \Omega} p(k) = 1.$$

Některá diskrétní rozdělení jsou tak důležitá, že daný matematický model má svůj vlastní název – těmito rozděleními se budeme dnes zabývat.

9.1 Rovnoměrné diskrétní rozdělení

Možná je dobré začít tím nejjednodušším:
<u>Rovnoměrné diskrétní rozdělení</u> je rozdělení veličiny nabývající n různých hodnot, a to všechny s pravděpodobností

$$P(X = k) = \frac{1}{n} \quad \textbf{pro } k = 1, 2, \ldots, n.$$

Je to vlastně rozdělení reprezentující klasickou pravděpodobnost (viz předn. číslo 7).

9.2 Binomické rozdělení pravděpodobnosti

Binomické rozdělení pravděpodobnosti studenti také znají z přednášky číslo 7, ovšem tehdy nebylo zmíněno, že veličina měřící počet šestek při pěti hodech kostkou má toto rozdělení (nebyl zmíněn jeho název).

Podívejme se na další příklad, kde binomické rozdělení je matematickým popisem dané náhodné veličiny.

Příklad 9.1. Veličina X udává, kolikrát při digitálním přenosu deseti bitů dojde k chybě při přenosu. Přitom pravděpodobnost chyby při přenosu jednoho

bitu je $p = 0{,}1$. Také předpokládejme, že přenosy jednotlivých bitů na sobě nezávisí (tj. pravděpodobnost chybně přeneseného bitu se nemění). Určete pravděpodobnost, že při přenosu deseti bitů dojde ke dvěma přenosovým chybám.

Řešení:

$$P(X = 0) = 0{,}9^{10} \doteq 0{,}3487,$$

neboť každý bit je přenesen správně s pravděpodobností $0{,}9$, a všech těchto deset pravděpodobností mezi sebou vynásobíme, protože popisují složitější situaci, kdy nás zajímá popis charakteru všech deseti přenosů v jedné pravděpodobnostní hodnotě.

$$P(X = 1) = 0{,}1 \cdot 0{,}9^9 + 0{,}9 \cdot 0{,}1 \cdot 0{,}9^8 + \cdots + 0{,}9^9 \cdot 0{,}1,$$

neboť musíme sečíst pravděpodobnosti všech „příznivých situací“, kdy dojde k jedné chybě v deseti přenosech. Tyto jednotlivé situace se navzájem vylučují (= nemohou nastat současně) – například chybný přenos při druhém bitu a všechny další bitiy správně se vylučuje se situací, kdy nastal chybný přenos při třetím bitu a všechny další bity byly přeneseny správně. Tak tedy protože se navzájem tyto situace vylučují, jejich pravděpodobnosti mezi sebou sečteme. Vyčísleno

$$P(X = 1) \doteq 0{,}3874.$$

Podobně

$P(X = 2) = 0{,}1^2 \cdot 0{,}9^8 \cdot$ **počet výběrů dvou pozic z deseti**,

tím způsobem zachytíme všechny možné navzájem se vylučující situace, kdy na dvou pozicich z deseti dojde k chybě, a zbylých osm bitů je přeneseno správně.

Kolik je všech možných výběrů dvou pozic z deseti? Přesně tolik, kolik je všech možných výběrů dvou lidí z deseti, zkrátka tolik, kolik je všech dvouprvkových podmnožin z desetiprvkové množiny – tak přesně se definuje kombinační číslo $\binom{10}{2}$:

$$P(X = 2) = 0{,}1^2 \cdot 0{,}9^8 \cdot \binom{10}{2} = 0{,}1^2 \cdot 0{,}9^8 \cdot \frac{10 \cdot 9}{1 \cdot 2} \doteq 0{,}1937$$

(předpokládám, že student si zopakoval vlastnosti a vyčíslení kombinačních čísel v kapitole 1 sbírky úloh [2] a že je schopen u písemek s nimi pracovat).

Podobně lze nyní vyčíslit pravděpodobnosti ostatních hodnot veličiny X – je jasné, že pravděpodobnosti většího počtu chyb budou nižší:

$$\begin{aligned}
P(X = 3) &= \binom{10}{3} 0{,}1^3 0{,}9^7 \doteq 0{,}0574;\\
P(X = 4) &= \binom{10}{4} 0{,}1^4 0{,}9^6 \doteq 0{,}0112;\\
P(X = 5) &= \binom{10}{5} 0{,}1^5 0{,}9^5 \doteq 0{,}0015;\\
P(X = 6) &= \binom{10}{6} 0{,}1^6 0{,}9^4 \doteq 0{,}0001;\\
P(X = 7) &= \binom{10}{7} 0{,}1^7 0{,}9^3 \doteq 0{,}0000;
\end{aligned}$$

$$
\begin{aligned}
P(X=8) &= \binom{10}{8} 0{,}1^8 0{,}9^2 \doteq 0{,}0000;\\
P(X=9) &= \binom{10}{9} 0{,}1^9 0{,}9^1 \doteq 0{,}0000;\\
P(X=10) &= \binom{10}{10} 0{,}1^{10} 0{,}9^0 \doteq 0{,}0000;
\end{aligned}
$$

Ve skutečnosti nejsou uvedené pravděpodobnosti přímo rovny nule – například poslední, nejmenší z nich, je rovna 10^{-10}, což je devět nul za desetinnou čárkou následováno jednou jedničkou. ALE pro naše praktické výpočty (= zaokrouhlování na čtyři desetinná místa) je posledních několik pravděpodobností

už rovno nule.

Nyní, pokud jsme počítali dobře, součet všech jedenácti hodnot pravděpodobnosti musí být roven jedné až na zaokrouhlovací chybu, která většinou ovlivní poslední uvažované desetinné místo.

Můžeme nyní shrnout vlastnosti veličiny, kterou lze popsat binomickým rozdělením pravděpodobnosti:

- **Každý z dílčích experimentů lze popsat dvěma výsledky – „úspěch“ nebo „neúspěch“.**
- **Jednotlivá opakování dílčího experimentu jsou navzájem nezávislá (tj. pravděpodobnost „úspě-**

chu" při každém opakování dílčího experimentu je stále stejná a je rovna číslu p).

- **Veličina X = počet „úspěchů" při n opakováních dílčího experimentu.**
- **Veličina X nabývá hodnot z množiny $\{0, 1, 2, \ldots, n\}$, a to s pravděpodobnostní funkcí**

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

- **Pro střední hodnotu a rozptyl binomicky rozdělené veličiny platí**

$$EX = np, \quad DX = np(1 - p)$$

(a tedy směrodatná odchylka $\sigma = \sqrt{DX} = \sqrt{np(1-p)}$).

Konkrétně v našem příkladu lze odhadnout, že při přenosu deseti bitů dojde průměrně k $EX = 10 \cdot 0{,}1 = 1$ chybě. Hodnotu DX už nelze dobře odhadnout úvahou – v našem příkladu $DX = 10 \cdot 0{,}1 \cdot 0{,}9 = 0{,}9$, a tedy $\sigma = \sqrt{0{,}9} \doteq 0{,}9487$. Pokud bychom uplatnili hrubé pravidlo šesti sigma, tak většina měření veličiny X v našem příkladu (= většina počtu chyb při přenosu deseti bitů) bude ležet v intervalu

$$(0{,}9 - 3\sigma; 0{,}9 + 3\sigma) = 0{,}9 \pm 2{,}846 = (-1{,}946; 3{,}746),$$

tedy v naprosté většině případů dojde k nule až třem chybám.

9.3 Geometrické rozdělení pravděpodobnosti

Podívejme se na další příklad, kde některé rysy zadání jsou stejné jako v předchozím oddílku, ale budeme měřit hodnoty jiné veličiny.

Příklad 9.2. Pravděpodobnost chyby při přenosu jednoho bitu je $p = 0{,}1$. Také předpokládejme, že přenosy jednotlivých bitů na sobě nezávisí (tj. pravděpodobnost chybně přeneseného bitu se nemění). Veličina X udává, kolik bitů digitálního přenosu bude přeneseno, dokud nedojde k první chybě (včetně bitu s touto chybou). Určete pravděpodobnost, že nastane $X = 5$.

Řešení: Nyní je pořadí chyby (N) a správnosti (A) při přenosu přesně dáno, což lze zaznačit sekvencí znázorňující situaci, jejíž pravděpodobnost počítáme:

$$P(X=5)=P(AAAAN)=0{,}9^4\cdot 0{,}1=0{,}06561.$$

Obecně veličina X může nabývat hodnot $1,2,\dots$ (teoreticky až do nekonečna, jenže se stále menší pravděpodobností), protože nás zajímají pouze sekvence, kde všechny znaky jsou A (= správný přenos), a až poslední je N (= chyba při přenosu), lze vlastně psát pravděpodobnostní funkci geometrického rozdělení:

$$P(X=k)=(1-p)^{k-1}\cdot p \quad \textbf{pro } k=1,2,\dots$$

Název geometrické rozdělení pochází z toho, že uspořádány do nekonečné posloupnosti, uvedené pravděpodobnosti vytvářejí geometrickou posloupnost, kde každý další člen této řady je $(1-p)$-násobkem předchozího členu (POZOR – je zde kolize pojmů, která v historii nikomu nevadila, ale při krátkém studiu tématu pravděpodobnost jí může člověk neznalý být zaskočen:

- **„geometrická pravděpodobnost“ je pravděpodobnost spojitá, kdy jednotlivé výsledky nastávají se stejnou frekvencí a je jich nespočetně nekonečně mnoho a pravděpodobnost souvisí s geometrickou mírou množiny příznivých případů (s délkou, obsahem, objemem – viz přednáška 7), kdežto**

- **„geometrické rozdělení pravděpodobnosti“ je diskrétní rozdělení v případě, kdy jednotlivé výsledky veličiny nastávají různě často, podle právě uvedeného vzorce, tedy příslušné pravděpodobnosti tvoří geometrickou posloupnost).**

Využitím vzorců pro střední hodnotu a ropztyl v diskrétním případě lze určit EX a DX u geometricky rozdělené veličiny (na základě jistých znalostí o sečítání nekonečných řad), ale studenti si mohou zapsat oficiálně povolený výsledek tohoto výpočtu

$$EX = \frac{1}{p}, \quad DX = \frac{1-p}{p^2}.$$

V našem příkladu

$$EX = \frac{1}{0{,}1} = 10, \quad DX = 90, \quad \sigma = \sqrt{DX} \doteq 9{,}4868,$$

tj. většina měření veličiny X bude ležet v intervalu $10 \pm 3 \cdot 9{,}4868 = 10 \pm 28{,}46 \doteq (0; 38{,}46)$**.**

9.4 Negativně binomické rozdělení pravděpodobnosti

Další příbuzné rozdělení, které má svůj název, je tzv. negativně (= záporně) binomické rozdělení pravděpodobnosti. Název už napovídá, že se jedná o rozdělení, které nějak souvisí s binomickým rozdělením. A ono souvisí s oběma předchozími oddílky. V podobné situaci měříme opět jinou veličinu.

Příklad 9.3. Pravděpodobnost chyby při přenosu jednoho bitu je $p = 0{,}1$. Také předpokládejme, že přenosy jednotlivých bitů na sobě nezávisí (tj. pravděpodobnost chybně přeneseného bitu se nemění). Veličina X udává, kolik bitů digitálního přenosu bude

přeneseno, dokud nedojde ke čtvrté chybě (včetně bitu s touto chybou). Určete pravděpodobnost, že nastane $X = 10$.

Řešení: Zde je řešení analogické příkladu 9.2, ovšem máme jen tu jistotu, že poslední znak v sekvenci popisující naši situaci je N (chybný přenost bitu); a víme také, že toto N je čtvrtým N v této sekvenci – ovšem pozice předchozích N mohou být libovolné, jaké jen si umíme představit. Tedy

$$\begin{aligned} P(X=10) &= P(NNNAAAAAAN) + P(NNANAAAAAN) + \cdots \\ &\quad + P(AAAAANANNN) + P(AAAAAANNNN). \end{aligned}$$

Čili pokud máme jisté, že posledním znakem v sekvenci je N, tak výskyt předchozích tří N na (10-1) pozicích je opět určen kombinačním číslem, čili sečteme pravděpodobnosti všech možných variant sekvencí, které mohou nastat – každá z těchto variant je stejně pravděpodobná ($0{,}9^6 \cdot 0{,}1^4$), jejich počet určíme kombinačním číslem

$$P(X = 10) = 0{,}9^6 \cdot 0{,}1^4 \cdot \binom{9}{3} \doteq 0{,}004464$$

(kombinační číslo udává, kolika způsoby při prvních devíti přenosech dojde ke třem chybám – výběr tří pozic z devíti).

Obecně veličina X udávající, kolik opakování dílčího experimentu musíme provést, abychom se „dočkali“ k-té „chyby“, nabývá hodnot z množiny $\{k, k+1, k+2, \ldots\}$ a

$$P(X = x) = p^k \cdot (1-p)^{x-k} \cdot \binom{x-1}{k-1},$$

dále lze odvodit, že

$$EX = \frac{k}{p}, \quad DX = \frac{k(1-p)}{p^2}.$$

Tedy v našem příkladu bychom dosáhli čtyř chyb přibližně po $EX = \frac{4}{0{,}1} = 40$ přenosech bitu, většina měření veličiny X by ležela v intervalu ($DX = 37,9472$)

$$40 \pm 3 \cdot \sigma = 40 \pm 3 \cdot 6{,}16 = 40 \pm 18{,}48.$$

9.5 Hypergeometrické rozdělení pravděpodobnosti

Uvažujme situaci, kdy z 850 vyrobených součástek jich 50 neodpovídá požadavkům zákazníka. Dvě součástky jsou náhodně vybrány z této denní produkce. Označme jevy

- A = první vybraná součástka neodpovídá požadavkům zákazníků;
- B = druhá vybraná součástka neodpovídá požadavkům zákazníků.

Na rozdíl ode všech předchozích typů rozdělení se nyní setkáváme s příkladem, kdy výběr druhé součástky je závislý na výběru předchozí, první sou-

částky (jev B je závislý na jevu A):

$$P(A) = \frac{50}{850}, \quad P(A \cap B) = P(A) \cdot P(B|A) = \frac{50}{850} \cdot \frac{49}{849}$$

(pravděpodobnost $P(B|A)$ **závisí na tom, zda první vybraná součástka odpovídala požadavkům kvality či nikoli, protože se mění počet „příznivých případů“ i počet všech případů – před výběrem druhé součástky jsme tu první nevraceli zpět do sady; tj. jedná se o tzv. výběr bez vracení).**

Všimněte si, že kdyby se každá vybraná součástka vrátila zpět, tak pravděpodobnost výběru nekvalitní součástky je stále stejná a je rovna $\frac{50}{850}$**, takže počet takto vybíraných nekvalitních součástek by bylo**

možné popsat rozdělením binomickým. V tomto příkladu je tedy klíčové, že vybranou součástku NEVRACÍME ZPĚT do sady denní výroby před výběrem další součástky.

Příklad 9.4. V právě popsané situaci výběru dvou součástek BEZ VRACENÍ ze sady 850 součástek, z nichž je 50 nekvalitních, veličina X měří počet nekvalitních součástek z těchto dvou vybraných. Určete její rozdělení pravděpodobnosti (tj. určete, jakých hodnot nabývá a s jakou pravděpodobností).

Řešení: $P(X = 0)$ **znamená, že obě vybrané součástky jsou kvalitní (tj. počet nekvalitních s. = 0):**

$$P(X = 0) = \frac{800}{850} \cdot \frac{799}{849} \doteq 0{,}886.$$

Při výpočtu $P(X = 1)$ **musíme sečít pravděpodobnosti dvou navzájem vylučujících se situací, kdy při výběru dvou součástek z daného počtu je jedna kvalitní a druhá ne:**

$$P(X = 1) = P(AN) + P(NA) = \frac{800}{850} \cdot \frac{50}{849} + \frac{50}{850} \cdot \frac{49}{849} \doteq 0{,}111.$$

A konečně, pravděpodobnost $P(X = 2)$ **jsme už počítali, ale neoznačili jako** $P(X = 2)$**, takže ještě jednou**

a pořádně:

$$P(X = 2) = \frac{50}{850} \cdot \frac{49}{849} \doteq 0{,}003.$$

Pokud se jedná o matematický popis náhodné veličiny a uvedené pravděpodobnosti vyčerpávají všechny možnosti, kterých náhodná veličina nabývá, jejich součet se musí rovnat jedné – krátce lze zkontrolovat, že to platí: $p(0) + p(1) + p(2) = 1$**.**

Pravděpodobnosti v uvedeném příkladu je možné počítat i jiným způsobem, a sice pomocí klasické výpočetní pravděpodobnosti a kombinačních čísel. Například $P(X = 0)$ znamená, že z uvedených 850 součástek byly vybrány dvě, takže $\binom{850}{2}$ je počet všech možných výběrů. Počet kvalitních výběrů, kde není žádná nekvalitní součástka, je $\binom{800}{2}$ (= dvě součástky byly vybrány z těch osmi set kvalitních). Tedy

$$P(X = 0) = \frac{\binom{800}{2}}{\binom{850}{2}} \doteq 0{,}886.$$

Podobně $P(X = 1)$ znamená určit pravděpodobnost, kdy „příznívé“ případy v čitateli spočívají v tom, že jednu součástku vybereme z těch nekvalitních 50,

kdežto druhou vybereme z těch kvalitních 800, a to vše popisuje jednu situaci, takže uvedená dvě kombinační čísla se mezi sebou násobí:

$$P(X=1)=\frac{\binom{50}{1}\cdot\binom{800}{1}}{\binom{850}{2}}\doteq 0{,}111.$$

A nakonec

$$P(X=2)=\frac{\binom{50}{2}}{\binom{850}{2}}\doteq 0{,}003.$$

Celkem vzato, <u>hypergeometrické rozdělení</u> popisuje následující situaci:

- **Sada N objektů obsahuje K objektů s požadovanou vlastností a $N-K$ objektů, které požadova-**

nou vlastnost nemají (platí $K \leq N$).

- **Ze sady náhodně vybereme vzorek n objektů (tedy $n \leq N$). Náhodná veličina X = počet objektů z těch n vybraných, které mají danou požadovanou vlastnost.**

- **Veličina X může nabývat celočíselných hodnot z intervalu $max\{0, n+K-N\}$ (= nabývá hodnoty 0, pokud n není větší než $N-K$) až $min\{K, n\}$ (nabývá hodnoty n, pokud n není větší než K), pravděpodobnostní funkce je daná vztahem**

$$P(X = x) = \frac{\binom{K}{x} \cdot \binom{N-K}{n-x}}{\binom{N}{n}}.$$

- **Lze odvodit, že pro EX a DX hypergeometrického rozdělení platí vztahy**

$$EX = np, \quad DX = np(1-p)\frac{N-n}{N-1}, \qquad \textbf{kde } p = \frac{K}{N}$$

(číslo p se interpertuje jako procento úspěšnosti v sadě N objektů). Člen $\frac{N-n}{N-1}$ ve výpočtu rozptylu se občas označuje jako korekční faktor konečné populace (pokud totiž N se blíží nekonečnu, hypergeometrické rozdělení splývá více a více s rozdělením binomickým – pro rostoucí N a K v konstatním poměru p výběr jednoho prvku ze sady N prvků stále méně ovlivňuje pravděpodobnost výběru druhého prvku, tj. lze říci, že pro velká

N, K **lze rozdíly v pravděpodobnostech binomického a hypergeometrického rozdělení zanedbat).**

9.6 Poissonovo rozdělení pravděpodobnosti

Dalším důležitým z těch naprosto základních typů diskrétních rozdělení je Poissonovo rozdělení pravděpodobnosti. Protože toto rozdělení souvisí velmi úzce s jistým spojitým rozdělením pravděpodobnosti, odložíme je na příští přednášku.

9.7 Otázky k opakování

Z dnešní přednášky vyvstávají čtyři otázky k opakování pro studenty, jejichž odpovědi je záhodno mít s sebou na písemky, a obsah minimálně otázky číslo 13 je nutné znát zpaměti:

otázka č. 13. **Binomické rozdělení – význačné diskrétní rozdělení číslo 1: V jaké situaci jej lze použít? Co měříme v proměnné** X**? Jakých hodnot veličina nabývá a s jakou pravděpodobností? Jaká je EX a DX?**

otázka č. 14. **Geometrické rozdělení – význačné diskrétní rozdělení číslo 2: podotázky viz ot. 13.**

otázka č. 15. **Záporně binomické rozdělení – význačné diskrétní rozdělení číslo 3: podotázky viz ot. 13.**

otázka č. 16. **Hypergeometrické rozdělení – význačné diskrétní rozdělení číslo 4: podotázky viz ot. 13.**

Literatura

[1] **Fajmon, B., Růžičková, I.: Matematika 3. Skriptum FEKT 2003 (identifikační číslo v informačním systému VUT: MAT103).** `http://www.rozhovor.cz/souvislosti/matematika3.pdf`.

[2] **Hlavičková, I., Hliněná, D.: Sbírka úloh z pravděpodobnosti. Skriptum FEKT 2008.**