5 DISKRÉTNÍ ROZDĚLENÍ PRAVDĚPODOBNOSTI

Čas ke studiu kapitoly: 120 minut

Cíl: Po prostudování tohoto odstavce budete umět:

- charakterizovat hypergeometrické rozdělení
- charakterizovat Bernoulliho pokusy a z nich odvozené jednotlivé typy diskrétních rozdělení: binomické, geometrické, negativně binomické
- charakterizovat Poissonův proces a z něj vycházející Poissonovo rozdělení
- popsat vzájemnou souvislost mezi diskrétními rozděleními

Výklad:

Rozdělení náhodné veličiny X je předpis, kterým definujeme pravděpodobnost jevů, jež lze touto náhodnou veličinou popsat. U diskrétní náhodné veličiny je tímto předpisem (rozdělením) většinou pravděpodobnostní funkce, rozdělení spojité náhodné veličiny je dáno distribuční funkci, popř. hustotou pravděpodobnosti.

Existuje mnoho typů diskrétních náhodných veličin. My si nyní shrneme základní poznatky o těch nejběžnějších.

5.1 Hypergeometrická náhodná veličina

Hypergeometrické rozdělení je základním pravděpodobnostním rozdělením při výběru bez vracení., který spočívá v tom, že náhodně vybrané prvky nevracíme zpět do základního souboru. Jednotlivé pokusy jsou pak závislé (pravděpodobnost výskytu jevu A v určitém pokusu závisí na výsledcích v předcházejících pokusech).

Předpokládejme, že v souboru **N** prvků je **M** prvků s danou vlastností a zbylých (**N-M**) prvků tuto vlastnost nemá. Postupně vybereme ze souboru **n** prvků, z nichž žádný nevracíme zpět. Nadefinujeme-li náhodnou veličinu X jako:

X … počet prvků se sledovanou vlastností ve výběru n prvků,

Pak má tato náhodná veličina hypergeometrické rozdělení s parametry N, M, n, což značíme:

$$X \rightarrow H(N;\mathrm{M};\mathrm{n})$$

Pravděpodobnostní funkce:

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}; \quad \text{pro } \max(\mathrm{n} - \mathrm{N} + \mathrm{m};0) \le \mathrm{k} \le \min(\mathrm{M};\mathrm{n})$$

Střední hodnota: $EX = n.\dfrac{M}{N}$

Rozptyl: $DX = n.\dfrac{M}{N}\left(1-\dfrac{M}{N}\right)\left(\dfrac{N-n}{N-1}\right)$

Hypergeometrické rozdělení hraje významnou roli při statistické kontrole jakosti v případech, kdy zkoumáme jakost malého počtu výrobku nebo když kontrola má ráz destrukční zkoušky (tj. výrobek je při zkoušce zničen).

Průvodce studiem:

- ***Odvození pravděpodobnostní funkce:***

Definice pravděpodobnostní funkce hypergeometrického rozdělení vychází z klasické definice pravděpodobnosti: počet příznivých možností ku počtu všech možností.

N (celkový počet prvků)

M (počet prvků s danou vlastností) **(N-M)** (počet prvků bez dané vlastností)

Počet všech možností*: vybíráme n prvků z N prvkové množiny, bez ohledu na pořadí , tj. jde o kombinace bez opakování n-tého řádu z N prvků*

$$C_n(N) = \binom{N}{n}$$

Počet příznivých možností: *vybíráme k prvků z M bez ohledu na pořadí (k prvků má mít danou vlastnost) a zároveň vybíráme (n-k) prvků z (N-M) bez ohledu na pořadí (zbylé prvky z vybírané* n-tice (tj, (n-k) prvků) *danou vlastnost mít nemají). Na základě kombinatorického pravidla o součinu můžeme tvrdit, že počet příznivých možností je:*

$$C_k(M) \cdot C_{n-k}(N-M) = \binom{M}{k}\binom{N-M}{n-k}$$

A proto na základě klasické definice pravděpodobnosti můžeme nadefinovat pravděpodobnostní funkci hypergeometrické náhodné veličiny:

$$P(X = k) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}; \quad \mathrm{pro}\ \max(\mathrm{n} - \mathrm{N} + \mathrm{m}; 0) \le \mathrm{k} \le \min(\mathrm{M}; \mathrm{n})$$

Řešený příklad:

Mezi 200 vajíčky určenými pro prodej v jisté maloobchodní prodejně je 50 vajíček prasklých. Jaká je pravděpodobnost, že vybereme-li si náhodně 20 vajec, bude 8 z nich prasklých?

Řešení:

Jde o výběr bez vracení (vybrané vajíčko nevracíme zpět), jednotlivé pokusy jsou závislé.

Nadefinujeme-li si náhodnou veličinu X jako:

X … počet prasklých vajíček mezi 20-ti vybranými

pak má tato náhodná veličina hypergeometrické rozdělení s parametry: N=200; M=50; n=20

$$X \to H(200;50;20)$$

200 (celkový počet vajec)

50 (počet prasklých vajec) **150** (počet dobrých vajec)

Vzorec pro pravděpodobnostní funkci hypergeometrického rozdělení si nemusíme pamatovat, hledanou pravděpodobnost určíme z klasické definice pravděpodobnosti.

Počet všech možností: vybíráme 20 vajec z 200 vajec (bez ohledu na pořadí)

$$C_{20}(200) = \binom{200}{20}$$

Počet příznivých možností: mezi vybranými 20-ti vejci má být 8 prasklých, tj. vybíráme 8 prasklých vajec z 50-ti prasklých a zároveň 12 (20-8) dobrých vajec ze 150-ti :

$$C_8(50) \cdot C_{12}(150) = \binom{50}{8}\binom{150}{12}$$

A proto:

$$\underline{\underline{P(X = 8)}} = \frac{\binom{50}{8} \cdot \binom{150}{12}}{\binom{200}{20}} = \underline{\underline{0{,}057 = 5{,}7\%}}$$

Pravděpodobnost, že mezi 20-ti vybranými vejci bude 8 prasklých je 0,057.

Výklad:

Dále se zmíníme o diskrétních náhodných veličinách, jejichž rozdělení je definováno za předpokladu, že jde o veličiny související s Bernoulliho pokusy.

Bernoulliho pokusy:

- posloupnost **nezávislých** pokusů majících pouze 2 možné výsledky (událost nastane-nenastane; úspěch-neúspěch; popřípadě 1-0)
- pravděpodobnost výskytu události (úspěchu) p je konstantní v každém pokuse

5.2 Binomická náhodná veličina:

Binomická náhodná veličina X je definována jako počet výskytu události (úspěchů) v n Bernoulliho pokusech.

Proto aby byla binomická náhodná veličina definována, musíme znát dva její parametry: celkový počet Bernoulliho pokusů – n a pravděpodobnost výskytu události (úspěchu) v každém z pokusů – p. Pak to, že má náhodná veličina binomické rozdělení zapisujeme:

$$X \rightarrow Bi\,(n, p)$$

Pravděpodobnostní funkce binomické náhodné veličiny stanovuje jaká je pravděpodobnost, že v **n** Bernoulliho pokusech dojde ke **k** úspěchům.

$$P\,(X = k) = \binom{n}{k} p^k (1 - p)^{n-k};\; 0 \le k \le n$$

Střední hodnota: $EX = n.p$

Rozptyl: $DX = n.p.(1 - p)$

5.2.1 Vztah hypergeometrického a binomického rozdělení:

Jestliže rozsah N je velký a n a M/N se nemění, blíží se hypergeometrické rozdělení binomickému. To znamená, že pro velká N můžeme zanedbat rozdíl mezi výběrem bez vracení a s vracením. V praxi se rozhodujeme podle hodnoty tzv. **výběrového poměru (n/N)**. Je-li tento poměr menší než 0,05, lze hypergeometrické rozdělení nahradit binomickým s parametry n a (M/N).

$$\left(\frac{n}{N} < 0{,}05\right) \Rightarrow \left[H(N; \mathrm{M}; \mathrm{n}) \rightarrow \mathrm{Bi}\left(\mathrm{n}; \frac{\mathrm{M}}{\mathrm{N}}\right)\right]$$

Průvodce studiem:

Tento průvodce studiem je určen pro zájemce o hlubší pochopení studované látky.

- ***Odvození vztahů pro výpočet střední hodnoty a rozptylu:***

$$\underline{\underline{EX}} = \sum_{k=0}^{n} k \cdot P(X=k) = \sum_{k=0}^{n} k.\frac{n!}{(n-k)!.k!}.p^k.(1-p)^{n-k} = \sum_{k=0}^{n} k.\frac{n.(n-1)!}{(n-k)!.(k.(k-1)!)}.\left(p.p^{k-1}\right).(1-p)^{n-k} =$$

$$= np\sum_{k=0}^{n}\frac{(n-1)!}{(n-k)!(k-1)!}.p^{k-1}.(1-p)^{n-k} = np.\sum_{k=0}^{n}\binom{n-1}{k-1}.p^{k-1}.(1-p)^{n-k} = \underline{\underline{np}}$$

$$EX^2 = \sum_{k=0}^{n} k^2 \cdot P(X=k) = \sum_{k=1}^{n} k^2 \cdot P(X=k) = \sum_{k=1}^{n}(k^2-k+k)\cdot\frac{n!}{(n-k)!\cdot k!}\cdot p^k\cdot(1-p)^{n-k} =$$

$$= \sum_{k=1}^{n} k(k-1).\frac{n!}{(n-k)!\cdot k!}\cdot p^k\cdot(1-p)^{n-k} + \sum_{k=1}^{n} k.\frac{n!}{(n-k)!\cdot k!}\cdot p^k\cdot(1-p)^{n-k} =$$

$$= \left[\sum_{k=2}^{n} k(k-1).\frac{n.(n-1).(n-2)!}{(n-k)!\cdot(k.(k-1).(k-2)!)}\cdot(p^2.p^{k-2})\cdot(1-p)^{n-k}\right] + EX =$$

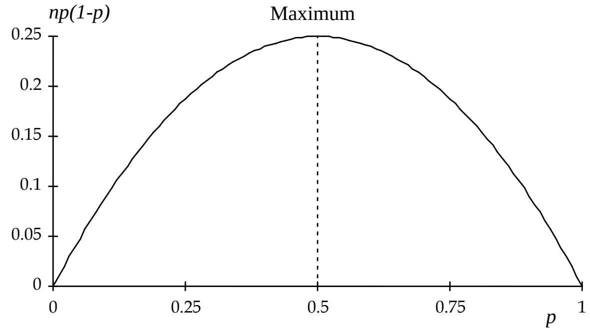
$$= \left[n.(n-1).p^2.\sum_{k=2}^{n}\frac{(n-2)!}{(n-k)!\cdot(k-2)!}\cdot p^{k-2}\cdot(1-p)^{n-k}\right] + EX = n.(n-1).p^2 + EX =$$

$$= n.(n-1).p^2 + n.p = (np)^2 - np^2 + np$$

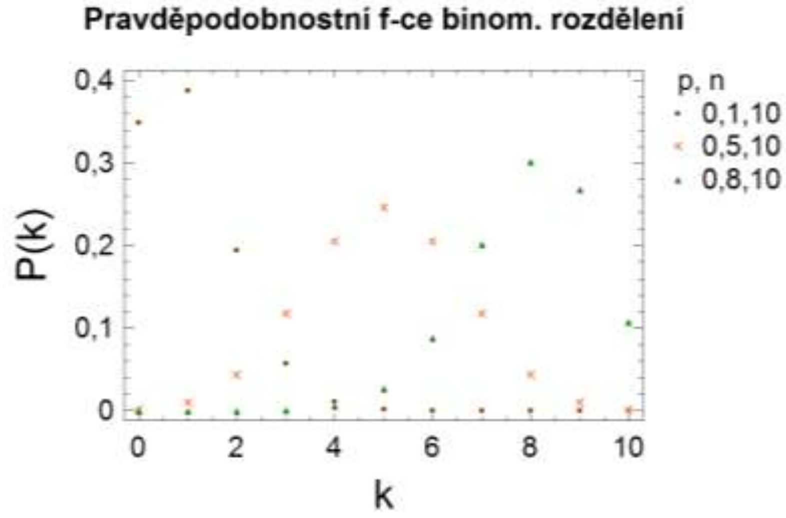
$$\underline{\underline{DX}} = EX^2 - (EX)^2 = \left[(np)^2 - np^2 + np\right] - (np)^2 = -np^2 + np = np(-p+1) = \underline{\underline{n.p.(1-p)}}$$

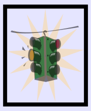
- *Z definice lokálních extrémů funkce lze jednoduše odvodit že rozptyl nabývá svého maxima pro p=0,5. To lze rovněž ukázat grafickým znázorněním funkce DX = f(p). (Jde o kvadratickou funkci, která nabývá nulových hodnot pro p = 0 a p = 1)*

- *Některé příklady demonstrace pravděpodobnostní funkce binomického rozdělení pro* **n =**



10 pokusů jsou znázorněny na následujícím obrázku. Všimněme si, že pokud p roste, rozdělení se posouvá k vyšším hodnotám na x-ové ose.





Řešený příklad:

Předpokládejme, že pravděpodobnost narození dívky je 0,49. Jaká je pravděpodobnost, že v rodině s 8 dětmi jsou:

a) právě 3 dívky
b) více než 2 dívky
c) méně než 3 dívky

Řešení:

Považujeme-li narození dítěte za náhodný pokus, pak studovanou náhodnou veličinou X je počet dívek v rodině s 8 dětmi.

Předpokládejme, že náhodné pokusy jsou nezávislé, tj. že znalost pohlaví prvního narozeného dítěte neovlivní pravděpodobnost narození dítěte určitého pohlaví při dalším „pokusu“, a mají pouze 2 možné výsledky (dívka, chlapec). Pak můžeme náhodnou veličinu X považovat za binomickou (určuje počet úspěchů (narození dívky) v n (8) pokusech, přičemž pravděpodobnost úspěchu je v každém pokusu konstantní (0,49).

X … počet dívek v rodině s 8 dětmi

$X \to Bi\,(n,p)$, tj. $X \to Bi\,(8;0{,}49)$

Rozdělení binomické náhodné veličiny: $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}$

Parametry binomického rozdělení z tohoto příkladu:

náhodný pokus	**úspěch**	**neúspěch**	**počet pokusů**	**pravděpodobnost úspěchu**	**počet úspěchu**
			n	p	k
narození dítěte	dívka	chlapec	8	0,49	

ada) k = 3

$$\underline{\underline{P(X=3)}} = \binom{8}{3}(0{,}49)^3(1-0{,}49)^{8-3} = \frac{8!}{5!.3!}(0{,}49)^3(0{,}51)^5 = \underline{\underline{0{,}23 = 23\%}}$$

adb) k > 2; tj. k = 3; 4; 5; 6; 7; 8

$$P(X>2) = P(\mathrm{X}=3) + P(\mathrm{X}=4) + P(\mathrm{X}=5) + P(\mathrm{X}=6) + P(\mathrm{X}=7) + P(\mathrm{X}=8) =$$

$$= \sum_{k=3}^{8}\binom{8}{k} 0{,}49^k (0{,}51)^{8-k}$$

Vzhledem k tomu, že tento výpočet je poněkud zdlouhavý, pokusíme se hledanou pravděpodobnost najít pomocí pravděpodobnosti doplňku.

$$\underline{\underline{P(X>2)}} = 1 - P(X \le 2) = 1 - [P(\mathrm{X}=0) + P(\mathrm{X}=1) + P(\mathrm{X}=2)] = 1 - \sum_{k=0}^{2}\binom{8}{k} 0{,}49^k (0{,}51)^{8-k} =$$

$$= 1 - 0{,}16 = \underline{\underline{0{,}84 = 84\%}}$$

adc) k < 3; tj. k = 0; 1; 2

$$\underline{\underline{P(X<3)}} = [P(\mathrm{X}=0) + P(\mathrm{X}=1) + P(\mathrm{X}=2)] = \sum_{k=0}^{2}\binom{8}{k} 0{,}49^k (0{,}51)^{8-k} = \underline{\underline{0{,}16 = 16\%}}$$

Výklad:

5.3 Alternativní náhodná veličina

Alternativní náhodná veličina X je speciální typem binomické náhodné veličiny pro jeden pokus (n = 1). Konáme náhodný pokus, při němž k výskytu události (úspěchu) dojde s pravděpodobnosti **p**. Tento náhodný pokus může mít pouze dva možné výsledky (úspěch, neúspěch).

Proto aby byla alternativní náhodná veličina definována, musíme znát pouze pravděpodobnost výskytu události (úspěchu) v každém z pokusů – **p**. Pak to, že má náhodná veličina alternativní rozdělení zapisujeme:

$$X \rightarrow A(p)$$

Pravděpodobnostní funkce alternativní náhodné veličiny stanovuje jaká je pravděpodobnost, že při pokusu dojde k úspěchu či neúspěchu.

$$P(X=1)=p$$
$$P(X=0)=1-p$$

Střední hodnota: $EX = p$
Rozptyl: $DX = p.(1-p)$

- **Odvození vztahů pro výpočet střední hodnoty a rozptylu:**

x_i	1	0
$P(X=x_i)$	p	1-p

$$\underline{\underline{EX}} = \sum_{i=1}^{2} x_i.P(X=x_i) = 1.p + 0.(1-p) = \underline{\underline{p}}$$

$$EX^2 = \sum_{i=1}^{2} x_i^2.P(X=x_i) = 1^2.p + 0^2.(1-p) = p$$

$$DX = EX^2 - (EX)^2$$
$$\underline{\underline{DX}} = p - (p)^2 = \underline{\underline{p(1-p)}}$$

5.4 Geometrická náhodná veličina:

Geometrická náhodná veličina X je definována jako **počet Bernoulliho pokusů do prvního výskytu události (úspěchu), včetně něj**. Zároveň se na ni můžeme dívat jako na speciální případ negativně binomické náhodné veličiny (pro k = 1), kterou si nadefinujeme v následujícím odstavci.

POZOR!!!!

Definice geometrické náhodné veličiny není jednoznačná. V některých publikacích (statistických softwarech) se můžeme setkat s tím, že 1. výskyt události se do počtu pokusů do 1. výskytu nezahrnuje. Pak se samozřejmě liší i příslušné pravděpodobnostní funkce, střední hodnoty a rozptyly. Pokud určujeme konkrétní hodnotu pravděpodobnostní (distribuční) funkce za pomoci statistického software, je nutné ověřit si jaká definice byla použita a podle toho modifikovat vstupní údaje pro požadovaný výpočet.

Proto aby byla geometrická náhodná veličina definována, musíme znát pouze pravděpodobnost výskytu události (úspěchu) v každém z pokusů – **p**. Pak to, že má náhodná veličina geometrické rozdělení zapisujeme:

$$X \rightarrow G(p)$$

Pravděpodobnostní funkce geometrické náhodné veličiny stanovuje jaká je pravděpodobnost, že pro dosažení prvního úspěchu musíme provést **n** pokusů (včetně toho úspěšného).

$$P(X=n)=p\,(1-p)^{n-1}; \qquad 1\le n<\infty$$

Střední hodnota: $EX=\dfrac{1}{p}$

Rozptyl: $DX=\dfrac{1-p}{p^2}$

Průvodce studiem:

Následující část výkladu je opět věnována zájemcům o matematické pozadí používaných vztahů:

- ***Odvození vztahů pro výpočet střední hodnoty a rozptylu:***

$$\underline{\underline{EX}}=\sum_{n=1}^{\infty}n\cdot P(X=n)=\sum_{n=1}^{\infty}n\cdot p\cdot(1-p)^{n-1}=p\cdot\sum_{n=1}^{\infty}n\cdot(1-p)^{n-1}=p\cdot\frac{\partial\left(\sum\limits_{n=1}^{\infty}(1-p)^n\right)}{\partial(1-p)}=$$

$$=p\cdot\frac{\partial\left(\frac{1-p}{p}\right)}{\partial(1-p)}=p\cdot\frac{1}{p^2}=\underline{\underline{\frac{1}{p}}}$$

Poznámka: $\sum\limits_{n=1}^{\infty}(1-p)^n$ *upravujeme jako součet geometrické řady*

$$E\,X^2=\sum_{n=1}^{\infty}n^2\,p\,(1-p)^{n-1}=p\left[\sum_{n=1}^{\infty}(n^2-n+n)(1-p)^{n-1}\right]=p\left[\sum_{n=2}^{\infty}(n^2-n)(1-p)^{n-1}+\sum_{n=1}^{\infty}n(1-p)^{n-1}\right]=$$

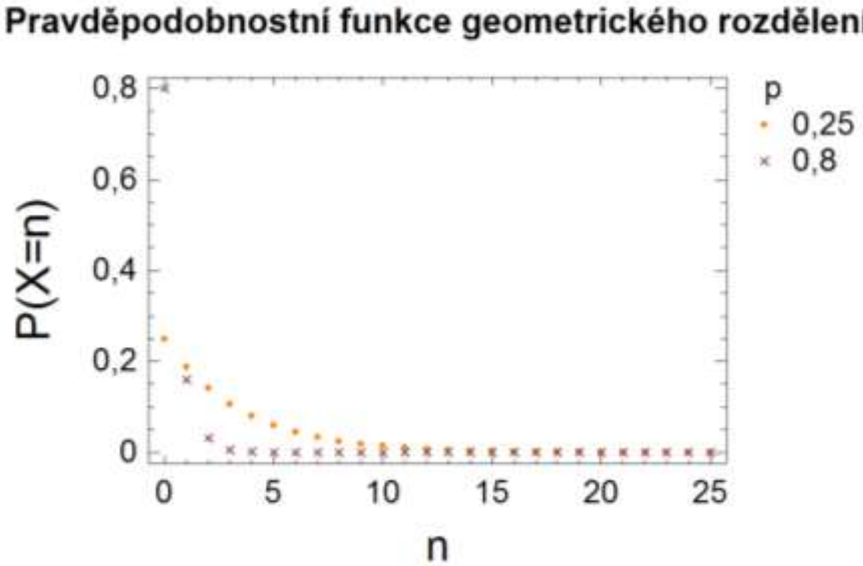
$$=p\left[\sum_{n=2}^{\infty}n.(n-1)(1-p)^{n-1}+\sum_{n=1}^{\infty}n(1-p)^{n-1}\right]=p(1-p)\sum_{n=2}^{\infty}n.(n-1)(1-p)^{n-2}+p\sum_{n=1}^{\infty}n(1-p)^{n-1}=$$

$$=p\,(1-p)\,\frac{\partial^2\left(\sum\limits_{k=1}^{\infty}(1-p)^k\right)}{\partial^2(1-p)}+p\,\frac{\partial\left(\sum\limits_{k=1}^{\infty}(1-p)^k\right)}{\partial(1-p)}=p\,(1-p)\,\frac{\partial^2\left(\frac{1-p}{p}\right)}{\partial^2(1-p)}+p\,\frac{\partial\left(\frac{1-p}{p}\right)}{\partial(1-p)}=$$

$$=p(1-p)\,\frac{2}{p^3}+p\,\frac{1}{p^2}=\frac{2(1-p)}{p^2}+\frac{1}{p}$$

$$\underline{\underline{DX}} = EX^2 - (EX)^2 = \left[\frac{2(1-p)}{p^2} + \frac{1}{p}\right] - \left(\frac{1}{p}\right)^2 = \frac{2-2p+p-1}{p^2} = \underline{\underline{\frac{1-p}{p^2}}}$$

- *Některé příklady geometrického rozdělení pro různé hodnoty p jsou ilustrovány níže. Z odvozeného vztahu pro rozptyl vyplývá, že s klesající pravděpodobností výskytu události rozptyl vzrůstá.*



Řešený příklad:

Jaká je pravděpodobnost, že proto aby nám padla na klasické kostce „6“, musíme házet:

a) právě 5x
b) více než 3x
c) Jaký je průměrný počet hodů nutných k padnutí „6“?

Řešení:

Považujeme-li za náhodný pokus hod kostkou (opakované hody tvoří Bernoulliho pokusy), pak počet hodů nutných k 1. úspěchu (padnutí „6“) je geometrickou náhodnou veličinou X s parametrem p = 1/6 (pravděpodobnost úspěchu v každém pokusu).

$$X \to G\left(\frac{1}{6}\right)$$

Pravděpodobnostní funkce geometrické náhodné veličiny je definována takto:

$$P(X=n) = p\,(1-p)^{n-1}; \qquad 1 \le n < \infty$$

Pravděpodobnost, že „6“ padne v 5. hodu určíme přímým dosazením do vztahu pro pravděpodobnostní funkci.

Poznámka: *V případě, že bychom hodnotu pravděpodobnostní funkce hledali pomocí software, který používá definici geometrické náhodné veličiny – počet pokusů před prvním úspěchem, museli bychom hledat pravděpodobnostní funkci ve „4“ (4 pokusy před prvním úspěšným)).*

ada) $\underline{\underline{P(X=5)}}=\left(\frac{1}{6}\right)\left(1-\frac{1}{6}\right)^{5-1}=\frac{1}{6}\cdot\left(\frac{5}{6}\right)^{4}=\underline{\underline{0{,}080=8{,}0\%}}$

Pravděpodobnost, že poprvé padne „6“ v 5. hodu je 8,0%.

adb)

$$\begin{aligned}\underline{\underline{P(X>3)}}&=P(X=4)+P(X=5)+\ldots=\\&=1-P(X\leq 3)=1-\left[P(X=1)+P(X=2)+P(X=3)\right]=\\&=1-\left[p+p(1-p)+p(1-p)^{2}\right]=\\&=1-\left[\frac{1}{6}+\frac{1}{6}\left(\frac{5}{6}\right)+\frac{1}{6}\left(\frac{5}{6}\right)^{2}\right]=\underline{\underline{0{,}578=57{,}8\%}}\end{aligned}$$

Pravděpodobnost, že poprvé padne „6“ nejdříve ve 4. hodu je 57,8%.

adc) $\underline{\underline{EX}}=\frac{1}{p}=\frac{1}{\frac{1}{6}}=\underline{\underline{6}}$

Do prvního padnutí „6“ musíme uskutečnit průměrně 6 hodů.

Výklad:

5.5 Negativně binomická náhodná veličina

Negativně binomická náhodná veličina X je definována jako počet Bernoulliho pokusů do k-tého výskytu události (úspěchu), včetně k-tého výskytu.

Z definice je tedy zřejmé, že se jedná o obecnější případ geometrické náhodné veličiny (geometrická náhodná veličina je speciálním případem negativně binomické náhodné veličiny pro k = 1).

POZOR!!!!

Obdobně jako u geometrické náhodné veličiny, ani v případě negativně binomické náhodné veličiny není definice jednoznačná. Někteří statistici (popř. statistický software) ji definují jako počet neúspěchů před k-tým úspěchem. Důsledek této nejednoznačnosti je stejný jako v případě geometrické náhodné veličiny. V případě srovnávacích výpočtů je vždy nutné ověřit, kterou definici autoři použili a tomu přizpůsobit další postup.

Proto aby byla negativně binomická náhodná veličina definována, musíme znát dva její parametry: celkový počet výskytu události (úspěchu) – **k** a pravděpodobnost výskytu události (úspěchu) v každém z pokusů – **p**. Pak to, že má náhodná veličina negativně binomické rozdělení zapisujeme:

$$X \rightarrow NB(k, p)$$

Pravděpodobnostní funkce negativně binomické náhodné veličiny stanovuje jaká je pravděpodobnost, že pro dosažení k výskytů události (úspěchu) musíme uskutečnit n Bernoulliho pokusů.

$$P(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k};\ k \leq n < \infty$$

Střední hodnota: $EX = \dfrac{k}{p}$

Rozptyl: $DX = \dfrac{k(1-p)}{p^2}$

5.5.1 Porovnání binomického a negativně binomického rozdělení

Ačkoliv se může na první pohled zdát, že obě rozdělení mají podobnou pravděpodobnostní funkci, existují významné rozdíly:

<u>Binomické rozdělení</u>

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}; 0 \leq k \leq n$$

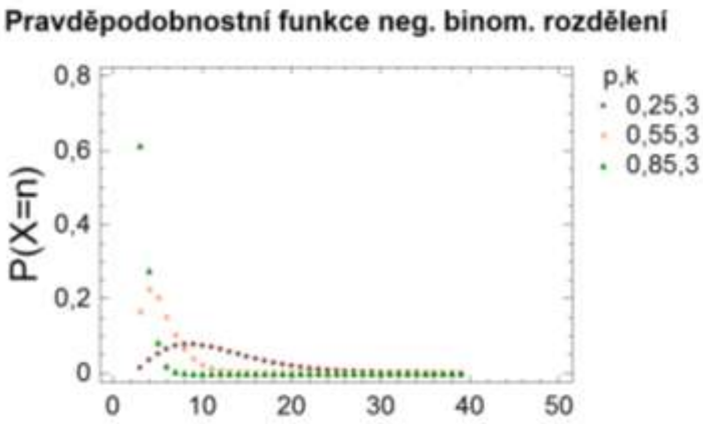
V tomto vztahu je ***k*** náhodné a ***n*** deterministické (předem známé).

<u>Negativně binomické rozdělení</u>

$$P(X = n) = \binom{n-1}{k-1} p^k (1-p)^{n-k}; k \leq n < \infty$$

V tomto vztahu je **n** náhodné a **k** deterministické (předem známé).

-



Průvodce studiem:

- ***Odvození vztahů pro výpočet střední hodnoty a rozptylu:***

Negativně binomickou náhodnou veličinu si můžeme představit jako součet nezávislých k geometrických náhodných veličin:

$$W_i \to G(p); 1 \leq i \leq k$$

potom

$$X = \sum_{i=1}^{k} W_i$$

$$E(X) = \sum_{i=1}^{k} E(W_i) = \frac{k}{p}$$

$$D(X) = \sum_{i=1}^{k} D(W_i) = \frac{k(1-p)}{p^2}$$

- *Následující obrázek ilustruje některé příklady NB rozdělení pro k = 3 a různé hodnoty p. Pokud p je v blízkosti hodnoty 0.5, NB rozdělení má jednoduchý modus poblíž hodnoty 5. Tento modus se vzdaluje směrem od počátku a přitom se jeho pravděpodobnostní hodnota zmenšuje, pokud p klesá, což znamená růst rozptylu pro klesající p. NB rozdělení má podobný tvar jako geometrické rozdělení pro velké hodnoty p.*

Řešený příklad:

Jaká je pravděpodobnost, že pro nalezení 3 dárců krevní skupiny A+, budeme muset vyšetřit:

a) právě 10 osob neznajících svou krevní skupinu
b) více než 9 osob neznajících svou krevní skupinu
c) více než 7 a méně než 12 osob neznajících svou krevní skupinu

Řešení:

Předpokládejme, že máme 8 krevních skupin (A+, A-, B+, B-, AB+, AB-, 0+, 0-), které se vyskytují se stejnou pravděpodobností. Za náhodný pokus budeme považovat vyšetření jedné osoby (2 možné výsledky - má krevní skupinu A+ (úspěch), nemá krevní skupinu A+). Definujeme-li si náhodnou veličinu X jako:

X … počet osob, které musíme vyšetřit, chceme-li najít 3 dárce s krevní skupinou A+

Pak můžeme X považovat za negativně binomickou náhodnou veličinu:

$$X \rightarrow NB(3, \frac{1}{8})$$

Pravděpodobnostní funkce X pak vypadá takto:

$$P(X = n) = \binom{n-1}{3-1}\left(\frac{1}{8}\right)^3 \left(1 - \frac{1}{8}\right)^{n-3} = \binom{n-1}{2}\left(\frac{1}{8}\right)^3 \left(\frac{7}{8}\right)^{n-3}; \qquad 3 \leq n < \infty$$

Nyní můžeme přistoupit k hledání konkrétních pravděpodobností:

ada) $\underline{\underline{P(X = 10)}} = \binom{9}{2} \cdot \left(\frac{1}{8}\right)^3 \cdot \left(\frac{7}{8}\right)^7 = \underline{\underline{0{,}028 = 2{,}8\%}}$

adb)

$$\underline{\underline{P(X > 9)}} = \mathrm{P}(9) + \mathrm{P}(10) + \mathrm{P}(11) + \ldots =$$

$$= 1 - P(X \leq 9) = 1 - \sum_{n=3}^{9} \binom{n-1}{2}\left(\frac{1}{8}\right)^3 \left(\frac{7}{8}\right)^{n-3} = \underline{\underline{0{,}908 = 90{,}8\%}}$$

adc)

$$\underline{\underline{P(7 < X < 12)}} = \mathrm{P}(8) + \mathrm{P}(9) + \mathrm{P}(10) + \mathrm{P}(11)$$

$$= \sum_{n=8}^{11} \binom{n-1}{2}\left(\frac{1}{8}\right)^3 \left(\frac{7}{8}\right)^{n-3} = \underline{\underline{0{,}103 = 10{,}3\%}}$$

Výklad:

- **Poissonův proces**

Poissonův proces je další z obecných modelů schémat sběru dat, který má široké využití v praxi. Lze ho chápat jako zobecnění Berhoulliho posloupnosti pokusů ve spojitém čase.

Poissonův proces **popisuje výskyt náhodných událostí na nějakém pevném časovém intervalu (popř. na vymezené prostorové oblasti - ploše)**. Obecným názvem pro takové procesy je **bodový proces**. Poissonův proces je speciálním případem bodového procesu.

U tohoto procesu musí být dodrženy dva předpoklady:

- rychlost výskytu událostí je konstantní v průběhu celého intervalu (popř. hustota výskytu je konstantní na vymezené ploše
- jednotlivé události musí být nezávislé

Rychlost výskytu události (hustotu výskytu události na ploše) **λ** je úměrná pravděpodobnosti výskytu jedné události za jednotku času.

Příklady Poissonova procesu:

- počet studentů vstupujících do budovy VŠB TUO od 8:00 do 9:00 hod.
- počet pacientů ošetřených během dopoledních ordinačních hodin
- počet mikrodefektů na zadaném vzorku materiálu, atd.

5.6 Poissonovo rozdělení pravděpodobnosti

Definujme si náhodný pokus jako Poissonův proces (nezávislé události probíhající v čase t, s rychlostí výskytu λ; popř. nezávislé události objevující se na ploše t s hustotou výskytu λ). Pokud si náhodnou veličinu X za těchto předpokladů nadefinujeme:

X … počet výskytu události v časovém intervalu t
nebo
X … počet výskytu události na ploše t

pak můžeme X považovat za náhodnou veličinu s Poissonovým rozdělením:

$$X \rightarrow Po(\lambda t)$$

Pravděpodobnostní funkce:

$$P(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}; \qquad 0 \le k < \infty$$

Střední hodnota: $EX = \lambda t$

Rozptyl: $DX = \lambda t$

Protože střední hodnota je rovna λt, můžeme tvrdit, že **parametr Poissonova rozdělení λt** je roven střednímu počtu události během časového intervalu t (popř. střednímu počtu výskytu události na ploše t).

Průvodce studiem:

A opět tu máme průvodní slovo pro zájemce o hlubší pochopení učiva:

- ***Odvození pravděpodobnostní funkce Poissonova rozdělení***

Uvažujme Poissonův proces, který je pozorován v průběhu času t. Předpokládejme, že rychlost výskytu událostí je λ. Potom pravděpodobnost výskytu událostí během intervalu (0;t) bude úměrná hodnotě λt.

Nyní rozdělíme interval délky t na n subintervalů stejné délky (t/n). Výskyt událostí v každém z těchto subintervalů bude nezávislý a pravděpodobnost výskytu událostí během jednoho tohoto malého intervalu bude úměrná hodnotě (λ.(t/n)). Pokud n je dostatečně velké číslo, pak délka intervalu (t/n) bude dostatečně malá - natolik, že pravděpodobnost výskytu více než jedné události v tomto intervalu je téměř nulová a pravděpodobnost výskytu jedné události je úměrná (λ.(t/n)).

Potom pravděpodobnostní rozdělení počtu událostí vyskytlých během celého intervalu délky t bude možno aproximovat binomickým rozdělením s parametry n a (λt /n) – za předpokladu, že (n→∞). Tedy:

$$P(X = k) = \lim_{n\to\infty}\left[\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k}\right]$$

Po úpravě dostáváme:

$$\underline{\underline{P(X = k)}} = \lim_{n\to\infty}\binom{n}{k}\left(\frac{\lambda t}{n}\right)^k\left(1-\frac{\lambda t}{n}\right)^{n-k} = \frac{(\lambda t)^k}{k!}\lim_{n\to\infty}\frac{n!}{(n-k)!n^k}\left(1-\frac{\lambda t}{n}\right)^{(n-k+k)-k}$$

$$= \frac{(\lambda t)^k e^{-\lambda t}}{k!}\lim_{n\to\infty}\frac{n!}{(n-k)!n^k} = \frac{(\lambda t)^k e^{-\lambda t}}{k!}\lim_{n\to\infty}\frac{n(n-1)(n-2)....(n-k+1)}{n^k} =$$

$$= \frac{(\lambda t)^k e^{-\lambda t}}{k!}\lim_{n\to\infty}\frac{n^k + k.n^{k-1}+...}{n^k} = \underline{\underline{\frac{(\lambda t)^k e^{-\lambda t}}{k!}}}$$

Pravděpodobnostní funkci Poissonova rozdělení tedy můžeme vyjádřit jako:

$$P(X=k)=\frac{(\lambda t)^k e^{-\lambda t}}{k!}; \qquad 0\le k<\infty$$

- ***Odvození vztahu pro výpočet střední hodnoty:***

$$\underline{\underline{E(X)}}=\sum_{k=0}^{\infty}kP(X=k)=\sum_{k=0}^{\infty}k\frac{(\lambda t)^k e^{-\lambda t}}{k!}=\lambda t\cdot e^{-\lambda t}\cdot\sum_{k=1}^{\infty}\frac{(\lambda t)^{k-1}}{(k-1)!}=\lambda t\cdot e^{-\lambda t}\cdot\sum_{l=01}^{\infty}\frac{(\lambda t)^l}{l!}=\underline{\underline{\lambda t}}$$

- ***Odvození vztahu pro výpočet rozptylu:***

$$DX=EX^2-(EX)^2$$

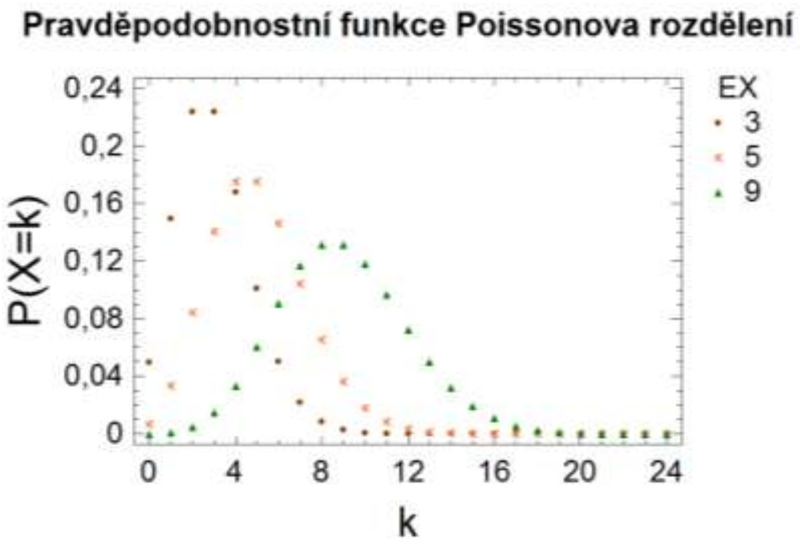
$$EX^2=\sum_{k=0}^{\infty}k^2P(X=k)=\sum_{k=0}^{\infty}(k^2-k+k)P(X=k)=\sum_{k=0}^{\infty}k(k-1)P(X=k)+\sum_{k=0}^{\infty}kP(X=k)=$$
$$=\sum_{k=2}^{\infty}k(k-1)\frac{(\lambda t)^k e^{-\lambda t}}{k!}+\mathrm{EX}=(\lambda t)^2\cdot e^{-\lambda t}\cdot\sum_{k=2}^{\infty}\frac{(\lambda t)^{k-2}}{(k-2)!}+\lambda t=(\lambda t)^2+\lambda t$$

Pro rozptyl pak dostáváme,

$$\underline{\underline{DX}}=EX^2-(EX)^2=(\lambda t)^2+\lambda t-(\lambda t)^2=\underline{\underline{\lambda t}}$$

Zajímavost tohoto rozdělení spočívá v tom, že střední hodnota je stejná jako rozptyl.

- *Následující obrázek ilustruje příklady Poissonova rozdělení pro různé hodnoty λ, při t=1.*
  Poznamenejme, že pro λ = 9 je rozdělení téměř symetrické.



Řešený příklad:

V nemocnici ABC se průměrně 30x ročně vyskytne porucha srdeční činnosti po určité operaci. Určete:

a) pravděpodobnost, že se v nemocnici ABC vyskytne příští měsíc právě 5 těchto poruch
b) pravděpodobnost, že se v nemocnici ABC vyskytne příští měsíc 2 a více těchto poruch
c) střední hodnotu a směrodatnou odchylku počtu těchto poruch během jednoho měsíce

Řešení:

Předpokládejme, že se jednotlivé poruchy srdeční činnosti po dané operaci vyskytují nezávisle na sobě, s konstantní rychlosti výskytu. Pak můžeme náhodnou veličinu

X … počet výskytu poruch srdeční činnosti během měsíce
(po dané operaci, v nemocnici ABC)

považovat za náhodnou veličinu s Poissonovým rozdělením. Její parametr – λt – určíme jako průměrný počet výskytu poruch srdeční činnosti během měsíce (střední hodnota Poissonova rozdělení je rovna λt).

$$t = 1 \text{ měsíc} \Rightarrow EX = \lambda t = \frac{30}{12} = 2{,}5 \left[mesic^{-1}\right] \Rightarrow X \to Po(2{,}5)$$

$$P(X = k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!}; \qquad 0 \le k < \infty$$

ada) Pravděpodobnost, že se v nemocnici ABC vyskytne příští měsíc právě 5 těchto poruch, určíme jednoduše dosazením do pravděpodobnostní funkce.

$$\underline{\underline{P(X = 5)}} = \frac{(2{,}5)^5 e^{-2{,}5t}}{5!} \underline{\underline{= 0{,}067 = 6{,}7\%}}$$

adb) Pravděpodobnost, že se v nemocnici ABC vyskytne příští měsíc 2 a více těchto poruch, bychom museli určit jako součet pravděpodobností pro počet výskytu (k) od 2 do ∞. Proto použijeme v tomto případě pravděpodobnost doplňku daného jevu:

$$\underline{\underline{P(X \ge 2)}} = 1 - P(X < 2) = 1 - \left[P(X = 0) + P(X = 1)\right] = 1 - \sum_{k=0}^{1} \frac{(\lambda t)^k e^{-\lambda t}}{k!} =$$
$$= 1 - \left[e^{-2{,}5} + 2{,}5e^{-2{,}5}\right] = 1 - 3{,}5e^{-2{,}5} \underline{\underline{= 0{,}713 = 71{,}3\%}}$$

adc) Střední hodnota i rozptyl náhodné veličiny X jsou rovny jejímu parametru, směrodatná odchylka je rovna odmocnině z rozptylu.

$$\underline{\underline{EX}} = DX =; \lambda t \underline{\underline{= 2{,}5}} \quad \underline{\underline{\sigma_X}} = \sqrt{DX} = \sqrt{2{,}5} \underline{\underline{\cong 1{,}6}}$$

Shrnutí:

Rozdělení náhodné veličiny X je předpis, kterým definujeme pravděpodobnost jevů, jež lze touto náhodnou veličinou popsat.

Základním rozdělením popisujícím výběry bez vracení je **hypergeometrické rozdělení**.

Název NV X	Popis	Pravděpodobnostní funkce
Hypergeometrická	Počet prvků se sledovanou vlastností ve výběru **n** prvků, který byl proveden ze základního souboru rozsahu **N** (v základním souboru má **M** prvků sledovanou vlastnost)	$P(X=k)=\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}};$ pro $\max(n - N + m;0) \le k \le \min(M;n)$

Bernoulliho pokusy:

- posloupnost **nezávislých** pokusů majících pouze 2 možné výsledky (událost nastane-nenastane; úspěch-neúspěch; popřípadě 1-0)
- pravděpodobnost výskytu události (úspěchu) p je konstantní v každém pokuse

Rozdělení diskrétní náhodné veličiny založené na Bernoulliho pokusech:

Název NV X	Popis	Pravděpodobnostní funkce	EX	DXl
Binomická	Počet úspěchů (k) v n pokusech	$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k};$ $0 \le k \le n$	np	$np(1-p)$
Alternativní	Počet úspěchů v jednom pokusu	$P(X=1)=p$ $P(X=0)=1-p$	p	$p(1-p)$
Geometrická	Počet pokusů (n) do 1. úspěchu	$P(X=n)=p(1-p)^{n-1};$ $1 \le n < \infty$	$\dfrac{1}{p}$	$\dfrac{(1-p)}{p^2}$
Negativně binomická	Počet pokusů (n) do k-tého úspěchu	$P(X=n)=\binom{n-1}{k-1}p^k(1-p)^{n-k};$ $k \le n < \infty$	$\dfrac{k}{p}$	$\dfrac{k(1-p)}{p^2}$

Poissonův proces popisuje výskyt náhodných událostí na nějakém pevném časovém intervalu (popř. na vymezené prostorové oblasti - ploše).

U tohoto procesu musí být dodrženy dva předpoklady:

- rychlost výskytu událostí je konstantní v průběhu celého intervalu (popř. hustota výskytu je konstantní na vymezené ploše
- jednotlivé události musí být nezávislé

Rozdělení diskrétní náhodné veličiny založené na Poissonově procesu:

Název NV X	Popis	Pravděpodobnostní funkce	EX	DXl
Poissonova	Počet události (k) v časovém intervalu (na ploše) (t)	$P(X=k)=\frac{(\lambda t)^k}{k!}e^{-\lambda t};$ $0\le k\le\infty$	λt	λt

Otázky

1. Co je to rozdělení pravděpodobnosti?
2. Jaká diskrétní rozdělení pravděpodobnosti znáte ?
3. Charakterizujte Bernoulliho pokusy a z nich odvozené jednotlivé typy diskrétních rozdělení
4. Odvoďte vztah pro výpočet střední hodnoty binomické náhodné veličiny.
5. Charakterizujte Poissonův proces
6. Charakterizujte náhodnou veličinu s Poissonovým rozdělením.

Úlohy k řešení

1. Pravděpodobnost úspěchu je 0.1. Určete pravděpodobnost, že do prvního úspěchu provedeme:

   a) méně než 5 pokusů
   b) více než 10 pokusů
   c) mezi 6 a 8 pokusy
   d) právě 7 pokusů.

2. Víme, že pravděpodobnost vady výrobku je 17%. Určete pravděpodobnost, že mezi 20 výrobky bude:

   a) více než 5 vadných výrobků
   b) méně než dva vadné výrobky
   c) mezi 4 a 8 vadnými výrobky
   d) právě 3 vadné výrobky

3. Kolikrát (průměrně) musíme hodit mincí, aby nám 5x padl lev?

4. Továrna produkuje integrované obvody XX. Při jedné fázi výroby dochází často k závadě, proto je 25% výrobků vadných. Jaká je pravděpodobnost, že mezi 12 integrovanými obvody budou:

   a) 4 vadné
   b) méně než 4 vadné
   c) Jaká je střední hodnota a rozptyl počtu vadných IO, budeme-li testovat 15 vzorků?
   d) Nyní uvažme, že bylo vyrobeno pouze 48 IO a my vybereme 12 z nich. Jaká je nyní pravděpodobnost, že mezi vybranými IO budou právě 4 vadné?

5. Distributor prodává knihu XY po telefonu. 12% hovorů je úspěšných (tj. objednají si knihu). Jaká je pravděpodobnost, že distributor předtím než bude úspěšný bude muset uskutečnit:

   a) 5 hovorů
   b) méně než 5 hovorů
   c) více než 8 hovorů

   Předpokládejme, že distributor musí splnit denní kvótu - prodat 10 knih.

   d) Jaká je pravděpodobnost, že distributor bude pro splnění denní kvóty potřebovat méně než 30 telefonátů?
   e) Určete střední hodnotu a rozptyl počtu telefonátů potřebných pro splnění denní kvóty.

   Uvažme nyní, že ne každý z těch, kdo si telefonicky objednají danou knihu, ji skutečně odebere. Přesněji řečeno - 65% osob objednanou knihu skutečně zaplatí. Distributor je podle této skutečnosti ohodnocen. Dostává 30,- Kč za každou objednávku a dalších 50,- Kč ve chvíli, kdy je objednávka převzata.

f) Jaká je pravděpodobnost, že výdělek distributora ve chvíli, kdy splní svou denní kvótu, bude vyšší než 500,- Kč?
g) Jaký je jeho průměrný výdělek (a směrodatná odchylka jeho výdělku) při splnění denní kvóty?

6. Celník na hranici se Slovenskem má za úkol kontrolovat projíždějící vozidla. Víme, že 25% vozidel veze kontraband a 40% z nich celník odhalí. Jaká je pravděpodobnost, že celník, předtím než objeví první vozidlo s kontrabandem, bude muset prohlédnout:

   a) 5 aut
   b) více než 10 aut
   c) Určete střední hodnotu a rozptyl počtu aut, jež musí celník prohlédnout předtím než objeví první automobil s kontrabandem.

   Nadřízený tohoto celníka vydal příkaz, že celník může jít domů poté co nalezne 5 aut s kontrabandem. Předpokládejme, že prohlédnutí jednoho auta trvá celníkovi 10 minut.

   d) Jaká je pravděpodobnost, že tento příkaz prodlouží celníkovi pracovní den (8 hodin) ?
   e) Jaká je nyní průměrná pracovní doba (a její směrodatná odchylka) celníka?

7. Bankovní úředník provádějící kontrolu návrhů půjček zjistil, že se v nich nachází 0.5 chyby na návrh. Jaká je pravděpodobnost, že úředník najde v deseti návrzích:

   a) 6 chyb
   b) více než 6 chyb
   c) ani jednu chybu.

   V 35% chyb je nutno chybu přičíst úmyslné chybné prezentaci dat.

   d) Jaký je průměrný počet chyb způsobených chybnou prezentací v celkovém množství 100 návrhů?
   e) Pokud všechny chybné návrhy vyřadíme, jaká je pravděpodobnost, že více než 2 návrhy z deseti budou vyřazeny vlivem úmyslné chybné prezentace dat?

8. Počet návštěvníků Fitness Centra VŠB je v průměru 10 na hodinu. Určete:

   a) pravděpodobnost, že v určitou hodinu je ve Fitcentru přesně 10 lidí
   b) pravděpodobnost, že v určitou hodinu je ve Fitcentru méně než 5 lidí
   c) pravděpodobnost, že v určitou hodinu je ve Fitcentru mezi 8 a 15 osobami

Řešení:

1. X … geometrická náhodná veličina
   a) 0,344
   b) 0,349
   c) 0,160
   d) 0,053

2. X … binomická náhodná veličina
   a) 0,110
   b) 0,123
   c) 0,446
   d) 0,236

3. 10x (negativně binomická náhodná veličina)

4. X … binomická náhodná veličina
   a) 0,190
   b) 0,650
   c) EX=3,75; DX=2,81
   d) (hypergeometrická náhodná veličina) 0,220

5. X … geometrická náhodná veličina
   a) 0,060
   b) 0,470
   c) 0,320
   Y … negativně binomická náhodná veličina
   d) 0,001
   e) EY=83,33; DY=611,11
   Z … binomická náhodná veličina, H = 300 + 50Z
   f) 0,910
   g) EH=625; σ_H=75

6. X … geometrická náhodná veličina
   a) 0,059
   b) 0,310
   c) EX=10; DX=90
   Y … negativně binomická náhodná veličina
   d) 0,470
   e) EY=8h 20min; σ_Y=3h 32min

7. X … Poissonova náhodná veličina
   a) 0,146

b) 0,238

c) 0,007

Y … binomická náhodná veličina

d) EY=17,5
e) 0,738

8. X … Poissonova náhodná veličina
   a) 0,125
   b) 0,029
   c) 0,731