4. Opakované pokusy a Bernoulliho schema

4.1. Poznámka: Sledujeme pravděpodobnosti výsledků opakovaného pokusu v sérii n jeho nezávislých opakování. Jednotlivé výsledky rozdělíme podle počtu výskytů sledovaného jevu. Např. počet 6 při opakovaných hodech kostkou. Počet chyb při přenosu určitého počtu symbolů.

4.2. Věta: Bernoulliho schema. Provádíme sérii n nezávislých náhodných pokusů, ve kterých nastává sledovaný výsledek, náhodný jev A, s pravděpodobností $P(A) = p,\ 0 < p < 1$. Pravděpodobnost $P_n(k)$ toho, že se v sérii vyskytne náhodný jev A právě $k-$krát, $k = 0, 1, 2, \ldots, n$ je rovna

$$(\spadesuit) \qquad P_n(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad 0 \le k \le n.$$

4.3. Příklad: Házíme $n-$ krát mincí a počítáme počet rubů. Určete pravděpodobnosti jednotlivých možných výsledků pokusu. Vyčíslete jejich hodnoty pro $n = 5$.

Řešení: Výsledky pokusu se řídí Bernoulliho schematem z odst. 4.2, kde je $p = 1 - p = \frac{1}{2}$. Sledovaný jev nastává v jednom ze dvou možných případů. Jestliže si označíme v souladu s 4.2 jako $P_n(k)$ pravděpodobnost náhodného jevu - v sérii padne $k-$ krát rub, $0 \le k \le n$, je

$$P_n(k) = \binom{n}{k} \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} = \binom{n}{k} \frac{1}{2^n}, \quad 0 \le k \le n.$$

Pro $n = 5$ je $\dfrac{1}{2^5} = \dfrac{1}{32} = 0,03125$ a po řadě je

$$\binom{5}{0} = \binom{5}{5} = 1, \qquad \binom{5}{1} = \binom{5}{4} = 5, \qquad \binom{5}{3} = \binom{5}{2} = 10.$$

Je tedy
$P_5(0) = P_5(5) = 0,03125, \qquad P_5(1) = P_5(4) = 5.0,03125 = 0,15625$
a $P_5(2) = P_5(3) = 10.0,03125 = 0,3125.$

4.4. Příklad: Házíme $n-$ krát hrací kostkou a počítáme počet šestek. Určete pravděpodobnosti jednotlivých možných výsledků pokusu. Vyčíslete jejich hodnoty pro $n = 5$.

Řešení: Výsledky pokusu se řídí Bernoulliho schematem z odst. 4.2, kde je $p = \frac{1}{6}$ a $1-p = \frac{5}{6}$. Sledovaný jev nastává v jednom ze šesti možných případů. Jestliže si označíme v souladu s 4.2 jako $P_n(k)$ pravděpodobnost náhodného jevu - v sérii padne $k-$ krát šestka, $0 \leq k \leq n$, je

$$P_n(k) = \binom{n}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k} = \binom{n}{k} \frac{5^{n-k}}{6^n}, \quad 0 \leq k \leq n.$$

Pro $n = 5$ je $\left(\frac{5}{6}\right)^5 = 0,401878$ a jestliže použijeme hodnot kombinačních čísel z příkladu 4.3 dostaneme po řadě

k	0	1	2	3	4	5
$P_5(k)$	0,40188	0,40188	0,16075	0,03215	0,00322	0,00013

Poznámka: Všimneme si, že v příkladě 4.3 je posloupnost pravděpodobností symetrická, sledovaný jev a jev opačný mají stejnou pravděpodobnost. V případě příkladu 4.4 jsou hodnoty asymetrické, pravděpodobnost opačného jevu je 5-krát větší. Větší pravděpodobnost mají hodnoty 0 a 1 pro počet šestek. Více si o rozdělení pravděpodobností řekneme v kapitole o náhodné veličině, kdy pravděpodobnosti z odst 4.2 odpovídají t.zv. *binomickému rozdělení.*

4.5. Věta: Vlastnosti čísel $P_n(k)$. Pro pravděpodobnosti $P_n(k)$ z odstavce 4.2 platí:

a) $\sum_{k=0}^{n} P_n(k) = 1$; b) $\sum_{k=0}^{n} k.P_n(k) = np$;

c) $\sum_{k=0}^{n} (k - np)^2.P_n(k) = np(1 - p)$.

Důkaz: a) Tvrzení vyplývá z toho, že uvedený součet je součtem pravděpodobností všech náhodných jevů. Jinak vyplývá tato identita z binomické věty pro $[p+(1-p)]^n = 1^n = 1$.

b) Ukážeme si obrat, pomocí kterého dokážeme uvedené tvzení. Je totiž

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

$$= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!.[(n-1)-(k-1)]!}.p^{k-1}(1-p)^{(n-1)-(k-1)} =$$

$$= np\sum_{m=0}^{n-1} \frac{(n-1)!}{m!.[(n-1)-m]!}.p^{m-1}(1-p)^{(n-1)-m} = np.$$

c) Tvrzení odvodíme obratem jako v případě b). Jenom si upravíme výraz
$(k-np)^2 = k^2 - 2knp + n^2p^2 = k(k-1) + k - 2knp + n^2p^2$.
Součet vyjádříme jako součet 4 sčítanců. Součet posledních tří vypočteme pomocí vztahů z a) a b) a v prvním provedeme krácení jako v odvození z b), jenom můžeme krátit výrazem $k(k-1)$ místo k.

Poznámka: Pravděpodobnosti $P_n(k)$ jednotlivých výsledků z Bernoulliho schematu mají určitou zákonitost rozdělení hodnot. Přestavují t.zv. *binomické rozdělení* a jejich hodnoty jsou předevší koncentrovány k hodnotě np, která má pravděpodobnost výskytu největší. Představu o jejich rozdělení získáme z *Bernoulliho nerovnosti,* která je *Čebyševovou nerovností* pro binomické rozdělení.

4.6. Věta: Bernoulliho nerovnost. Je-li $P(A) = p,\ 0 < p < 1$, pak pro počet k výskytů náhodného jevu A v sérii n nezávislých pokusů a číslo $\varepsilon > 0$ platí odhad:

$$P^* = P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = P\left(|k-np| < n\varepsilon\right) \geq 1 - \frac{p(1-p)}{n\varepsilon^2},$$

tedy

$$P\left(\left|\frac{k}{n}-p\right|\geq\varepsilon\right)=P\left(|k-np|\geq n\varepsilon\right)\leq\frac{p(1-p)}{n\varepsilon^2}.$$

Tudíž

$$\lim_{n\to\infty}P\left(\left|\frac{k}{n}-p\right|<\varepsilon\right)=1.$$

Důkaz: Znázorníme si na obrázku, kde leží hodnoty k, které odpovídají počtu výskytů náhodného jevu A splňující podmínku $0\leq k\leq n$ a

$\delta(\varepsilon):\ \left|\frac{k}{n}-p\right|<\varepsilon\Leftrightarrow|k-np|<n\varepsilon\Leftrightarrow np-n\varepsilon<k<np+n\varepsilon.$

Potom z vlastností uvedených v odstavci 4.5 úpravami dostaneme:

$$np(1-p)=\sum_{k=0}^{n}(k-np)^2.P_n(k)\geq\sum_{k\notin\delta(\varepsilon)}(k-np)^2.P_n(k)\geq$$

$$\geq n^2\varepsilon^2\sum_{k\notin\delta(\varepsilon)}P_n(k)=n^2\varepsilon^2\left(1-\sum_{k\in\delta(\varepsilon)}P_n(k)\right)=$$

$$=n^2\varepsilon^2P\left(|\frac{k}{n}-p|\geq\varepsilon\right)=n^2\varepsilon^2\left(1-P\left(|\frac{k}{n}-p|<\varepsilon\right)\right)$$

Uvážíme nerovnici mezi prvním a posledním členem ve vztahu a po vydělení výrazem $n^2\varepsilon^2$ dostaneme první z nerovnic ve větě. Druhá je jejím přepisem, neboť obsahuje pravděpodobnost opačného jevu, tedy doplněk do 1. Třetí vlastnost bezprostředně vyplývá z první, jestliže přejdeme ve vztahu k limitě a uvážíme, že se jedná o posloupnost s kladnými členy.

Poznámka: Nerovnost ve větě uvadí vztah mezi třemi veličinami. Jsou to odhad pravděpodobnosti, velikost intervalu a počet pokusů v sérii. Jestliže si dvě zadáme můžeme určit třetí. Poznamenejme, že vycházíme z odhadů a tudíž získáme i odhad hodnoty počítané veličiny. Ukažme si příklady takového využití odhadu z věty 4.6.

4.7. Příklad: Házíme 1000-krát mincí. Odhadněte pravděpodobnost P^*, s jakou se bude počet rubů vyskytovat v intervalu $(450, 550)$.

Řešení: V souladu se zněním věty 4.6 je : $n = 1000,\ p = 0,5$, $np = 1000.0,5 = 500$ a $n\varepsilon = 1000\varepsilon = 50 \Rightarrow \varepsilon = 0,05$. Podle odhadu z věty je

$$P^* = P\left(\left|\frac{k}{1000} - 0,5\right| < 0,05\right) \geq 1 - \frac{0,5.0,5}{1000.(0,05)^2} = 1 - \frac{1}{10} = 0,9.$$

4.8. Příklad: V náhodném pokusu má jev A pravděpodobnost výskytu $P(A) = 0,3$. V jakém intervalu se bude počet jevů A vyskytovat s pravděpodobností $P^* = 0,98$, pokud budeme pokus opakovat 500-krát.

Řešení: V souladu s větou 4.6 je: $p = 0,3,\ n = 500,\ np = 150$, $P^* \geq 0,98,\ \varepsilon =?$ Aby byla pravděpodobnost popsaného jevů větší nebo rovna 0,98, musí být větší její odhad. Odtud dostaneme podmínku pro hledaný interval. Je

$$P^* = P\left(\left|\frac{k}{500} - 0,3\right| < \varepsilon\right) \geq 1 - \frac{0,3.0,7}{500.\varepsilon^2} \geq 0,98 \Rightarrow$$

$$0,02 \geq \frac{0,21}{500\varepsilon^2} \Rightarrow \varepsilon \geq \sqrt{\frac{0,21}{10}} = 0,145.$$

Potom pro toleranční interval dostaneme: $n\varepsilon = 500.0,145 = 72,5$ tedy $np - n\varepsilon \leq k \leq np + n\varepsilon \Rightarrow 77,5 \leq k \leq 222,5$.

4.9. Příklad: Házíme hrací kostkou a počítáme počet šestek. Kolik musíme provést hodů, aby se s pravděpodobností $P^* = 0,99$ lišila relativní četnost od hodnoty $p = \frac{1}{6}$ nejvýše o $\varepsilon = 0,1$.

Řešení: V souladu se značením ve větě 4.6 je : $p = \frac{1}{6} = 0,1666\overline{6}$, $\varepsilon = 0,1,\ P^* \geq 0,9$ a $n =?$ Pravděpodobnost P^* bude větší než 0,99, jestliže bude větší než 0,99 její odhad. Je tedy

$$P^* = P\left(\left|\frac{k}{n} - \frac{1}{6}\right| < 0,1\right) \geq 1 - \frac{\frac{1}{6}\cdot\frac{5}{6}}{n.(0,1)^2} \geq 0,99 \Rightarrow n \geq \frac{5}{36}.10^4 = 1388,89.$$

Musíme provést alespoň 1389 hodů. Určeme si interval pro počet šestek. Je $np = 1389.0,16667 = 231,5$ a $n\varepsilon = 1389.0,1 = 138,9$, tedy $np - n\varepsilon \leq k \leq np + n\varepsilon \Rightarrow 92,6 \leq k \leq 370,4$.

Poznámka: Komentujme ještě výsledek příkladu 4.9. Ten říká, že budeme-li opakovaně házet kostkou série o 1389 hodech a počítat šestky, tak v přibližně jedné ze sta sérií bude počet šestek menší než 92 či větší než 370.

Poznámka: Bernoulliho nerovnost nám dovoluje odhadnout pravděpodobnost P^*, což je souhrn pravděpodobností výskytů sledovaného náhodného jevu A kolem hodnoty np, která má největší pravděpodobnost. Směrem k oběma krajním hodnotám 0 a n tato pravděpodobnost klesá. Pro velké hodnoty n je výpočet těchto pravděpodobností velmi pracný. Jedná se o velké množství malých čísel. Budeme si uvádět formule, které umožní jejich jednoduchý výpočet. Na ukázku uvedeme jednu z nich, která vychází z t.zv. *Poissonova rozdělení.*

4.10. Věta: Poissonova věta. Nechť v sérii n nezávislých pokusů nastává náhodný jev A s pravděpodobností $P(A) = p_n$ a nechť $\lim_{n\to\infty} np_n = \lambda > 0$. Potom je

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!}\, \mathrm{e}^{-\lambda}, \quad k = 0, 1, 2, \dots .$$

Důkaz: nebudeme provádět. Pro využití tvrzení ve statistice není zajímavý. Opírá se o známou identitu $e^z = \lim_{n\to\infty} \left(1 + \frac{z}{n}\right)^n$.

4.11. Poznámka: Tvrzení věty říká, že pro velké hodnoty n, $n \geq 30$ a malé hodnoty p, $p \leq 0,1$, platí:

$$P_n(k) = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{(np)^k}{k!}\, \mathrm{e}^{-np}, \quad 0 \leq k \leq n.$$

Volíme $\lambda = np$ a uvedená aproximace je dobrá jak pro jednotlivé hodnoty, tak i pro jejich souhrn.

Příklad: Sdělovacím kanálem přenášíme symboly s chybovostí $0,1\%$. Jaká je pravděpodobnost, že ve zprávě, která má 2000 symbolů budou nejvýše 3 chyby.

Řešení: Přenos symbolů je opakovaný pokus, $n = 2000$, ve které sledovaný náhodný jev A, výskyt chyby, se objevuje s pravděpodobností $p = 0,1.0,01 = 0,001$. Pravděpodobnost výskytu daného počtu chyb se řídí Bernoulliho schematem z věty 4.4. Hledaná pravděpodobnost je rovna

$$P = P_{2000}(0) + P_{2000}(1) + P_{2000}(2) + P_{2000}(3),$$

kde $P_{2000}(k) = \binom{2000}{k}(0,001)^k(0,999)^{2000-k}, \quad 0 \le k \le 2000$. Snadno tyto pravděpodobnosti vyčíslíme pomocí aproximace z poznámky 4.11. Je $p = 0,001 < 0,1,\ n = 2000 > 30$ a $\lambda = np = 2000.0,001 = 2$. Je pak

$$P_{2000}(k) \approx \frac{2^k}{k!}\,\mathrm{e}^{-2}, \quad 0 \le k.$$

Odtud pro hledanou pravděpodobnost dostaneme

$$P = \left(1 + 2 + \frac{4}{2} + \frac{8}{6}\right)\mathrm{e}^{-2} = \frac{19}{3}\,\mathrm{e}^{-2} = 6,3333.0,13533 = 0,85712.$$

■