Příklady rozdělení náhodných veličin

Definice: Posloupnost nezávislých pokusů, z nichž v každém může nastat úspěch s pravděpodobností p a neúspěch s pravděpodobností $q = 1 - p$, se nazývá posloupnost *bernoulliovských pokusů*. Pojmenování odkazuje na Jacoba Bernoulliho (1655–1705).

Definice: *Binomické rozdělení* má náhodná veličina X, která udává celkový počet úspěchů v n bernoulliovských pokusech. Lehce se zjistí, že platí $P(X = k) = \binom{n}{k} p^k q^{n-k}$ pro $k = 0, \dots, n$. Budeme zkracovat $X \sim \mathrm{Bi}(n, p)$.

Zadání: (Newtonův-Pepysův problém) Na binomické rozdělení jsme už narazili dříve. Chceme, aby při $6m$ hodech kostkou padlo aspoň m šestek, $m = 1, 2, 3$. Pro které m je to nejpravděpodobnější?

Řešení: Označíme-li X_{6m} počet šestek v $6m$ hodech, pak $X_{6m} \sim \mathrm{Bi}(6m, 1/6)$. Platí $P(X_6 \geq 1) = \sum_{k=1}^{6} \binom{6}{k} \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{6-k}$, nebo lépe přes doplňkovou pravděpodobnost: $P(X_6 \geq 1) = 1 - P(X_6 = 0) = 1 - \left(\frac{5}{6}\right)^6$. Dohromady zjistíme, že

$$P(X_6 \geq 1) \doteq 0{,}665 > P(X_{12} \geq 2) \doteq 0{,}619 > P(X_{18} \geq 3) \doteq 0{,}597.$$

Zadání: Určitý let má kapacitu 150 pasažérů. Letecká společnost se chrání proti tomu, že někteří pasažéři se nedostaví k letu, tím, že letadlo přeobsadí. V tomto případě prodala 160 letenek. Pravděpodobnost, že se pasažér nedostaví, je 0,1. Předpokládejme, že pasažéři se chovají nezávisle na sobě. Jaká je pravděpodobnost, že některý pasažér bude muset být přemístěn na jiný let z důvodu nedostatku místa v letadle?

Řešení: Na problém se můžeme dívat jako na 160 bernoulliovských pokusů, které s pravděpodobností 0,9 skončí úspěchem (pasažér se dostaví) a s pravděpodobností 0,1 neúspěchem. Pak počet lidí (označme X), kteří skutečně dorazí na letiště, má binomické rozdělení s parametry $n = 160$ a $p = 0{,}9$. Výpočtem zjistíme, že $P(X > 150) = \sum_{k=151}^{160} \binom{160}{k} 0{,}9^k 0{,}1^{160-k} \doteq 0{,}036$.

Spočtěme střední hodnotu náhodné veličiny s binomickým rozdělením:

$$\begin{aligned}
\mathbb{E}X &= \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} \\
&= \sum_{k=1}^{n} n \binom{n-1}{k-1} p^k q^{n-k} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^j q^{n-1-j} \\
&= np(p+q)^{n-1} = np.
\end{aligned}$$

Předpokládejme, že $np = \lambda > 0$ a my zvětšujeme n (tím zároveň zmenšujeme p), pak

$$\begin{aligned}
\lim_{n\to\infty} \binom{n}{k} p^k q^{n-k} &= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \\
&= \frac{\lambda^k}{k!} \cdot 1 \cdot \mathrm{e}^{-\lambda} \cdot 1 = \mathrm{e}^{-\lambda} \frac{\lambda^k}{k!}.
\end{aligned}$$

Definice: Náhodná veličina X má *Poissonovo rozdělení* s parametrem $\lambda > 0$, když $P(X = k) = \mathrm{e}^{-\lambda}\frac{\lambda^k}{k!}$ pro $k = 0, 1, 2, \dots$ Je pojmenováno po Siméonu Denisi Poissonovi (1781–1840).

Střední hodnota Poissonova rozdělení je

$$\mathbb{E}X = \sum_{k=0}^{\infty} k\mathrm{e}^{-\lambda}\frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!} = \lambda\mathrm{e}^{-\lambda}\sum_{j=0}^{\infty}\frac{\lambda^j}{j!} = \lambda\mathrm{e}^{-\lambda}\mathrm{e}^{\lambda} = \lambda.$$

Příklad: V New Jersey se stalo, že jistá žena (Evelyn Adamsová) vyhrála hlavní cenu v loterii dvakrát během 4 měsíců. Nejprve 23. října 1985 uhádla všech 6 vylosovaných čísel z 39, pak 14. února 1986 trefila 6 čísel ze 42. Vyhrála celkem 5,4 milionů dolarů. Podle zástupců loterie je pravděpodobnost, že se něco takového stane, rovna přibližně 1 ku 17 biliónům. Jejich úvaha byla založena na tom, že $\frac{1}{\binom{39}{6}}\cdot\frac{1}{\binom{42}{6}} \doteq \frac{1}{1{,}71\cdot 10^{13}}$. To je pravděpodobnost, že vyhraje jackpot ve dvou konkrétních hrách. Když si ale uvědomíme obrovský počet lidí, kteří sází po dlouhou dobu, tak pravděpodobnost, že se někdy stane, že někdo vyhraje dvakrát, není až tak vysoce zanedbatelná. Příběh ženy z New Jersey má smutný konec. O výhry ji připravili výherní automaty a příbuzní, dnes žije na mizině a bydlí s rodinou v přívěsu.

Pro názornost se přenesme do poměrů v ČR. Uvažujme, že Sportku si vsadí 500 000 lidí, každý přitom vyplní 6 sloupečků. Pravděpodobnost, že určitý člověk vyhraje v jednom losování jackpot, je $p = 6 \cdot 2/\binom{49}{6}$, protože ve Sportce probíhají 2 tahy (zanedbáváme situaci, že oba tahy dopadnou stejně). Za období 5 let proběhne zhruba 500 losování. Pravděpodobnost, že během té doby daný člověk vyhraje jackpot aspoň dvakrát, je $p_2 = \sum_{k=2}^{500}\binom{500}{k}p^k(1-p)^{500-k} = 1 - (1-p)^{500} - 500p(1-p)^{499} \doteq 9{,}184 \cdot 10^{-8}$. Použijeme-li aproximaci Poissonovým rozdělením, tak počet vyhraných jackpotů má přibližně Poissonovo rozdělení s parametrem $\lambda = 500p$, proto pravděpodobnost aspoň dvou je přibližně $p_2 \doteq 1 - \mathrm{e}^{-\lambda} - \lambda\mathrm{e}^{-\lambda} \doteq 9{,}202 \cdot 10^{-8}$. Počet lidí, kteří vyhrají během dané doby dvakrát jackpot, má binomické rozdělení s parametry 500 000 a p_2, což se dá velmi dobře aproximovat Poissonovým rozdělením s parametrem $\lambda_2 = 500\,000 p_2$. Pravděpodobnost, že bude aspoň jeden takový člověk, je $1 - (1-p_2)^{500\,000} \doteq 1 - \mathrm{e}^{-\lambda_2} \doteq 0{,}045$.

Zadání: Jistá loterijní společnost vydala 10 000 stíracích losů v Andoře a 10 000 000 ve Španělsku. Na každém losu je jiná viditelná kombinace čísel (čtyřciferných v Andoře a sedmiciferných ve Španělsku). Dále každý los obsahuje ještě jednu kombinaci čísel, která je překryta stříbrným proužkem. Přiřazení bylo provedeno náhodně a to tak, že žádné dva losy neobsahují stejnou skrytou kombinaci. Majitel losu setře stříbrný proužek a vyhrává, pokud se odkrytá kombinace shoduje s tou vytištěnou. Myslíte si, že se pravděpodobnost, že bude alespoň jeden výherce v Andoře, výrazně liší od pravděpodobnosti, že bude alespoň jeden výherce ve Španělsku?

Řešení: Označme X náhodnou veličinu, která označuje počet výherců, pokud bylo vydáno n losů. Tento počet je stejný jako počet lidí, kteří dostanou svůj kabát v problému šatnářky. Její rozdělení se dá přesně určit. Pro tuto chvíli si vystačíme s tím, že se dá dobře aproximovat pomocí Poissonova rozdělení s parametrem 1. Náhodná veličina X představuje počet úspěchu v n pokusech, v každém pokuse je pravděpodobnost úspěchu $1/n$. Pokusy ovšem nejsou nezávislé. Pro velká n jsou ale závislosti mezi pokusy nepatrné a aproximace je dostatečná. V obou případech je tedy pravděpodobnost přibližně $1 - \mathrm{e}^{-1} \doteq 0{,}632$.

Definice: Říkáme, že náhodná veličina X, která označuje počet pokusů před prvním úspěchem v bernoulliovských pokusech, má *geometrické rozdělení.* Je dáno pravděpodobnostmi $P(X = k) = q^k p$ pro $k = 0, 1, 2, \ldots$

Střední hodnota geometrického rozdělení je

$$\mathbb{E}X = \sum_{k=0}^{\infty} kq^k p = pq \sum_{k=0}^{\infty} kq^{k-1} = pq \frac{1}{(1-q)^2} = \frac{q}{p}.$$

Využili jsme toho, že $\sum_{k=0}^{\infty} kq^{k-1} = 1/(1-q)^2$, což lze dostat derivováním známého vztahu pro součet geometrické řady $\sum_{k=0}^{\infty} q^k = 1/(1-q)$.

Příklad: Při ruské ruletě se počet výstřelů naprázdno řídí geometrickým rozdělením s parametrem $p = 1/6$. Střední hodnota je $q/p = 5$, tedy ostrý náboj vyjde v průměru při šestém výstřelu.

Příklad: Při martingalové strategii v ruletě sázíme na barvu tak dlouho, dokud se poprvé netrefíme. Počet kol, které proběhnou, je náhodná veličina X. Platí, že $X - 1$ má geometrické rozdělení s parametrem $p = 18/37$. Střední hodnota počtu kol je

$$\mathbb{E}X = 1 + \frac{q}{p} = \frac{1}{p} = \frac{37}{18} \doteq 2{,}06.$$