PRAVDĚPODOBNOST A STATISTIKA

Modely diskrétní náhodné veličiny

Katedra kvantitativních metod FVL UO Brno

6. PŘEDNÁŠKA

Modely diskrétní náhodné veličiny

- Poissonovo rozdělení
- Alternativní rozdělení
- Binomické rozdělení
- Hypergeometrické rozdělení

Poissonovo rozdělení $Po(\lambda)$ je možné použít jako model náhodné veličiny, která nabývá hodnot $0, 1, 2, \ldots$ a udává buď počet událostí, k nimž dojde v časovém intervalu délky t nebo počet výskytů daných prvků v geometrické oblasti o pevné velikosti, jestliže k událostem či výskytům dochází jednotlivě a nezávisle na sobě. Parametr rozdělení $\lambda > 0$ udává střední počet událostí resp. výskytů.

Definice

Náhodná veličina X má *Poissonovo rozdělení* $Po(\lambda)$, právě když má pravděpodobnostní funkce tvar

$$p(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & x = 0, 1, 2, \ldots, \\ 0 & \text{jinak.} \end{cases}$$

Následující tabulka uvádí hodnoty některých číselných charakteristik Poisssonova rozdělení.

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$	$Mo(X)$
λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$	$\lambda - 1 \leq Mo(X) \leq \lambda$

Příklady náhodných veličin s Poissonovým rozdělením:
počet poruch stroje za směnu, počet nehod na jistém místě za rok, počet zákazníků v obchodě během 1 hodiny, počet vad na povrchu výrobku, počet vad v balíku látky, počet bublin na tabuli skla apod. Hodnoty pravděpodobnostní funkce Poissonova rozdělení jsou pro některé hodnoty λ tabelovány.

Následující tabulka uvádí hodnoty některých číselných charakteristik Poisssonova rozdělení.

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$	$Mo(X)$
λ	λ	$\frac{1}{\sqrt{\lambda}}$	$\frac{1}{\lambda}$	$\lambda - 1 \leq Mo(X) \leq \lambda$

Příklady náhodných veličin s Poissonovým rozdělením:
počet poruch stroje za směnu, počet nehod na jistém místě za rok, počet zákazníků v obchodě během 1 hodiny, počet vad na povrchu výrobku, počet vad v balíku látky, počet bublin na tabuli skla apod. Hodnoty pravděpodobnostní funkce Poissonova rozdělení jsou pro některé hodnoty λ tabelovány.

Příklad

Během 1 hodiny spojí sekretářka řediteli v průměru 6 hovorů. Potřebujeme sledovat zatížení sekretářky ve 20-ti minutových intervalech. Popište náhodnou veličinu udávající počet spojených telefonních hovorů během 20 minut pomocí pravděpodobností a distribuční funkce. Dále určete pravděpodobnost, že během 20 minut sekretářka spojí
a) alespoň 1 hovor, b) nejvýše 2 hovory, c) jeden nebo 2 hovory.

Určete střední hodnotu, rozptyl, směrodatnou odchylku, modus, koeficient šikmosti a špičatosti sledované náhodné veličiny.

Příklad

Náhodná veličina X udává počet spojených telefonních hovorů za 20 minut. Může nabývat hodnot $0, 1, 2, \ldots$. Předpokládejme, že je možné ji modelovat pomocí Poissonova rozdělení. Parametr λ udává střední hodnotu náhodné veličiny s Poissonovým rozdělením, tedy střední počet telefonátů během 20 minut, což je 2 (za 1 hodinu je jich průměrně 6). Náhodná veličina X má Poissonovo rozdělení $X \sim Po(2)$.

Pravděpodobnostní funkce má tvar

$$p(x) = \begin{cases} \frac{2^x}{x!} e^{-2} & x = 0, 1, 2, \ldots, \\ 0 & \text{jinak.} \end{cases}$$

Příklad

Náhodná veličina X udává počet spojených telefonních hovorů za 20 minut. Může nabývat hodnot $0, 1, 2, \ldots$. Předpokládejme, že je možné ji modelovat pomocí Poissonova rozdělení. Parametr λ udává střední hodnotu náhodné veličiny s Poissonovým rozdělením, tedy střední počet telefonátů během 20 minut, což je 2 (za 1 hodinu je jich průměrně 6). Náhodná veličina X má Poissonovo rozdělení $X \sim Po(2)$.

Pravděpodobnostní funkce má tvar

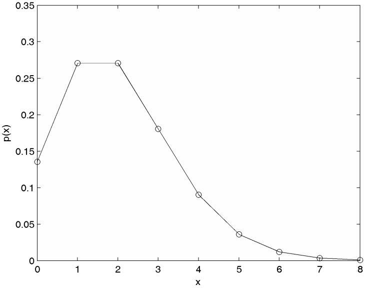
$$p(x) = \begin{cases} \frac{2^x}{x!} e^{-2} & x = 0, 1, 2, \ldots, \\ 0 & \text{jinak.} \end{cases}$$

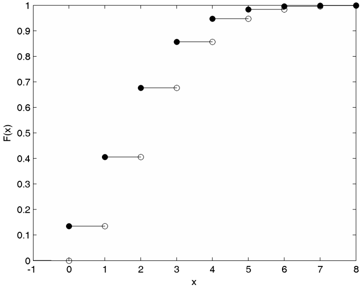
Příklad

x	0	1	2	3	4	5	6
$p(x)$	0,1353	0,2707	0,2707	0,1804	0,0902	0,0361	0,0120
$F(x)$	0,1353	0,4060	0,6767	0,8571	0,9473	0,9834	0,9955

Tabulka: Vybrané hodnoty pravděpodobnostní a distribuční funkce $Po(2)$.

Příklad





Obrázek: Pravděpodobnostní a distribuční funkce rozdělení $Po(2)$

Příklad

Nyní spočítáme že během 20 minut sekretářka spojí

a) alespoň 1 hovor
$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) = \\ = 1 - 0{,}1353 \doteq 0{,}865,$$

b) nejvýše 2 hovory
$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \\ = p(0) + p(1) + p(2) = 0{,}1353 + 0{,}2707 + 0{,}2707 = \\ = F(2) \doteq 0{,}677,$$

c) jeden nebo 2 hovory
$$P(X = 1 \vee X = 2) = P(X = 1) + P(X = 2) = p(1) + p(2) = \\ = 0{,}2707 + 0{,}2707 = 0{,}541.$$

Příklad

Nyní spočítáme že během 20 minut sekretářka spojí

a) alespoň 1 hovor

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) = \\ = 1 - 0{,}1353 \doteq 0{,}865,$$

b) nejvýše 2 hovory

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \\ = p(0) + p(1) + p(2) = 0{,}1353 + 0{,}2707 + 0{,}2707 = \\ = F(2) \doteq 0{,}677,$$

c) jeden nebo 2 hovory

$$P(X = 1 \vee X = 2) = P(X = 1) + P(X = 2) = p(1) + p(2) = \\ = 0{,}2707 + 0{,}2707 = 0{,}541.$$

Příklad

Nyní spočítáme že během 20 minut sekretářka spojí

a) alespoň 1 hovor
$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) = \\ = 1 - 0{,}1353 \doteq 0{,}865,$$

b) nejvýše 2 hovory
$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2) = \\ = p(0) + p(1) + p(2) = 0{,}1353 + 0{,}2707 + 0{,}2707 = \\ = F(2) \doteq 0{,}677,$$

c) jeden nebo 2 hovory
$$P(X = 1 \vee X = 2) = P(X = 1) + P(X = 2) = p(1) + p(2) = \\ = 0{,}2707 + 0{,}2707 = 0{,}541.$$

Příklad

Dále určíme některé číselné charakteristiky:

- střední hodnota Poissonova rozdělení je rovna $E(X) = \lambda = 2$,
- rozptyl má hodnotu $D(X) = \lambda = 2$,
- směrodatná odchylka $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1{,}414$.
- pro modus platí $\lambda - 1 \leq Mo(X) \leq \lambda$, tedy $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ a 2 (viz tabulka pravděpodobnostní funkce),
- koeficient šikmosti $\alpha_3 = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{2}} \doteq 0{,}707$,
- koeficient špičatosti $\alpha_4 = \frac{1}{\lambda} = \frac{1}{2} = 0{,}5$.

Příklad

Dále určíme některé číselné charakteristiky:

- střední hodnota Poissonova rozdělení je rovna $E(X) = \lambda = 2$,
- rozptyl má hodnotu $D(X) = \lambda = 2$,
- směrodatná odchylka $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1{,}414$.
- pro modus platí $\lambda - 1 \leq Mo(X) \leq \lambda$, tedy $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ a 2 (viz tabulka pravděpodobnostní funkce),
- koeficient šikmosti $\alpha_3 = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{2}} \doteq 0{,}707$,
- koeficient špičatosti $\alpha_4 = \frac{1}{\lambda} = \frac{1}{2} = 0{,}5$.

Příklad

Dále určíme některé číselné charakteristiky:

- střední hodnota Poissonova rozdělení je rovna $E(X) = \lambda = 2$,
- rozptyl má hodnotu $D(X) = \lambda = 2$,
- směrodatná odchylka $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1{,}414$.
- pro modus platí $\lambda - 1 \leq Mo(X) \leq \lambda$, tedy $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ a 2 (viz tabulka pravděpodobnostní funkce),
- koeficient šikmosti $\alpha_3 = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{2}} \doteq 0{,}707$,
- koeficient špičatosti $\alpha_4 = \frac{1}{\lambda} = \frac{1}{2} = 0{,}5$.

Příklad

Dále určíme některé číselné charakteristiky:

- střední hodnota Poissonova rozdělení je rovna $E(X) = \lambda = 2$,
- rozptyl má hodnotu $D(X) = \lambda = 2$,
- směrodatná odchylka $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1{,}414$.
- pro modus platí $\lambda - 1 \leq Mo(X) \leq \lambda$, tedy $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ a 2 (viz tabulka pravděpodobnostní funkce),
- koeficient šikmosti $\alpha_3 = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{2}} \doteq 0{,}707$,
- koeficient špičatosti $\alpha_4 = \frac{1}{\lambda} = \frac{1}{2} = 0{,}5$.

Příklad

Dále určíme některé číselné charakteristiky:

- střední hodnota Poissonova rozdělení je rovna $E(X) = \lambda = 2$,
- rozptyl má hodnotu $D(X) = \lambda = 2$,
- směrodatná odchylka $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1{,}414$.
- pro modus platí $\lambda - 1 \leq Mo(X) \leq \lambda$, tedy $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ a 2 (viz tabulka pravděpodobnostní funkce),
- koeficient šikmosti $\alpha_3 = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{2}} \doteq 0{,}707$,
- koeficient špičatosti $\alpha_4 = \frac{1}{\lambda} = \frac{1}{2} = 0{,}5$.

Dále určíme některé číselné charakteristiky:

- střední hodnota Poissonova rozdělení je rovna $E(X) = \lambda = 2$,
- rozptyl má hodnotu $D(X) = \lambda = 2$,
- směrodatná odchylka $\sigma = \sqrt{D(X)} = \sqrt{\lambda} = \sqrt{2} \doteq 1{,}414$.
- pro modus platí $\lambda - 1 \leq Mo(X) \leq \lambda$, tedy $2 - 1 \leq Mo(X) \leq 2$, $Mo(X) = 1$ a 2 (viz tabulka pravděpodobnostní funkce),
- koeficient šikmosti $\alpha_3 = \frac{1}{\sqrt{\lambda}} = \frac{1}{\sqrt{2}} \doteq 0{,}707$,
- koeficient špičatosti $\alpha_4 = \frac{1}{\lambda} = \frac{1}{2} = 0{,}5$.

Alternativní rozdělení

Některé náhodné pokusy mohou mít je 2 různé výsledky: pokus je úspěšný a pokus je neúspěšný. Náhodná veličina udávající počet úspěchů v jednom pokusu se nazývá alternativní. Tato veličina nabývá hodnot 0 a 1. Pravděpodobnost úspěchu je dána parametrem π $(0 < \pi < 1)$.

Definice

Náhodná veličina X má *alternativní rozdělení* $A(\pi)$, právě když má pravděpodobnostní funkce tvar

$$p(x) = \begin{cases} \pi^x (1-\pi)^{1-x} & x = 0, 1, \\ 0 & \text{jinak.} \end{cases}$$

Některé náhodné pokusy mohou mít je 2 různé výsledky: pokus je úspěšný a pokus je neúspěšný. Náhodná veličina udávající počet úspěchů v jednom pokusu se nazývá alternativní. Tato veličina nabývá hodnot 0 a 1. Pravděpodobnost úspěchu je dána parametrem π $(0 < \pi < 1)$.

Definice

Náhodná veličina X má *alternativní rozdělení* $A(\pi)$, právě když má pravděpodobnostní funkce tvar

$$p(x) = \begin{cases} \pi^x(1-\pi)^{1-x} & x = 0, 1, \\ 0 & \text{jinak.} \end{cases}$$

Následující tabulka uvádí hodnoty některých číselných charakteristik alternativního rozdělení.

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$
π	$\pi(1-\pi)$	$\frac{1-2\pi}{\sqrt{\pi(1-\pi)}}$	$\frac{1-6\pi(1-\pi)}{\pi(1-\pi)}$

Příklady: počet zmetků při náhodném výběru 1 výrobku, počet zásahů při jednom výstřelu, počet spojení při 1 telefoním volání, indikuje nastoupení či nenastoupení náhodného jevu.

Následující tabulka uvádí hodnoty některých číselných charakteristik alternativního rozdělení.

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$
π	$\pi(1-\pi)$	$\frac{1-2\pi}{\sqrt{\pi(1-\pi)}}$	$\frac{1-6\pi(1-\pi)}{\pi(1-\pi)}$

Příklady: počet zmetků při náhodném výběru 1 výrobku, počet zásahů při jednom výstřelu, počet spojení při 1 telefoním volání, indikuje nastoupení či nenastoupení náhodného jevu.

Příklad

Určete střední hodnotu a rozptyl náhodné veličiny s alternativním rozdělením $X \sim A(\pi)$.

Pro střední hodnotu diskrétní náhodné veličiny s alternativním rozdělením platí

$$E(X) = \sum_{x \in M} xp(x) = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi.$$

Rozptyl nespojité náhodné veličiny daného rozdělení získáme ze vztahu

$$\begin{aligned} D(X) &= \sum_{x \in M} [x - E(X)]^2 p(x) = (0 - \pi)^2 (1 - \pi) + (1 - \pi)^2 \pi = \\ &= \pi^2 (1 - \pi) + (1 - \pi)^2 \pi = \pi(1 - \pi)(\pi + 1 - \pi) = \pi(1 - \pi). \end{aligned}$$

Příklad

Určete střední hodnotu a rozptyl náhodné veličiny s alternativním rozdělením $X \sim A(\pi)$.

Pro střední hodnotu diskrétní náhodné veličiny s alternativním rozdělením platí

$$E(X) = \sum_{x \in M} xp(x) = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi.$$

Rozptyl nespojité náhodné veličiny daného rozdělení získáme ze vztahu

$$\begin{aligned} D(X) &= \sum_{x \in M} [x - E(X)]^2 p(x) = (0 - \pi)^2 (1 - \pi) + (1 - \pi)^2 \pi = \\ &= \pi^2 (1 - \pi) + (1 - \pi)^2 \pi = \pi(1 - \pi)(\pi + 1 - \pi) = \pi(1 - \pi). \end{aligned}$$

Příklad

Určete střední hodnotu a rozptyl náhodné veličiny s alternativním rozdělením $X \sim A(\pi)$.

Pro střední hodnotu diskrétní náhodné veličiny s alternativním rozdělením platí

$$E(X) = \sum_{x \in M} xp(x) = 0 \cdot (1 - \pi) + 1 \cdot \pi = \pi.$$

Rozptyl nespojité náhodné veličiny daného rozdělení získáme ze vztahu

$$D(X) = \sum_{x \in M} [x - E(X)]^2 p(x) = (0 - \pi)^2 (1 - \pi) + (1 - \pi)^2 \pi =$$
$$= \pi^2 (1 - \pi) + (1 - \pi)^2 \pi = \pi(1 - \pi)(\pi + 1 - \pi) = \pi(1 - \pi).$$

Náhodná veličina, kterou je možné modelovat pomocí binomického rozdělení, udává počet úspěchů v posloupnosti n nezávislých alternativních pokusů, přičemž úspěch v každém pokusu nastává s pravděpodobností π $(0 < \pi < 1)$.

Definice

Náhodná veličina X má *binomické rozdělení* $B(n, \pi)$, právě když má pravděpodobnostní funkce tvar

$$p(x) = \begin{cases} \binom{n}{x}\pi^x(1-\pi)^{n-x} & x = 0, 1, \ldots, n, \\ 0 & \text{jinak.} \end{cases}$$

Náhodná veličina, kterou je možné modelovat pomocí binomického rozdělení, udává počet úspěchů v posloupnosti n nezávislých alternativních pokusů, přičemž úspěch v každém pokusu nastává s pravděpodobností π $(0 < \pi < 1)$.

Definice

Náhodná veličina X má *binomické rozdělení* $B(n, \pi)$, právě když má pravděpodobnostní funkce tvar

$$p(x) = \begin{cases} \binom{n}{x}\pi^x(1-\pi)^{n-x} & x = 0, 1, \ldots, n, \\ 0 & \text{jinak.} \end{cases}$$

Následující tabulka uvádí hodnoty některých číselných charakteristik binomického rozdělení.

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$	$Mo(X)$
$n\pi$	$n\pi(1-\pi)$	$\frac{1-2\pi}{\sqrt{n\pi(1-\pi)}}$	$\frac{1-6\pi(1-\pi)}{n\pi(1-\pi)}$	$(n+1)\pi-1 \leq Mo(X) \leq (n+1)\pi$

Příklady náhodných veličin s binomickým rozdělením: počet padnutých šestek v pěti hodech hrací kostkou, počet vadných výrobků z celkového počtu 100 výrobků, je-li pravděpodobnost výskytu vadného výrobku 0,005, počet spojení při n telefonních voláních, počet zásahů při n výstřelech apod.

Následující tabulka uvádí hodnoty některých číselných charakteristik binomického rozdělení.

$E(X)$	$D(X)$	$\alpha_3(X)$	$\alpha_4(X)$	$Mo(X)$
$n\pi$	$n\pi(1-\pi)$	$\frac{1-2\pi}{\sqrt{n\pi(1-\pi)}}$	$\frac{1-6\pi(1-\pi)}{n\pi(1-\pi)}$	$(n+1)\pi-1 \leq Mo(X) \leq (n+1)\pi$

Příklady náhodných veličin s binomickým rozdělením: počet padnutých šestek v pěti hodech hrací kostkou, počet vadných výrobků z celkového počtu 100 výrobků, je-li pravděpodobnost výskytu vadného výrobku 0,005, počet spojení při n telefonních voláních, počet zásahů při n výstřelech apod.

Mají-li veličiny $X_1, \ldots, X_n$ stejné alternativní rozdělení s parametrem π a jsou nezávislé, potom veličina $M = X_1 + X_2 + \cdots + X_n$ má binomické rozdělení $B(n, \pi)$, s parametry n a π. Alternativní rozdělení je tedy speciálním případem binomického rozdělení pro $n = 1$.

Poissonovo rozdělení je limitním případem binomického rozdělení. Jestliže $n \to \infty$ a $\pi \to 0$, pak $n\pi \to \lambda$. Hodnoty pravděpodobnostní funkce binomického rozdělení je možné aproximovat pomocí hodnot pravděpodobnostní funkce Poissonova rozdělení. Při řešení úloh je pak dostačující, aby $n > 30$, $\pi < 0{,}1$, pak platí

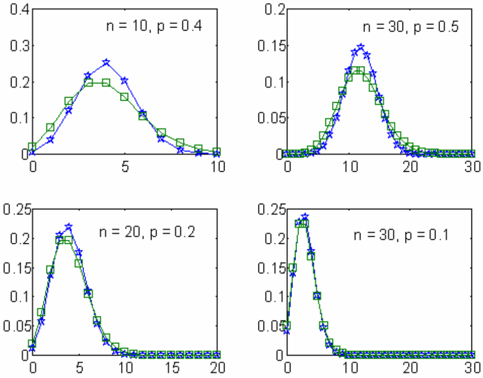
$$\binom{n}{x} \pi^x (1-\pi)^{n-x} \approx \frac{\lambda^x}{x!} e^{-\lambda}.$$

Mají-li veličiny $X_1, \ldots, X_n$ stejné alternativní rozdělení s parametrem π a jsou nezávislé, potom veličina $M = X_1 + X_2 + \cdots + X_n$ má binomické rozdělení $B(n, \pi)$, s parametry n a π. Alternativní rozdělení je tedy speciálním případem binomického rozdělení pro $n = 1$.

Poissonovo rozdělení je limitním případem binomického rozdělení. Jestliže $n \to \infty$ a $\pi \to 0$, pak $n\pi \to \lambda$. Hodnoty pravděpodobnostní funkce binomického rozdělení je možné aproximovat pomocí hodnot pravděpodobnostní funkce Poissonova rozdělení. Při řešení úloh je pak dostačující, aby $n > 30$, $\pi < 0{,}1$, pak platí

$$\binom{n}{x} \pi^x (1-\pi)^{n-x} \approx \frac{\lambda^x}{x!} e^{-\lambda}.$$

Binomické rozdělení



Obrázek: čtverec – Poissonovo rozdělení, hvězda – binomické rozdělení

Příklad

Pravděpodobnost, že narozené dítě je chlapec je 0,51. Jaká je pravděpodobnost,že mezi pěti po sobě narozenými dětmi budou
a) právě 3 děvčata, b) nejvýše 3 chlapci?
Určete pravděpodobnostní a distribuční funkci náhodné veličiny udávající počet chlapců mezi pěti po sobě narozenými dětmi. Jaký je nejpravděpodobnější počet narozených chlapců?

Určete střední hodnotu, rozptyl a směrodatnou odchylku dané náhodné veličiny.

Příklad

Náhodná veličina X udává počet chlapců mezi pěti po sobě narozenými dětmi. Tato náhodná veličina může nabývat hodnot $0, 1, 2, \ldots, 5$. Považujme narození dítěte za nezávislý náhodný pokus, ve kterém se narodí chlapec s pravděpodobností 0,51. Náhodnou veličinu může popsat pomocí binomického rozdělení $X \sim B(5; 0{,}51)$.

Pravděpodobnostní funkci můžeme zapsat ve tvaru

$$p(x) = \begin{cases} \binom{5}{x} 0{,}51^x 0{,}49^{5-x} & x = 0, 1, \ldots, 5, \\ 0 & \text{jinak.} \end{cases}$$

Příklad

Náhodná veličina X udává počet chlapců mezi pěti po sobě narozenými dětmi. Tato náhodná veličina může nabývat hodnot $0, 1, 2, \ldots, 5$. Považujme narození dítěte za nezávislý náhodný pokus, ve kterém se narodí chlapec s pravděpodobností 0,51. Náhodnou veličinu může popsat pomocí binomického rozdělení $X \sim B(5; 0{,}51)$.

Pravděpodobnostní funkci můžeme zapsat ve tvaru

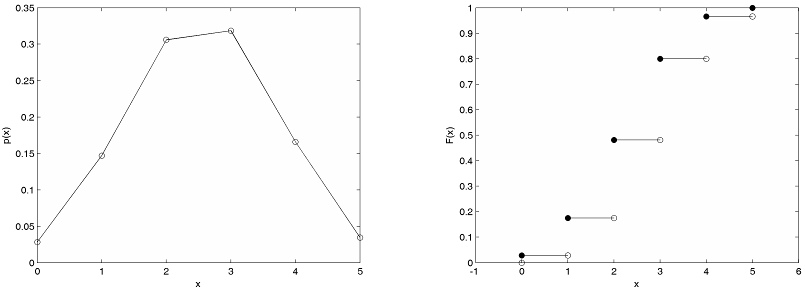
$$p(x) = \begin{cases} \binom{5}{x} 0{,}51^{x} 0{,}49^{5-x} & x = 0, 1, \ldots, 5, \\ 0 & \text{jinak.} \end{cases}$$

Příklad

x	0	1	2	3	4	5
$p(x)$	0,0282	0,1470	0,3060	0,3185	0,1657	0,0345
$F(x)$	0,0282	0,1752	0,4813	0,7998	0,9655	1,0000

Tabulka: Pravděpodobnostní funkce a vybrané hodnoty distribuční funkce $B(5; 0{,}51)$.

Příklad



Obrázek: Pravděpodobnostní a distribuční funkce rozdělení $B(5; 0{,}51)$

Nyní určíme pravděpodobnost, že mezi pěti po sobě narozenými dětmi budou

a) právě 3 děvčata, tzn. právě 2 chlapci
$P(X = 2) = p(2) \doteq 0{,}306,$

b) nejvýše 3 chlapci
$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + \\ &\quad + P(X = 3) = p(0) + p(1) + p(2) + p(3) = \\ &= 0{,}0282 + 0{,}147 + 0{,}3060 + 0{,}3185 = \\ &= F(3) \doteq 0{,}800. \end{aligned}$$

Příklad

Nyní určíme pravděpodobnost, že mezi pěti po sobě narozenými dětmi budou

a) právě 3 děvčata, tzn. právě 2 chlapci
$P(X = 2) = p(2) \doteq 0{,}306$,

b) nejvýše 3 chlapci

$$\begin{aligned} P(X \le 3) &= P(X = 0) + P(X = 1) + P(X = 2) + \\ &\quad + P(X = 3) = p(0) + p(1) + p(2) + p(3) = \\ &= 0{,}0282 + 0{,}147 + 0{,}3060 + 0{,}3185 = \\ &= F(3) \doteq 0{,}800. \end{aligned}$$

Nejpravděpodobnější počet narozených chlapců určuje modus a ten můžeme určit ze vztahu $(n+1)\pi - 1 \leq Mo(X) \leq (n+1)\pi$, tedy $(5+1)\cdot 0{,}51 - 1 \leq Mo(X) \leq (5+1)\cdot 0{,}51$, což je $2{,}06 \leq Mo(X) \leq 3{,}06$ odkud dostáváme $Mo(X) = 3$. Nejpravděpodobnější hodnotu můžeme samozřejmě najít přímo v tabulce pravděpodobnostní funkce.

Střední hodnota je pro binomické rozdělení rovna $E(X) = n\pi = 5\cdot 0{,}51 = 2{,}55$, rozptyl je roven $D(X) = n\pi(1-\pi) = 5\cdot 0{,}51\cdot(1-0{,}51) \doteq 1{,}250$ a směrodatná odchylka $\sigma = \sqrt{D(X)} \doteq 1{,}119$.

Příklad

Nejpravděpodobnější počet narozených chlapců určuje modus a ten můžeme určit ze vztahu $(n+1)\pi - 1 \leq Mo(X) \leq (n+1)\pi$, tedy $(5+1)\cdot 0{,}51 - 1 \leq Mo(X) \leq (5+1)\cdot 0{,}51$, což je $2{,}06 \leq Mo(X) \leq 3{,}06$ odkud dostáváme $Mo(X) = 3$. Nejpravděpodobnější hodnotu můžeme samozřejmě najít přímo v tabulce pravděpodobnostní funkce.

Střední hodnota je pro binomické rozdělení rovna $E(X) = n\pi = 5\cdot 0{,}51 = 2{,}55$, rozptyl je roven $D(X) = n\pi(1-\pi) = 5\cdot 0{,}51\cdot(1-0{,}51) \doteq 1{,}250$ a směrodatná odchylka $\sigma = \sqrt{D(X)} \doteq 1{,}119$.

Máme N objektů mezi nimiž je M se sledovanou vlastností (např. 4 vadné výrobky v sérii 200 kusů, 6 čísel ze 49, na která sázející Sportky vsadil, ...). Vybereme náhodně bez vracení n objektů. Náhodná veličina X, která udává počet vybraných objektů se sledovanou vlastností má hypergeometrické rozdělení.

Definice

Náhodná veličina X má *hypergeometrické rozdělení* $Hg(N, M, n)$, právě když má pravděpodobnostní funkce tvar

$$p(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \max\{0, n-N+M\} \le x \le \min\{n, M\}, \\ 0 & \text{jinak.} \end{cases}$$

Máme N objektů mezi nimiž je M se sledovanou vlastností (např. 4 vadné výrobky v sérii 200 kusů, 6 čísel ze 49, na která sázející Sportky vsadil, ...). Vybereme náhodně bez vracení n objektů. Náhodná veličina X, která udává počet vybraných objektů se sledovanou vlastností má hypergeometrické rozdělení.

Definice

Náhodná veličina X má *hypergeometrické rozdělení* $Hg(N, M, n)$, právě když má pravděpodobnostní funkce tvar

$$p(x) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} & \max\{0, n-N+M\} \leq x \leq \min\{n, M\}, \\ 0 & \text{jinak.} \end{cases}$$

Následující tabulka uvádí hodnoty některých číselných charakteristik binomického rozdělení.

$E(X)$	$D(X)$	$\alpha_3(X)$	$Mo(X)$	pozn.
$n\pi$	$n\pi(1-\pi)\frac{N-n}{N-1}$	$\frac{(1-2\pi)(N-2n)}{(N-2)\sigma}$	$a-1 \leq Mo(X) \leq a$	$\pi = \frac{M}{N}$, $a = \frac{(M+1)(n+1)}{N+2}$

Příklady: počet vadných výrobků mezi n náhodně vybranými výrobky z dodávky, Sportka, 5 ze 40, 10 šťastných čísel, apod.

Následující tabulka uvádí hodnoty některých číselných charakteristik binomického rozdělení.

$E(X)$	$D(X)$	$\alpha_3(X)$	$Mo(X)$	pozn.
$n\pi$	$n\pi(1-\pi)\frac{N-n}{N-1}$	$\frac{(1-2\pi)(N-2n)}{(N-2)\sigma}$	$a-1 \leq Mo(X) \leq a$	$\pi = \frac{M}{N}, \ a = \frac{(M+1)(n+1)}{N+2}$

Příklady: počet vadných výrobků mezi n náhodně vybranými výrobky z dodávky, Sportka, 5 ze 40, 10 šťastných čísel, apod.

Hypergeometrické rozdělení

Zlomek $\frac{n}{N}$ vyjadřuje tzv. *výběrový podíl*. Je-li tento podíl menší než 0,05, lze hypergeometrické rozdělení aproximovat binomickým rozdělením s parametry n a $\pi = \frac{M}{N}$, tedy

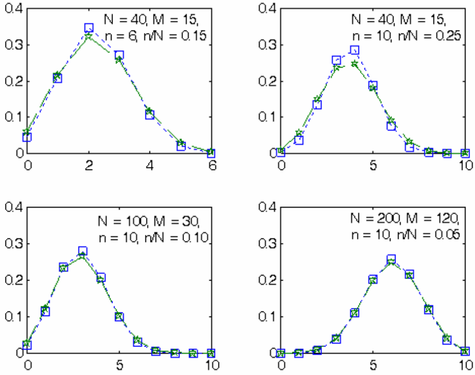
$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \approx \binom{n}{x} \pi^x (1-\pi)^{n-x}.$$

Je-li rozsah N velký a n relativně malé, potom rozdíl mezi výběrem bez vracení (rozdělení $Hg(N, M, n)$) a s vracením (rozdělení $B(n, \pi)$) je zanedbatelný.

Je-li navíc $\pi = \frac{M}{N} < 0{,}1$ a $n > 30$, je možné hypergeometrické rozdělení aproximovat Poissonovým rozdělením, kde $\lambda = n\frac{M}{N}$, tedy

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} \approx \frac{\lambda^x}{x!}e^{-\lambda}.$$

Hypergeometrické rozdělení



Obrázek: čtverec – hypergeometrické rozdělení, hvězda – binomické rozdělení

Příklad

Výrobky jsou dodávány v sériích po 100 kusech. Výstupní kontrola prohlíží z každé série 5 různých náhodně vybraných výrobků a přijímá ji, jestliže mezi nimi není žádný zmetek. Očekáváme, že série obsahuje 4 % zmetků. Určete pravděpodobnostní a distribuční funkci náhodné veličiny udávající počet zmetků ve výběru. S jakou pravděpodobnostní nebude série přijata? Spočítejte střední hodnotu a směrodatnou odchylku této náhodné veličiny. Zjistěte, zda jsou splněny podmínky aproximace binomickým rozdělením.

Příklad

V sériích po 100 kusech se očekává 4% zmetků, což je 4. Náhodnou veličinu udávající počet zmetků mezi 5 vybranými výrobky můžeme popsat pomocí hypergeometrického rozdělení $X \sim Hg(100, 4, 5)$. Tato náhodná veličina může nabývat hodnot $0, 1, 2, 3$ a 4.

Pravděpodobnostní funkce má tvar

$$p(x) = \begin{cases} \dfrac{\binom{4}{x}\binom{96}{5-x}}{\binom{100}{5}} & \text{pro } 0, 1, 2, 3, 4, \\ 0 & \text{jinak.} \end{cases}$$

Příklad

V sériích po 100 kusech se očekává 4% zmetků, což je 4. Náhodnou veličinu udávající počet zmetků mezi 5 vybranými výrobky můžeme popsat pomocí hypergeometrického rozdělení $X \sim Hg(100, 4, 5)$. Tato náhodná veličina může nabývat hodnot $0, 1, 2, 3$ a 4.

Pravděpodobnostní funkce má tvar

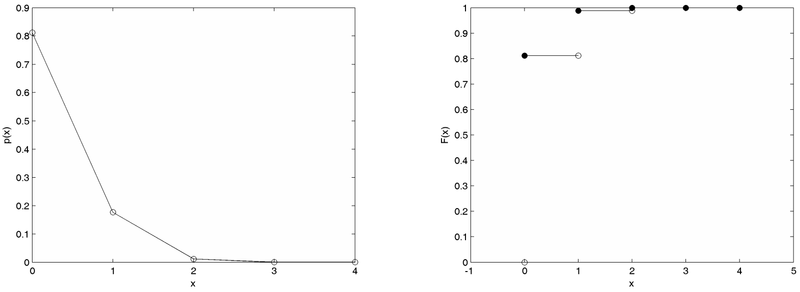
$$p(x) = \begin{cases} \frac{\binom{4}{x}\binom{96}{5-x}}{\binom{100}{5}} & \text{pro } 0, 1, 2, 3, 4, \\ 0 & \text{jinak.} \end{cases}$$

Příklad

x	0	1	2	3	4
$p(x)$	0,8119	0,1765	0,0114	0,0002	$1{,}3 \cdot 10^{-6}$
$F(x)$	0,8119	0,9884	0,9998	1	1

Tabulka: Hodnoty pravděpodobnostní funkce a distribuční funkce $Hg(100, 4, 5)$

Příklad



Obrázek: Pravděpodobnostní a distribuční funkce rozdělení $Hg(100, 4, 5)$

Příklad

Určíme pravděpodobnost, se kterou nebude série přijata, tzn. že bude obsahovat alespoň 1 zmetek, tedy
$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) \doteq 0{,}188.$

Střední hodnota hypergeometrického rozdělení je $E(X) = n\frac{M}{N} = 0{,}2$,
směrodatná odchylka $\sigma = \sqrt{D(X)} = \sqrt{n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}} \doteq 0{,}429.$

Příklad

Určíme pravděpodobnost, se kterou nebude série přijata, tzn. že bude obsahovat alespoň 1 zmetek, tedy
$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - p(0) \doteq 0{,}188.$

Střední hodnota hypergeometrického rozdělení je $E(X) = n\frac{M}{N} = 0{,}2$, směrodatná odchylka $\sigma = \sqrt{D(X)} = \sqrt{n\frac{M}{N}(1 - \frac{M}{N})\frac{N-n}{N-1}} \doteq 0{,}429.$

Příklad

Jelikož je výběrový podíl $\frac{n}{N} = 0{,}05$, můžeme hypergeometrické rozdělení aproximovat binomickým rozdělením $B(5; 0{,}04)$.

Pomocí aproximace tímto rozdělením by série nebyla přijata s pravděpodobností

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) =$$
$$= 1 - \binom{5}{0} 0{,}04^0 \cdot 0{,}96^5 \doteq 0{,}185.$$

Příklad

Jelikož je výběrový podíl $\frac{n}{N} = 0{,}05$, můžeme hypergeometrické rozdělení aproximovat binomickým rozdělením $B(5; 0{,}04)$.

Pomocí aproximace tímto rozdělením by série nebyla přijata s pravděpodobností

$$P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = \\ = 1 - \binom{5}{0} 0{,}04^0 \cdot 0{,}96^5 \doteq 0{,}185.$$

Základní literatura

- MANN, P.S. Introductory Statistics. 6th edition. Hoboken: Wiley, 2007. ISBN 978-0-471-75530-2.
- MOUČKA, J., RÁDL, P. Matematika pro studenty ekonomie. 1. vyd. Grada 2010. ISBN 978-80-247-3260-2.
- NEUBAUER, J., SEDLAČÍK, M., KŘÍŽ, O. Základy statistiky – Aplikace v technických a ekonomických oborech. Grada 2012.ISBN: 978-80-247-4273-1.
- ŘEZANKOVÁ, H. Analýza dat z dotazníkových šetření. 2. vydání, Professional Publishing, 2010. ISBN: 9788074310195.

Doporučená literatura

- AGRESTI, A. Categorical Data Analysis. Second Edition. Wiley 2002. ISBN: 0-471-36093-7.
- ANDĚL, J. Statistické metody. 3. vydání. Praha: Matfyzpress, 2003. ISBN 80-86732-08-8.
- ANDĚL, J. Základy matematické statistiky. 2. vyd. Praha: Matfyzpress, 2007, 358 s. ISBN 978-80-7378-001-2.
- VÁGNER, M. Integrální počet funkcí jedné proměnné. 1. vydání. Brno: UO, 2005,126 s. ISBN 80-7231-025-9.
- VÁGNER, M., KAŠTÁNKOVÁ, V. Posloupnosti a řady. 1. vydání. Brno: UO, 2006. ISBN 80-7231-131-X.