Vybraná rozdělení diskrétní náhodné veličiny

Martina Litschmannová

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY

Opakování: Náhodná veličina

- Co je to náhodná veličina?

  Funkce, která každému (dále nedělitelnému) výsledku náhodného pokusu (tj. elementárnímu jevu) přiřazuje reálné číslo.

  Zjednodušeně: číselný výsledek náhodného pokusu

- Co je to rozdělení náhodné veličiny X?

  Předpis, který umožňuje určit pravděpodobnost, že náhodná veličina X nabývá hodnoty z libovolné podmnožiny reálných čísel.

- Jaké základní typy náhodných veličin znáte?

  Diskrétní a spojité náhodné veličiny.

Opakování: Náhodná veličina

- Jak lze popsat rozdělení diskrétní náhodné veličiny?

  Distribuční funkcí $F(x) = P(X < x)$, resp. pravděpodobnostní funkcí $P(x_i)$.

- Jak lze popsat rozdělení spojité náhodné veličiny?

  Distribuční funkcí $F(x) = P(X < x)$, resp. hustotou pravděpodobností $f(x)$.

- Jaké základní číselné charakteristiky používáme pro popis náhodné veličiny?

  Střední hodnota $E(X)$, rozptyl $D(X)$, směrodatná odchylka $\sigma(X)$, p-kvantily x_p.

Opakování: Náhodná veličina

- Co nám říká Čebyševova nerovnost?

$$\forall k > 0: P(\mu - k\sigma \leq X \leq \mu + k\sigma) > 1 - \frac{1}{k^2}$$

Např.: $P(\mu - \sigma \leq X \leq \mu + \sigma) > 0$

$P(\mu - 2\sigma \leq X \leq \mu + 2\sigma) > 0{,}75$

$P(\mu - 3\sigma \leq X \leq \mu + 3\sigma) > 0{,}89$

Poznámka: $\mu = E(X)$, $\sigma = \sigma(X)$.

Diskrétní náhodná veličina (příklady)

Technika:

- počet šroubů typu M10 mezi 10 vybranými (víme-li, že do dodávky 100 šroubů bylo omylem zařazeno 20 šroubů typu M50)

Biomedicína:

- počet pacientů (z 10 očkovaných) u nichž byla použita prošlá očkovací látka (víme-li, že v balení bylo 20 dávek očkovací látky, přičemž 5 z nich bylo prošlých)

Obecně: počet úspěchů v n (závislých) pokusech

Diskrétní náhodná veličina (příklady)

Technika:

- počet správně přenesených bitů předtím než dojde ke 4. chybě (víme-li, že pravděpodobnost chybného přenosu bitu je 0,12)

Biomedicína:

- počet dobrovolníků, které budeme muset testovat dříve než najdeme 5 dárců s krevní skupinou AB (předpokládejme, že dobrovolníci neznají svou krevní skupinu, pravděpodobnost výskytu krevní skupiny (populační frekvence) krevní skupiny AB je 0,05)

Obecně: počet (nezávislých) pokusů do k. úspěchů

Diskrétní náhodná veličina (příklady)

Technika:

- počet škrábanců na 1 m^2 lakovaného povrchu (víme-li, že průměrně lze očekávat 3 škrábance na 10 m^2)

Biomedicína:

- počet červených krvinek v 10 ml krve ženy (víme-li, že průměrně lze pozorovat $4{,}8 \cdot 10^{12}$ červených krvinek v 1l krve (u žen))

Obecně: počet události v časovém intervalu, na ploše, v objemu

Vybraná rozdělení diskrétní náhodné veličiny

- Alternativní rozdělení
- Hypergeometrické rozdělení
- Binomické rozdělení
- Negativně binomické rozdělení
- Geometrické rozdělení
- Poissonovo rozdělení

Alternativní rozdělení

- nastane-li sledovaný náhodný jev A (budeme říkat, že došlo k výskytu události, resp. že došlo k úspěchu), bude mít náhodná veličina X hodnotu $x = 1$,
- nenastane-li jev A, bude mít náhodná veličina X hodnotu $x = 0$.

X … počet výskytů daného náhodného jevu (úspěchů) v jednom pokusu

$$X \sim A(\pi)$$

pravděpodobnost úspěchu

(parametr rozdělení)

Pravděpodobnostní funkce:
$P(X = 1) = \pi$
$P(X = 0) = 1 - \pi$

Střední hodnota: $E(X) = \pi$
Rozptyl: $D(X) = \pi(1 - \pi)$

(dokažte!)

Bernoulliho pokusy

Posloupnost nezávislých pokusů majících pouze dva možné výsledky

(tj. takových pokusů, kdy úspěch v libovolné skupině pokusů neovlivňuje pravděpodobnost úspěchu v pokusu, který do této skupiny nepatří),

v nichž jev A (úspěch) nastává s pravděpodobností π
a neúspěch s pravděpodobností $1 - \pi$.

Příklad 1

Předpokládejme, že pravděpodobnost narození dívky je 0,49. Jaká je pravděpodobnost toho, že mezi čtyřmi dětmi v rodině je

a) je právě jedna dívka?

X… počet dívek mezi 4 dětmi

D… narodí se dívka, $P(D) = \pi$

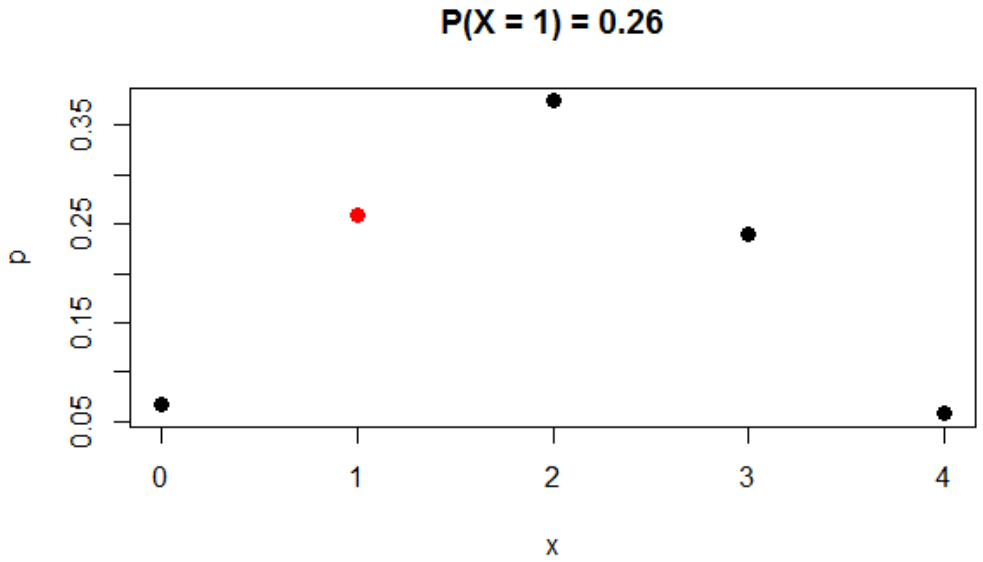
$\bar{D}$… narodí se kluk, $P(\bar{D}) = 1 - \pi$

$(X = 1)$ … $\{D\bar{D}\bar{D}\bar{D}, \bar{D}D\bar{D}\bar{D}, \bar{D}\bar{D}D\bar{D}, \bar{D}\bar{D}\bar{D}D\}$

$\mathrm{P}(\bar{D}D\bar{D}\bar{D}) = \mathrm{P}(\bar{D}\bar{D}D\bar{D}) = P(\bar{D}\bar{D}\bar{D}D) = \pi \cdot (1 - \pi)^3$

$P(X = 1) = P(D\bar{D}\bar{D}\bar{D} \cup \bar{D}D\bar{D}\bar{D} \cup \bar{D}\bar{D}D\bar{D} \cup \bar{D}\bar{D}\bar{D}D) = \binom{4}{1} \cdot \pi \cdot (1 - \pi)^3$

$P(X = 1) = \mathbf{0,260}$



Binomické rozdělení

X … počet úspěchů v n Bernoulliho pokusech

$$X \sim Bi(n; \pi)$$

počet pokusů pravděpodobnost úspěchu

Počet příznivých realizací v posloupnosti Bernoulliho pokusů.

Pravděpodobnostní funkce: $P(X = k) = \binom{n}{k} \pi^k (1 - \pi)^{n-k}$

Pravděpodobnost výskytu jedné příznivé realizace v posloupnosti Bernoulliho pokusů.

Příklad 1

Předpokládejme, že pravděpodobnost narození dívky je 0,49. Jaká je pravděpodobnost toho, že mezi čtyřmi dětmi v rodině je

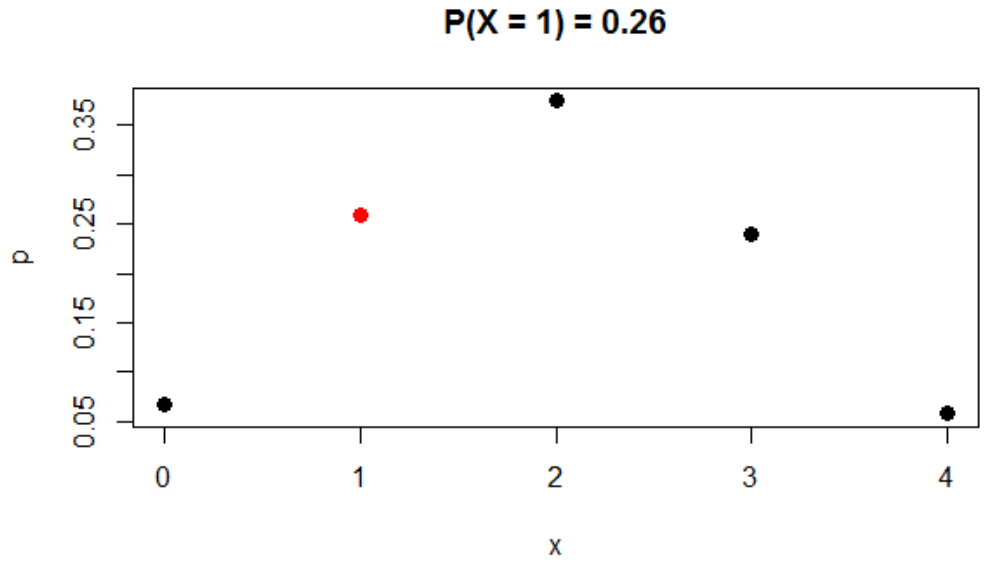
a) je právě jedna dívka?

X… počet dívek mezi 4 dětmi
$X \sim Bi(n = 4, \pi = 0{,}49)$

V R: $P(X = k)$ … $dbinom(k, n, \pi)$

$P(X \leq k)$ … $pbinom(k, n, \pi)$

$P(X = 1) = \mathbf{0{,}260}$ $(dbinom(1{,}4{,}0.49))$



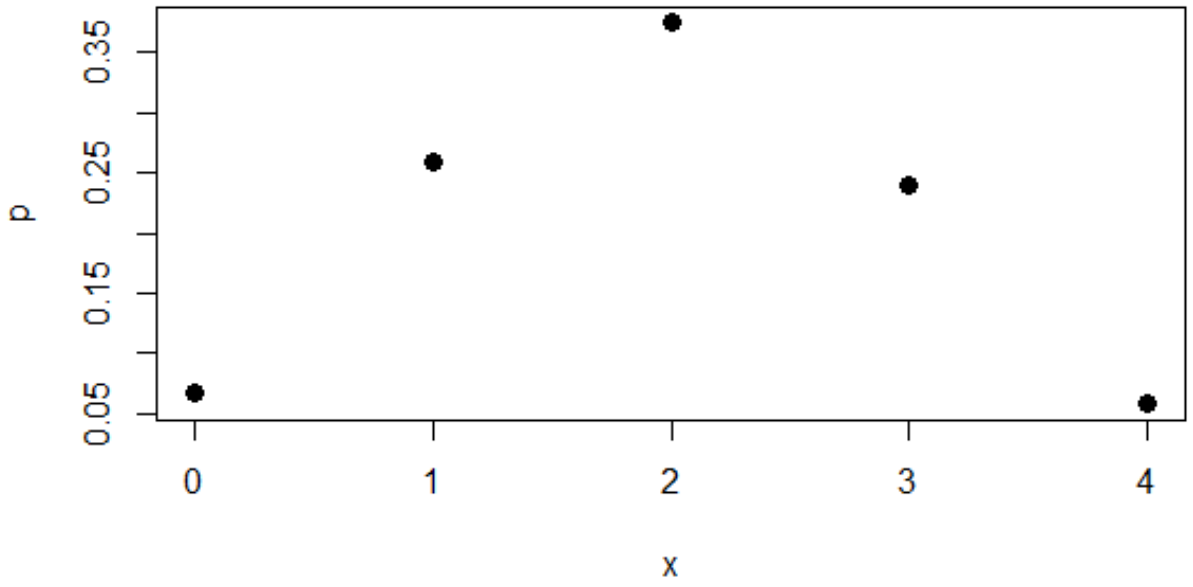
Příklad 1

Předpokládejme, že pravděpodobnost narození dívky je 0,49. Jaká je pravděpodobnost toho, že mezi čtyřmi dětmi v rodině je

a) je právě jedna dívka?

b) jsou méně než dvě dívky?

c) je více než jedna dívka?



V R: $P(X = k)$ ... $dbinom(k, n, \pi)$

$P(X \leq k)$ ... $pbinom(k, n, \pi)$

a) $P(X = 1) = \mathbf{0,260}$ $(dbinom(1,4,0.49))$

b) $P(X < 2) = P(X \leq 1) = \mathbf{0,328}$ $(pbinom(1,4,0.49))$

c) $P(X > 1) = 1 - P(X \leq 1) = \mathbf{0{,}672}$ =$(1 - pbinom(1,4,0.49))$

Binomické rozdělení

X … počet úspěchů v n Bernoulliho pokusech

$$X \sim Bi(n; \pi)$$

počet pokusů pravděpodobnost úspěchu

Pravděpodobnostní funkce: $P(X = k) = \binom{n}{k} \pi^k (1-\pi)^{n-k}$

Střední hodnota: $E(X) = n\pi$

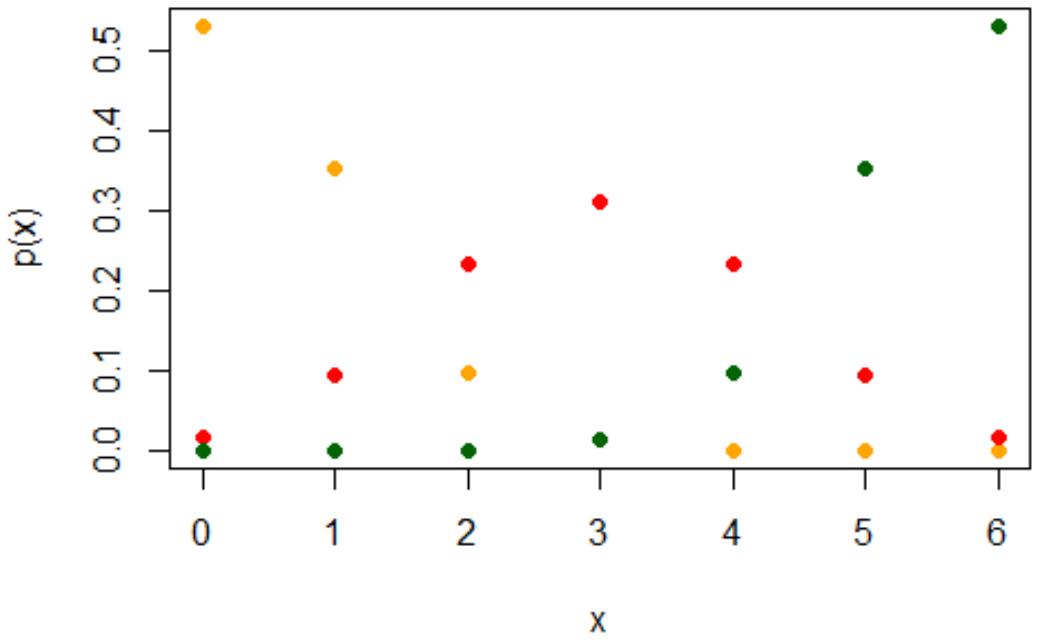
Rozptyl: $D(X) = n\pi(1-\pi)$ (dokažte!)

Příklady: počet chlapců mezi 10 000 novorozenci, počet vadných výrobků mezi 30 testovanými, počet nevzrostlých rostlin ze 100 zasazených cibulek…

Binomické rozdělení

X … počet úspěchů v n Bernoulliho pokusech

$$X \sim Bi(n; \pi)$$



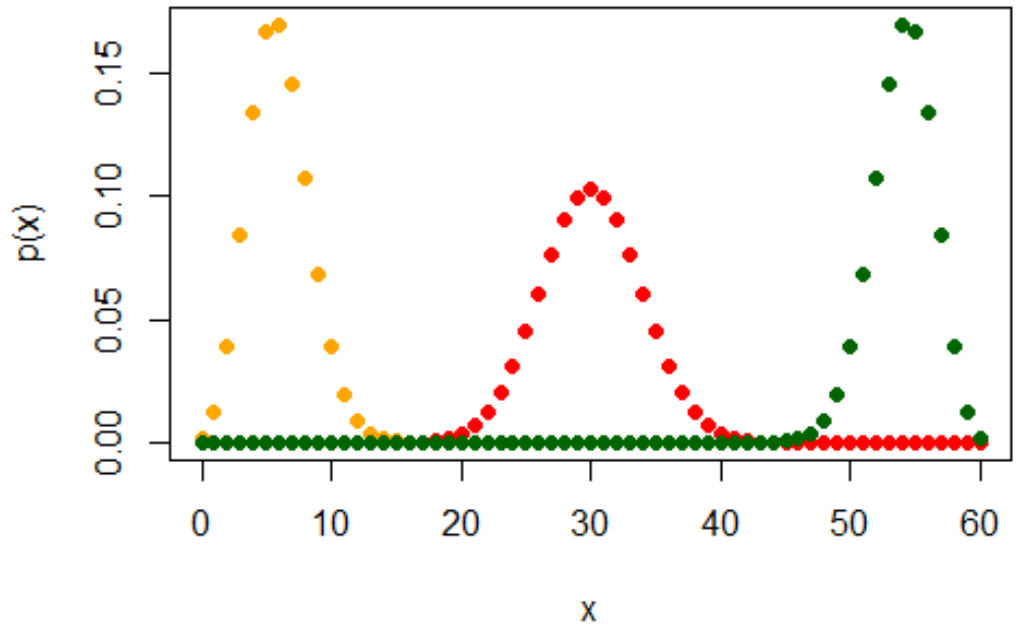
$X_1 \sim Bi(n = 6; \pi = 0{,}1)$
$X_2 \sim Bi(n = 6; \pi = 0{,}5)$
$X_3 \sim Bi(n = 6; \pi = 0{,}9)$

$\pi < 0{,}5$, malé n $\Rightarrow$ poz. zešikmené r.
$\pi = 0{,}5$ $\Rightarrow$ symetrické r.
$\pi > 0{,}5$, malé n $\Rightarrow$ neg. zešikmené r.

Binomické rozdělení

X … počet úspěchů v n Bernoulliho pokusech

$$X \sim Bi(n; \pi)$$



$X_1 \sim Bi(n = 60; \pi = 0{,}1)$
$X_2 \sim Bi(n = 60; \pi = 0{,}5)$
$X_3 \sim Bi(n = 60; \pi = 0{,}9)$

$\pi \neq 0{,}5$, s rostoucím n se rozdělení stává více a více symetrickým

Příklad 2

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?

X … počet prasklých vajec z 20 vybraných (výběr bez vracení)

celkem vajec
200

prasklých vajec
50

neprasklých vajec
150

$$P(X = 8) = \frac{\binom{50}{8}\binom{150}{12}}{\binom{200}{20}}$$

počet příznivých možností
– vybíráme 8 prasklých z 50 prasklých a zároveň 12 ze 150 neprasklých

počet všech možností
– vybíráme 20 vajec z 200

Hypergeometrické rozdělení

V souboru $\boldsymbol{N}$ prvků je $\boldsymbol{M}$ prvků s danou vlastností a zbylých $(\boldsymbol{N}-\boldsymbol{M})$ prvků tuto vlastnost nemá.

Postupně vybereme ze souboru $\boldsymbol{n}$ prvků, z nichž žádný nevracíme zpět.

X … počet prvků se sledovanou vlastností ve výběru n prvků

$$X \sim H(N; M; n)$$

rozsah základního souboru

počet prvků s danou vlastností

rozsah výběru

Hypergeometrické rozdělení

V souboru $\boldsymbol{N}$ prvků je $\boldsymbol{M}$ prvků s danou vlastností a zbylých $(\boldsymbol{N}-\boldsymbol{M})$ prvků tuto vlastnost nemá.

Postupně vybereme ze souboru $\boldsymbol{n}$ prvků, z nichž žádný nevracíme zpět.

X … počet prvků se sledovanou vlastností ve výběru n prvků

$$X \sim H(N; M; n)$$

Pravděpodobnostní funkce: $P(X=k)=\dfrac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$

Počet příznivých možností, tj. počet možností
jak vybrat k prvků s danou vlastností z M
a zároveň $(n-k)$ prvků, které uvedenou vlastnost nemají z $(N-M)$ prvků.

Počet všech možností, jak vybrat n prvků z N (nezáleží na pořadí).

rozsah základního souboru
N

Počet prvků
s danou vlastností
M

Počet prvků,
které danou vlastnost nemají
$N-M$

Příklad 2

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?

X … počet prasklých vajec z 20 vybraných (výběr bez vracení)

$X \sim H(N = 200, M = 50, n = 20)$

celkem vajec
200

prasklých vajec
50

neprasklých vajec
150

V R: $P(X = k)$ … $dhyper(k, M, N - M, n)$

$P(X \leq k)$ … $phyper(k, M, N - M, n)$

$P(X = 8) = \mathbf{0,057}$ $(dhyper(8,50,150,20))$

Hypergeometrické rozdělení

V souboru $\boldsymbol{N}$ prvků je $\boldsymbol{M}$ prvků s danou vlastností a zbylých $(\boldsymbol{N}-\boldsymbol{M})$ prvků tuto vlastnost nemá.

Postupně vybereme ze souboru $\boldsymbol{n}$ prvků, z nichž žádný nevracíme zpět.

X … počet prvků se sledovanou vlastností ve výběru n prvků

$$X \sim H(N; M; n)$$

Pravděpodobnostní funkce: $P(X=k)=\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$

Střední hodnota: $E(X)=n\cdot\frac{M}{N}$

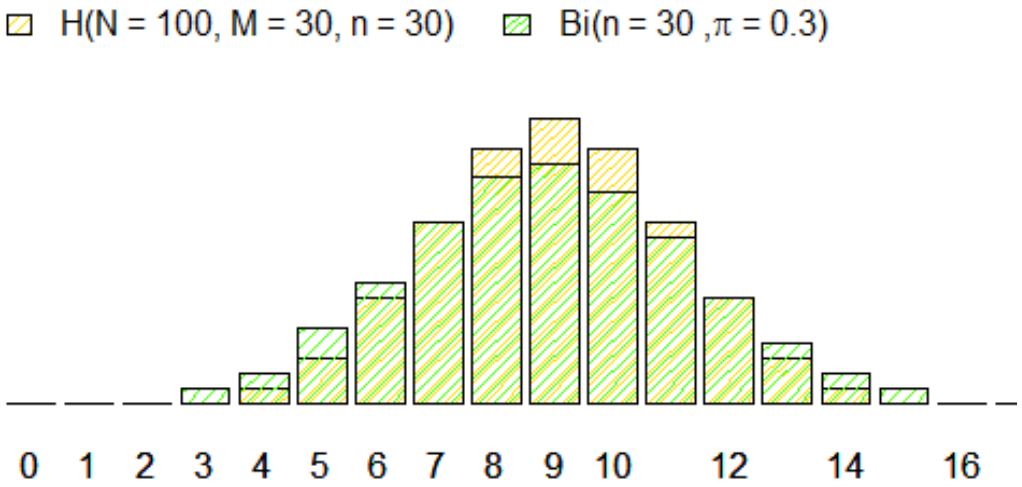
Rozptyl: $D(X)=n\cdot\frac{M}{N}\left(1-\frac{M}{N}\right)\left(\frac{N-n}{N-1}\right)$

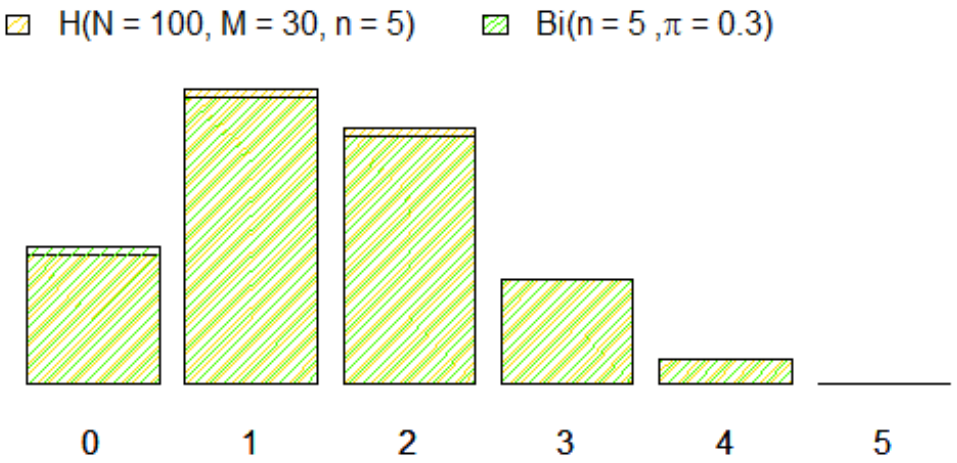
Příklady: počet vadných výrobků mezi 10 vybranými z dodávky 30 výrobků, mezi nimiž bylo 7 vadných, počet dívek v náhodně vybrané skupině 4 dětí ze třídy, v níž je 6 chlapců a 8 dívek, …

Hypergeometrické rozdělení – aproximace bin. rozdělením

Je-li n/N, tzv. **výběrový poměr**, menší než 0,05, lze hypergeometrické rozdělení nahradit binomickým s parametry n a M/N.

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





Negativně binomické (Pascalovo) rozdělení

X … počet Bernoulliho pokusů do k. výskytu události (úspěchu), včetně k. výskytu

$$X \sim NB(k;\ \pi)$$

požadovaný počet úspěchů (výskytů události)

pravděpodobnost úspěchu

Pravděpodobnostní funkce:

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

pravděpodobnost k. úspěchu

pravděpodobnost $k-1$ úspěchů v $n-1$ Bernoulliho pokusech

Negativně binomické (Pascalovo) rozdělení

X … počet Bernoulliho pokusů do k. výskytu události (úspěchu), včetně k. výskytu

$$X \sim NB(k;\ \pi)$$

požadovaný počet úspěchů (výskytů události)

pravděpodobnost úspěchu

Pravděpodobnostní funkce: $P(X = n) = \binom{n-1}{k-1} \pi^k (1-\pi)^{(n-k)}$

Střední hodnota: $E(X) = \frac{k}{\pi}$ **Rozptyl:** $D(X) = \frac{k(1-\pi)}{\pi^2}$

Příklady: počet dárců neznajících svou krevní skupinu, které musíte testovat proto, abyste našli 4 dárce s krevní skupinou 0, počet cestujících, které musí revizor zkontrolovat do chvíle, než najde 10 černých pasažérů, …

Příklad 3

Dle http://ksicht.natur.cuni.cz/serialy/detektivni-chemie/3 je pravděpodobnost výskytu krevní skupiny A+ 35%. V polní nemocnici nutně potřebují najít 3 dárce krve s touto krevní skupinou. Potenciálních dárců je dostatek, nikdo z nich však nezná svou krevní skupinu. Jaká je pravděpodobnost, že pro nalezení 3 dárců krevní skupiny A+, budeme muset postupně vyšetřit více než 7 a méně než 12 potenciálních dárců?

X … počet osob, které musíme vyšetřit, chceme-li najít 3 dárce s krevní skupinou A+
$X \sim NB(k = 3; \pi = 0{,}35)$

$$P(7 < X < 12) = \sum_{n=8}^{11} \binom{n-1}{2} 0{,}35^3 \cdot 0{,}65^{n-3} \cong 0{,}332$$

POZOR!

V případě negativně binomické náhodné veličiny není definice ustálená. Někteří statistici (popř. statistický software) ji definují jako počet neúspěchů před k. úspěchem.

Jak se to projeví v praktickém výpočtu pravděpodobnostní funkce?

Příklad 3

Dle http://ksicht.natur.cuni.cz/serialy/detektivni-chemie/3 je pravděpodobnost výskytu krevní skupiny A+ 35%. V polní nemocnici nutně potřebují najít 3 dárce krve s touto krevní skupinou. Potenciálních dárců je dostatek, nikdo z nich však nezná svou krevní skupinu. Jaká je pravděpodobnost, že pro nalezení 3 dárců krevní skupiny A+, budeme muset postupně vyšetřit více než 7 a méně než 12 potenciálních dárců?

X … počet osob, které musíme vyšetřit, chceme-li najít 3 dárce s krevní skupinou A+
$X \sim NB(k = 3; \pi = 0{,}35)$

V R:

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

$Y = X - 3$

$P(7 < X < 12) = P(X < 12) - P(X \leq 7) = P(X \leq 11) - P(X \leq 7) = P(Y \leq 8) - P(Y \leq 4) = \mathbf{0{,}332}$

$(pnbinom(8{,}3{,}0.35) - pnbinom(4{,}3{,}0.35))$

Geometrické rozdělení

Geometrické rozdělení je speciálním typem negativně binomického rozdělení pro $k = 1$.

X … počet Bernoulliho pokusů do prvního výskytu události (úspěchu), včetně.

$$X \sim Ge(\pi)$$

pravděpodobnost úspěchu

Pravděpodobnostní funkce: $P(X = n) = \pi(1 - \pi)^{(n-1)}$

Střední hodnota: $E(X) = \frac{1}{\pi}$ **Rozptyl:** $D(X) = \frac{1-\pi}{\pi^2}$

Příklady: počet volání nutných k tomu, abychom se dovolali do televizní soutěže, počet řidičů, kteří podstoupí test na obsah alkoholu v krvi do doby, než bude nalezen první podnapilý řidič, …

Poissonův proces

Bodový proces

Sledujeme chod procesu, v němž čas od času dochází k nějaké význačné události.

Registrujeme počet události $N(s;t)$ v časovém intervalu $\langle s; s+t \rangle$

Jako **Poissonův proces** označujeme proces, který je:

- ordinární – tj. pravděpodobnost výskytu více než jedné události v limitně krátkém časovém intervalu $(t \longrightarrow 0)$ je nulová. (tzv. **řídké jevy**),
- stacionární – tj. rychlost výskytu událostí λ je konstantní v průběhu celého sledovaného intervalu,
- beznásledný – tj. pravděpodobnost výskytu události není závislá na čase, který uplynul od minulé události.

Poissonův proces

- Rychlost výskytu události λ je parametrem Poissonova procesu.
- Poissonův proces lze obdobně jako v časovém intervalu definovat na libovolné uzavřené prostorové oblasti (na ploše, v objemu).

Poissonovo rozdělení

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*, resp. v objemu *t* s hustotou výskytu λ). Pak

X … počet výskytu události v uzavřené oblasti (v čase, na ploše, v objemu)

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

Pravděpodobnostní funkce: $P(X = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$

Střední hodnota, rozptyl: $E(X) = D(X) = \lambda t$

Příklady: počet pacientů ošetřených během dopoledních ordinačních hodin,

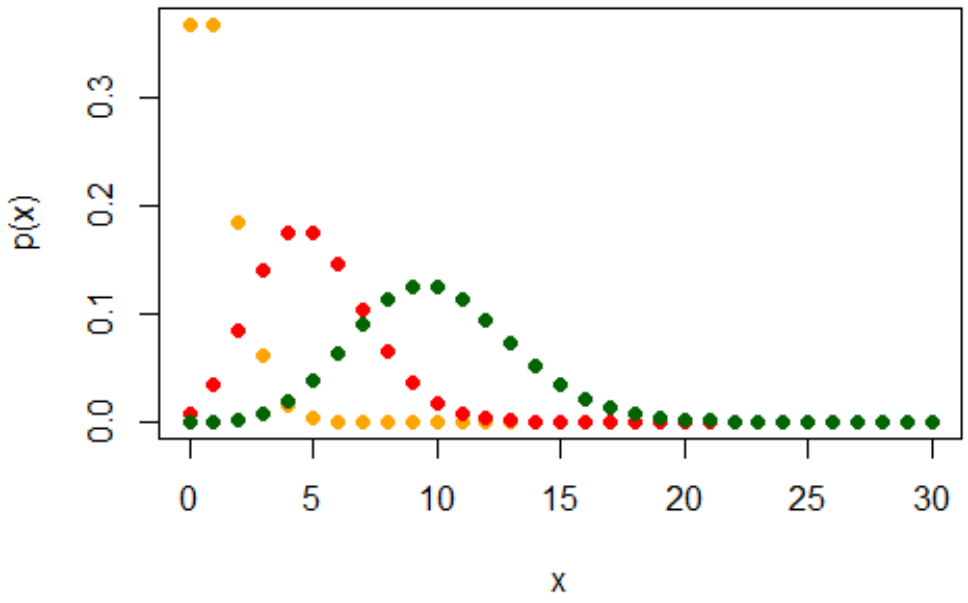
počet mikrodefektů na zadaném vzorku materiálu, počet mikroorganismů v 1 dl vody, …

Poissonovo rozdělení

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*, resp. v objemu *t* s hustotou výskytu *λ*). Pak

X … počet výskytu události v uzavřené oblasti (v čase, na ploše, v objemu)

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

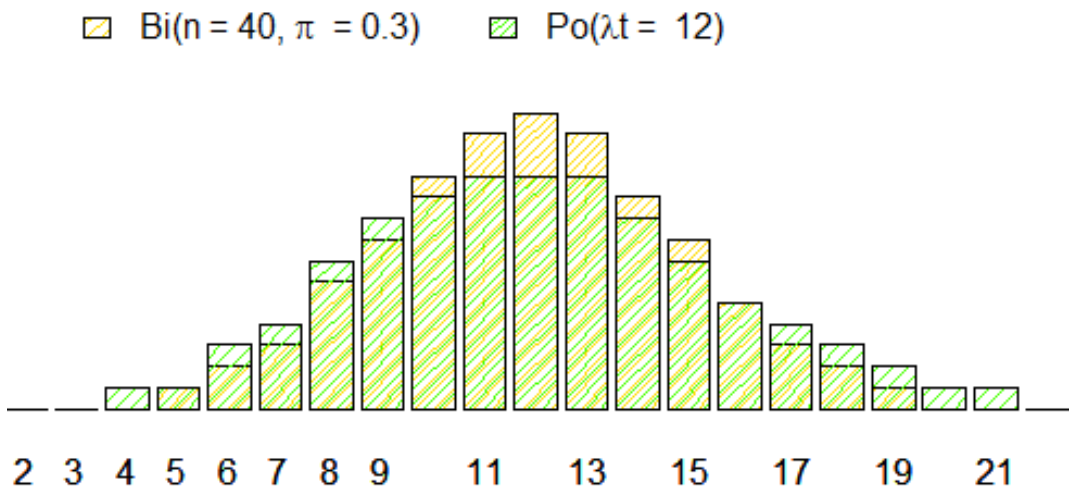


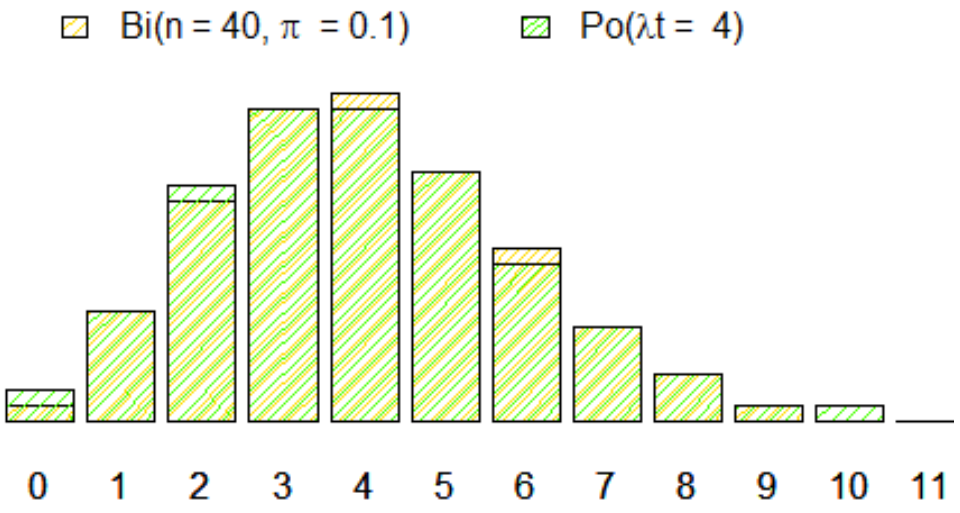
$X_1 \sim Po(\lambda t = 1)$
$X_2 \sim Po(\lambda t = 5)$
$X_3 \sim Po(\lambda t = 10)$

Poissonovo rozdělení – aproximace bin. rozdělením

Poissonovým rozdělením lze velmi dobře aproximovat binomické rozdělení pro případ, že počet pokusů n je dostatečně velký ($n > 30$) a pravděpodobnost výskytu události (úspěchu) je dostatečně malá ($\pi < 0{,}1$). V takovém případě je $\lambda t = n\pi$.

$$(n > 30 \ \wedge \ \pi < 0{,}1) \Rightarrow Bi(n, \pi) \approx Po(n\pi)$$





Příklad 4

Telefonní ústředna zapojí během hodiny průměrně 15 hovorů. Jaká je pravděpodobnost, že během 4 minut zapojí ústředna

a) právě jeden hovor?

b) alespoň dva hovory?

X … počet hovorů, které telefonní ústředna spojí během 4 minut
$X \sim Po(\lambda t = ?)$

$E(X) = \lambda t = 1$ (průměrně 15 hovorů za hodinu, tj. průměrně 1 hovor za 4 minuty)

Další příklad – video.

Příklad 4

Telefonní ústředna zapojí během hodiny průměrně 15 hovorů. Jaká je pravděpodobnost, že během 4 minut zapojí ústředna

a) právě jeden hovor?

b) alespoň dva hovory?

X … počet hovorů, které telefonní ústředna spojí během 4 minut
$X \sim Po(\lambda t = 1)$

$E(X) = \lambda t = 1$ (průměrně 15 hovorů za hodinu, tj. průměrně 1 hovor za 4 minuty)

V R:

a) $P(X = 1) = \mathbf{0,368}$ (*dpois(1,1)*)

b) $P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) = \mathbf{0,264}$ (*1-ppois(1,1)*)

Vybraná rozdělení diskrétní náh. veličiny - přehled

Popis	Podmínky		Název náhodné veličiny - symbolický zápis
počet úspěchů v n pokusech	nezávislé pokusy	$k = 1$	**Alternativní** - $A(\pi)$
			Binomická - $Bi(n, \pi)$
	závislé pokusy		**Hypergeometrická** - $H(N, M, n)$
počet pokusů do k. úspěchu (včetně)	nezávislé pokusy	$k = 1$	**Geometrická** - $Ge(\pi)$
			Negativně binomická (Pascalova) -$NB(k, \pi)$
počet událostí v uzavřené oblasti (v čase, na ploše, v objemu)	ordinarita, stacionarita, beznáslednost procesu		**Poissonova** - $\mathrm{Po}(\lambda t)$

Vybraná rozdělení diskrétní náhodné veličiny - přehled

Adéla Vrtková, Martina Litschmannová: R-tahák

Rozdělení pravděpodobnosti - Předpony	
r-	generování hodnot z určitého rozdělení
d-	hustota pravděpodobnosti $f(x)$ nebo pravděpodobnostní funkce $P(X = x)$
p-	$P(X \leq x)$
q-	kvantilová funkce $F^{-1}(x)$

Rozdělení pravděpodobnosti - Diskrétní	
-binom	Binomické rozdělení
-hyper	Hypergeometrické rozdělení
-nbinom	Negativně binomické rozdělení !jiná definice - počet neúspěšných pokusů!
-geom	Geometrické rozdělení !jiná definice - počet neúspěchů před prvním úspěchem!
-pois	Poissionovo rozdělení

Děkuji za pozornost!

martina.litschmannova@vsb.cz

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | FAKULTA ELEKTROTECHNIKY A INFORMATIKY | KATEDRA APLIKOVANÉ MATEMATIKY