

Využití MATLABu při práci s Poissonovým rozložením

Základní poznatky o Poissonově rozložení $Po(\lambda)$

Náhodná veličina $X \sim Po(\lambda)$ udává počet událostí, které nastanou v jednotkovém časovém intervalu případně v jednotkové oblasti, jestliže k událostem dochází náhodně, jednotlivě a vzájemně nezávisle. Parametr λ je střední hodnota počtu těchto událostí. Pokud sledujeme náhodnou veličinu, která udává počet událostí v intervalu délky t , pak uvedená náhodná veličina má rozložení $Po(t\lambda)$.

Pravděpodobnostní funkce:
$$P(X=x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda} & \text{pro } x = 0, 1, 2, \dots \\ 0 & \text{jinak} \end{cases}$$
, distribuční funkce:
$$F(x) = \sum_{t=0}^x \frac{\lambda^t}{t!} e^{-\lambda}$$
, $E(X) = \lambda$, $D(X) = \lambda$

Aproximace binomického rozložení pomocí Poissonova rozložení:

Nechť náhodná veličina $X \sim Bi(n, \theta)$. Za předpokladu, že $n \geq 30$ a $\theta \leq 0,1$, lze pravděpodobnostní funkci této náhodné veličiny uspokojivě aproximovat pravděpodobnostní funkcí rozložení $Po(n\theta)$:

$$P(X=x) \approx \frac{(n\theta)^x}{x!} e^{-n\theta}$$

Interval spolehlivosti pro střední hodnotu: Nechť X_1, \dots, X_n je náhodný výběr z rozložení $Po(\lambda)$ a necht' m je realizace výběrového průměru. Pak meze $100(1-\alpha)\%$ přibližného empirického intervalu spolehlivosti pro střední hodnotu λ jsou:

$$d = m - \sqrt{\frac{m}{n}} u_{1-\frac{\alpha}{2}}, \quad h = m + \sqrt{\frac{m}{n}} u_{1-\frac{\alpha}{2}}$$

MATLAB počítá meze $100(1-\alpha)\%$ přibližného empirického intervalu spolehlivosti pro střední hodnotu λ podle vzorce:

$$d = \frac{1}{2n} \chi^2_{\alpha/2, 2nm} - \frac{1}{2n}, \quad h = \frac{1}{2n} \chi^2_{1-\alpha/2, 2nm} + \frac{1}{2n}$$

Testování hypotézy o shodě parametrů dvou Poissonových rozložení: Necht' X_{11}, \dots, X_{1n_1} je náhodný výběr z rozložení $Po(\lambda_1)$ a necht' X_{21}, \dots, X_{2n_2} je na něm nezávislý náhodný výběr z rozložení $Po(\lambda_2)$.

Označme M_* vážený průměr výběrových průměrů, tj. $M_* = \frac{n_1 M_1 + n_2 M_2}{n_1 + n_2}$.

Na hladině významnosti α testujeme hypotézu $H_0: \lambda_1 = \lambda_2$ proti alternativě $H_1: \lambda_1 \neq \lambda_2$.

Testová statistika: $T_0 = \frac{\sqrt{n_1 (M_1 - M_*)^2 + n_2 (M_2 - M_*)^2}}{M_*}$.

Platí-li H_0 , $T_0 \approx N(0,1)$.

Kritický obor: $W = (-\infty, -u_{1-\alpha/2}] \cup [u_{1-\alpha/2}, \infty)$

Jestliže realizace t_0 testové statistiky T_0 patří do kritického oboru, H_0 zamítáme na asymptotické hladině významnosti α .

a) Kreslení grafu pravděpodobnostní a distribuční funkce rozložení Po(2)

```
x=[0:10]';  
pf=poisspdf(x,2);  
plot(x,pf,'o')  
df=poisscdf(x,2);  
figure  
stairs(x,df)
```

(Samostatný úkol: Jak nakreslit graf distribuční funkce bez svislých čar?)

Jedno z možných řešení:

```
hold on  
for i=1:(length(x)-2) plot([i,i],[0,1],'w'); end
```

b) Generování 100 realizací náhodné veličiny s rozložením Po(2) a kreslení histogramu

```
r=poissrnd(2,100,1);  
hist(r,x)
```

c) Odhad střední hodnoty a výpočet mezí intervalu spolehlivosti pro střední hodnotu na základě proměnné r

Hodnoty uložené v proměnné r považujeme za realizace náhodného výběru rozsahu 100 z rozložení Po(2)

```
[m,meze]=poissfit(r)
```

d) Výpočet střední hodnoty a rozptylu rozložení Po(2)

```
[m,v]=poisstat(2)
```

Příklady na využití Poissonova rozložení

Příklad 1.: Při provozu balicího automatu vznikají během směny náhodné poruchy, které se řídí rozložením $Po(2)$. Jaká je pravděpodobnost, že během směny dojde k aspoň jedné poruše?

Řešení: X – počet poruch během směny, $X \sim Po(2)$, $P(X \geq 1) = 1 - P(X < 1) = 1 - P(X = 0) = 1 - \frac{2^0}{0!} e^{-2} = 0,8647$.

V MATLABu: $p = 1 - \text{poisspdf}(0,2)$

Příklad 2.: Telefonní ústředna zapojí během hodiny průměrně 15 hovorů. Jaká je pravděpodobnost, že během 4 minut ústředna zapojí a) právě jeden hovor, b) aspoň dva hovory?

Řešení: X – počet zapojených hovorů během 4 minut = 1/15 hodiny, $X \sim Po(t\lambda)$, kde $t = 1/15$ a $\lambda = 15$, tedy $X \sim Po(1)$.

ad a) $P\{X = 1\} = e^{-1} = 0,36788$,

ad b) $P\{X \geq 2\} = 1 - P\{X \leq 1\} = 1 - P\{X = 0\} - P\{X = 1\} = 1 - 2e^{-1} = 1 - 2 \cdot 0,36788 = 0,264242$

V MATLABu: a) $p = \text{poisspdf}(1,1)$, b) $p = 1 - \text{poisscdf}(1,1)$

Příklad 3.: Ze zkušenosti víme, že při správné obsluze stroje je v průměru 0,1% výrobků zmetkových. Ke stroji nastoupil nový pracovník. Za týden vyrobil 5 000 kusů, z nichž 11 bylo zmetkových. Lze takto vysoký počet zmetků vysvětlit působením náhodných vlivů?

Řešení: Počítáme pravděpodobnost, že pracovník vyrobil aspoň 11 zmetků za předpokladu, že stroj je obsluhován správně. X – počet vyrobených zmetků za týden, $X \sim Bi(5000, 0,001)$. Při splnění podmínek dobré aproximace lze rozložení veličiny X aproximovat rozložením $Po(5)$.

$P\{X \geq 11\} = 1 - P\{X \leq 10\} = 1 - \sum_{i=0}^{10} \frac{5^i}{i!} e^{-5} = 1 - 0,9863 = 0,0137$.

Je zřejmé, že nový pracovník nepracuje správně.

V MATLABu: $p = 1 - \text{poisscdf}(10,5)$

Přesný výpočet v MATLABu: $p = 1 - \text{binocdf}(10,5000,0.001)$

Příklad 4.: Pro $n = 30$ a $p = 0,1$ ilustrujte aproximaci binomického rozložení $Bi(n, p)$ Poissonovým rozložením $Po(np)$. Vypočtené hodnoty obou pravděpodobnostních funkcí v bodech $x = 0, 1, \dots, 30$ zapište do tabulky.

Řešení:

```
x=[0:1:30];
pf1=binopdf(x,30,0.1);
pf2=poisspdf(x,3);
[x pf1 pf2]
```

Příklad 5.: Na výrobní lince se zhruba každé dvě hodiny vyskytne porucha. Sestrojte tabulku uvádějící, s jakou pravděpodobností se na této lince během osmihodinové pracovní směny nevyskytne žádná porucha, vyskytne jedna porucha, vyskytnou dvě poruchy atd. až vyskytne deset poruch. S jakou pravděpodobností nastane více než deset poruch? Určete nejpravděpodobnější počet poruch během osmihodinové směny.

Řešení:

X - počet poruch během osmi hodin, $X \sim Po(t\lambda)$, kde $t = 4$ a $\lambda = 1$, tedy $X \sim Po(4)$.

$$P\{X = 0\} = \frac{e^{-4} 4^0}{0!} = 0,0183$$

$$P\{X = 1\} = \frac{e^{-4} 4^1}{1!} = 0,0733$$

.....

$$P\{X = 10\} = \frac{e^{-4} 4^{10}}{10!} = 0,0053$$

Přehledný zápis všech výsledků viz následující tabulka (včetně kumulovaných hodnot čili distribuční funkce). Všimněte si, že ačkoli jsme samozřejmě nevyčerpali všechny možné počty poruch (těch je nekonečně mnoho), dosahuje součet všech vypočtených pravděpodobností téměř 1. Najdeme ho jako hodnotu distribuční funkce $\Phi(10)=0,9972$ a jde o pravděpodobnost, že počet poruch bude nejvýše deset. Celková pravděpodobnost, že by byl počet poruch naopak větší než deset, je tedy $1 - \Phi(10)$ a činí pouze necelá tři promile ($1 - 0,9972 = 0,0028$).

n	0	1	2	3	4	5	6	7	8	9	10
$\pi(n)$	0,0183	0,0733	0,1465	0,1954	0,1954	0,1563	0,1042	0,0595	0,0298	0,0132	0,0053
$\Phi(n)$	0,0183	0,0916	0,2381	0,4335	0,6288	0,7851	0,8893	0,9489	0,9786	0,9919	0,9972

Nejpravděpodobnější počet poruch během osmihodinové směny je tři až čtyři (hledáme modus, jde o bimodální případ).

V MATLABu:

```
x=[0:10]';
pf=poisspdf(x,4);
df=poisscdf(x,4);
[x pf df]
```

Grafické znázornění:

```
plot(x,pf,'o')
stairs(x,df)
```

Příklad 6.: Semena rostlin určitého druhu jsou znečištěna malým množstvím plevelů. Je známo, že na jedné jednotce plochy vyrostou po osetí v průměru 4 rostliny plevelů. Vypočítejte pravděpodobnost, že na dané jednotce plochy:

- Nebude žádný plevel,
- Vyrostou nejvýše 3 rostliny plevelů,
- Vyrostou aspoň 5, ale nejvýše 7 rostlin plevelů.

Výsledek: ad a) 0,0183, ad b) 0,4335, ad c) 0,32.

Příklad 7.: V prodejně posunuli zavírací dobu ve všední dny z 18 na 19 hodin. Sestrojte 90% přibližný empirický interval spolehlivosti pro střední hodnotu počtu zákazníků v této době, navštívilo-li prodejnu ve 30 náhodně zvolených dnech ve sledované době celkem 225 zákazníků. Přitom předpokláme, že počet zákazníků v určitém časovém intervalu má Poissonovo rozložení. Meze intervalu spolehlivosti vypočtete jak pomocí aproximace normálním rozložením, tak bez této aproximace.

Výsledek: ad a) $6,68 < \lambda < 8,32$, ad b) $6,70 < \lambda < 8,38$

Příklad 8.: Na jisté VŠ si v podzimním semestru roku 2010 předmět Statistika zapsalo 711 osob, z toho 351 mužů a 360 žen. V průběhu zkouškového období bylo zaznamenáno u mužů 68 neúspěšných pokusů o složení zkoušky z tohoto předmětu a u žen 108 neúspěšných pokusů. Na hladině významnosti 0,05 testujte hypotézu, že střední hodnota počtu neúspěšných pokusů připadajících na jednoho muže je stejná jako střední hodnota počtu neúspěšných pokusů připadajících na jednu ženu.

Výsledek: Průměrný počet neúspěšných pokusů připadajících na jednoho muže je 0,1937, průměrný počet neúspěšných pokusů připadajících na jednu ženu je 0,3. Testová statistika se realizuje hodnotou $t_0 = 2,8474$, kritický obor má tvar $w = (-\infty, -1,96) \cup (1,96, \infty)$, H_0 zamítáme na asymptotické hladině významnosti 0,05.