DÚ č.4 - Poissonův proces, systémy hromadné obsluhy

Marek Nevole, Jan Novotný

ČVUT - FIT

{nevolmar, novot103}@fit.cvut.cz

21. března 2022

1 Úvod

Ve čtvrtém úkolu z předmětu vybrané statistické metody jsme se zabývali Poissonovými procesy a systémy hromadné obsluhy. Za reprezentanta byl zvolen Marek Nevole.

Úkol jsme vypracovali pomocí programovacího jazyka Python[1] v prostředí Jupyter Notebook[2] s volně dostupnými knihovnami SciPy[3], NumPy[4], Seaborn[5] a Matplotlib[6].

2 Popis problému

Uvažujte model hromadné obsluhy $M|G|\infty$.

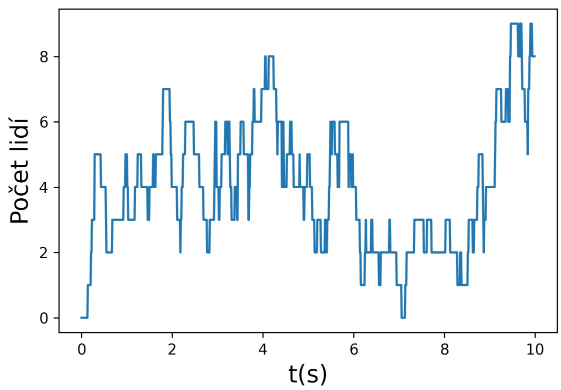
- *Požadavky přichází podle Poissonova procesu s intenzitou $\lambda = 10\ \mathrm{s}^{-1}$.*
- *Doba obsluhy jednoho požadavku (v sekundách) má rozdělení $S \sim \mathrm{Ga}(4, 2)$, tj. Gamma s parametry $a = 4, p = 2$.*
- *Časy mezi příchody a časy obsluhy jsou nezávislé.*
- *Systém má (teoreticky) nekonečně paralelních obslužných míst (každý příchozí je rovnou obsluhován).*

Označme N_t počet zákazníků v systému v čase t. Předpokládejme, že na začátku je systém prázdný, tj. $N_0 = 0$.

3 Úloha č.1

Simulujte jednu trajektorii $\{N_t(\omega) \mid t \in (0, 10\ \mathrm{s})\}$. Průběh trajektorie graficky znázorněte.

Zákazníci přichází podle Poissonova procesu s intenzitou $\lambda = 10\ s^{-1}$. Počet příchozích zákazníků v intervalu $[s, t]$ odpovídá Poissonovu rozdělení přírůstků $N_t - N_s \sim \mathrm{Poisson}(\lambda(t - s))$, tedy počet zákazníků této úlohy je z rozdělení Poisson(100). Toto rozdělení je implementováno v knihovně SciPy jako *poisson* a pro náhodný výběr obsahuje metodu *rvs*, které jsme předali parametr *mu=100*. Náhodný výběr z tohoto rozdělení vrátil hodnotu $n = 98$. Časy jednotlivých příchodů zákazníků odpovídají rovnoměrnému rozdělení $U(0, t)$. Tedy jsme udělali 95 náhodných výběrů z rozdělení $U(0, 10)$, pomocí třídy *uniform* a metody *rvs* s parametry *scale=t, size=n*. Doba obsloužení těchto zákazníků je z rozdělení $\mathrm{Ga}(4, 2)$ s parametry $a = 4, p = 2$. Gamma rozdělení je implementováno jako *gamma* s metodou pro náhodný výběr *rvs*. Parametry pro tuto metodu jsou *shape* a *scale*, v našem studijním textu používáme parametry, které odpovídají parametrizaci *shape* a *rate*. Pro $\mathrm{Ga}(a, p)$ je *shape* = p a *rate* = a. Mezi *scale* a *rate* lze převádět pomocí vzorce scale $= \frac{1}{\text{rate}}$. Tedy po 98 náhodných výběrech z $\mathrm{Ga}(4, 2)$ jsme dostali intervaly všech zákazníků v čase a výslednou trajektorii lze pozorovat na obrázku 1.



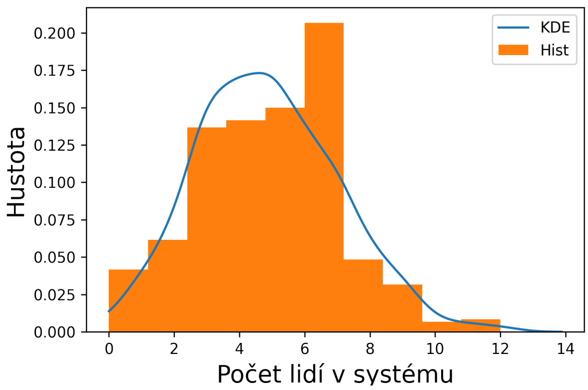
Obrázek 1: Jedna trajektorie $\{N_t(\omega) \mid t \in (0, 10\ \mathrm{s})\}$.

4 Úloha č.2

Simulujte $n = 500$ nezávislých trajektorií pro $t \in (0, 100)$. Na základě těchto simulací odhadněte rozdělení náhodné veličiny N_{100}.

V této úloze jsme využili kód z předchozí úlohy

[1] python.org
[2] jupyter.org
[3] scipy.org
[4] numpy.org
[5] seaborn.pydata.org
[6] matplotlib.org



Obrázek 2: Odhad rozdělení N_{100} pomocí histogramu a jádrového odhadu.

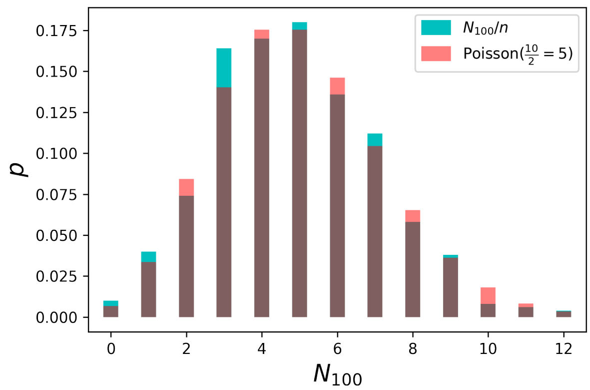
a simulovali jsme 500 trajektorií. Na základě hodnot z posledního časového kroku, tedy v $t = 100$, jsme sestavili histogram a odhad hustoty rozdělení pomocí jádrového odhadu, toto lze pozorovat na obrázku 2. Dle zkoumané veličiny můžeme tvrdit, že pozorované rozdělení bude diskrétní a svojí podobou připomíná Poissonovo rozdělení s parametrem $\lambda = 5$.

5 Úloha č.3

Diskutujte, jaké je limitní rozdělení tohoto systému pro $t \to +\infty$. Pomocí vhodného testu otestujte na hladině významnosti 5 %, zda výsledky simulace N_{100} odpovídají tomuto rozdělení.

Zákazníci přicházejí dle Poissonova rozdělení s intenzitou $\lambda = 10\ s^{-1}$. Střední hodnota doby obsluhy odpovídá $E\ Ga(4, 2) = 0.5$, tedy $\mu = \frac{1}{EGa} = 2$, z přednášky 23 víme, že z dlouhodobého hlediska $t \to +\infty$ má počet zákazníků v systému Poissonovo rozdělení s intenzitou λ/μ. Tedy pro tuto úlohu $N_t \sim \text{Poisson}(10/2)$.

Úlohou bylo otestovat, zda rozdělení výsledků simulace odpovídá teoretickému rozdělení. K tomuto jsme využili test dobré shody, v tomto případě se známými parametry. Za nulovou hypotézu jsme postavili rovnost pravděpodobností $H_0 : \boldsymbol{p} = \boldsymbol{p'}$, kde $\boldsymbol{p}$ jsou teoretické pravděpodobnosti Poissonova procesu s intenzitou $\lambda = 5$ a $\boldsymbol{p'}$ jsou relativní četnosti z 500 simulací N_{100}, alternativní hypotézou byla nerovnost $H_A : \boldsymbol{p} \neq \boldsymbol{p'}$. Překryv těchto pravděpodobností lze pozorovat na obrázku 3. Testovou statistikou byla χ^2 statistika. Jelikož naměřená data nesplňovala všude podmínku teoretických četností $np_i \geq 5$, tak jsme místo pravděpodobností $P(N_{100} = 0)$ a $P(N_{100} = 1)$ vybrali pouze $P(N_{100} \leq 1)$ a pro všechna $i > 10$ nahradili



Obrázek 3: Graf relativních pozorovaných četností N_{100} a pravděpodobnosti hodnot z rozdělení Poisson(5).

$P(N_{100} = i)$ za $P(N_{100} > 10) = 1 - P(N_{100} \leq 10)$. Knihovna SciPy nabízí funkci *chisquare*, která na vstupu bere naměřené a teoretické četnosti a vrátí testovou statistiku společně s p-hodnotou. Dle $\hat{p} > \alpha$ nemůžeme zamítnout hypotézu shody.

$$\chi^2 = 8.293$$
$$\hat{p} = 0.600$$
$$\chi^2_{0.05,10} = 18.307$$