Zápočtová úloha z predmetu Pravdepodobnosť a štatistika
akad. rok 2008/2009
Radoslav Glinský

Pred tým, než som začal rôzne výpočty, som si najprv nasimuloval príchody ľudí do staníc metra. Vieme, že doba medzi príchodmi ľudí do staníc A a B sa riadi exponenciálnym rozdelením a priemerne na stanicu príde 40 resp. 30 ľudí. Použijeme teda funkciu rexp(), ktorá mi náhodne generuje čas príchodu ľudí. Na základe tohto si vytvoríme vektory vektor_prichodov_A a vektor_prichodov_B s časmi príchodov (v minútach).
Okrem tohto som si z nasimulovaných dát odvodil aj pre každého cestujúceho poradie vlaku, ktorým sa odviezol. Toto poradie dostaneme jednoducho tak, že čas príchodu cestujúceho na stanicu vydelíme bez zbytku periódou, akou vlaky jazdia – v našom prípade 2 minúty (k tomu ešte musíme pripočítať číslo 1). Na základe poradia vlaku zvýšim počet cestujúcich v tomto vlaku s každým cestujúcim o 1. Tieto počty cestujúcich mám uloženú vo vektore, konkrétne vektor_poctu_ludi_vo_vlaku.
Použil som seed(25121987).

Riešenia 1. časti:

1. Počet cestujúcich v prvom vlaku medzi stanicami B a C

Na začiatku som si zastrojil vektor_poctu_ludi_vo_vlaku, ktorý má dlžku 180 = počet všetkých dopoludnajších vlakov. Počet cestujúcich v prvom vlaku získam jednoducho tak, že vypíšem vektor_poctu_ludi_vo_vlaku[1], tj. prvý index vektoru.

2. Odhady strednej hodnoty počtu cestujúcich v dopoludnajšom vlaku

Bodovým odhadom strednej hodnoty je aritmetický priemer. Na zostrojený vektor použijem funkciu mean(vektor_poctu_ludi_vo_vlaku). Pre moje nasimulované data dostaneme, že bodový odhad počtu ľudí v jednom dopoludnajšom vlaku je 140,2889.

Pre intervalový odhad platí:

$$\left(\overline{x} - u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}; \overline{x} + u_{1-\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Podľa tohto vzorca:

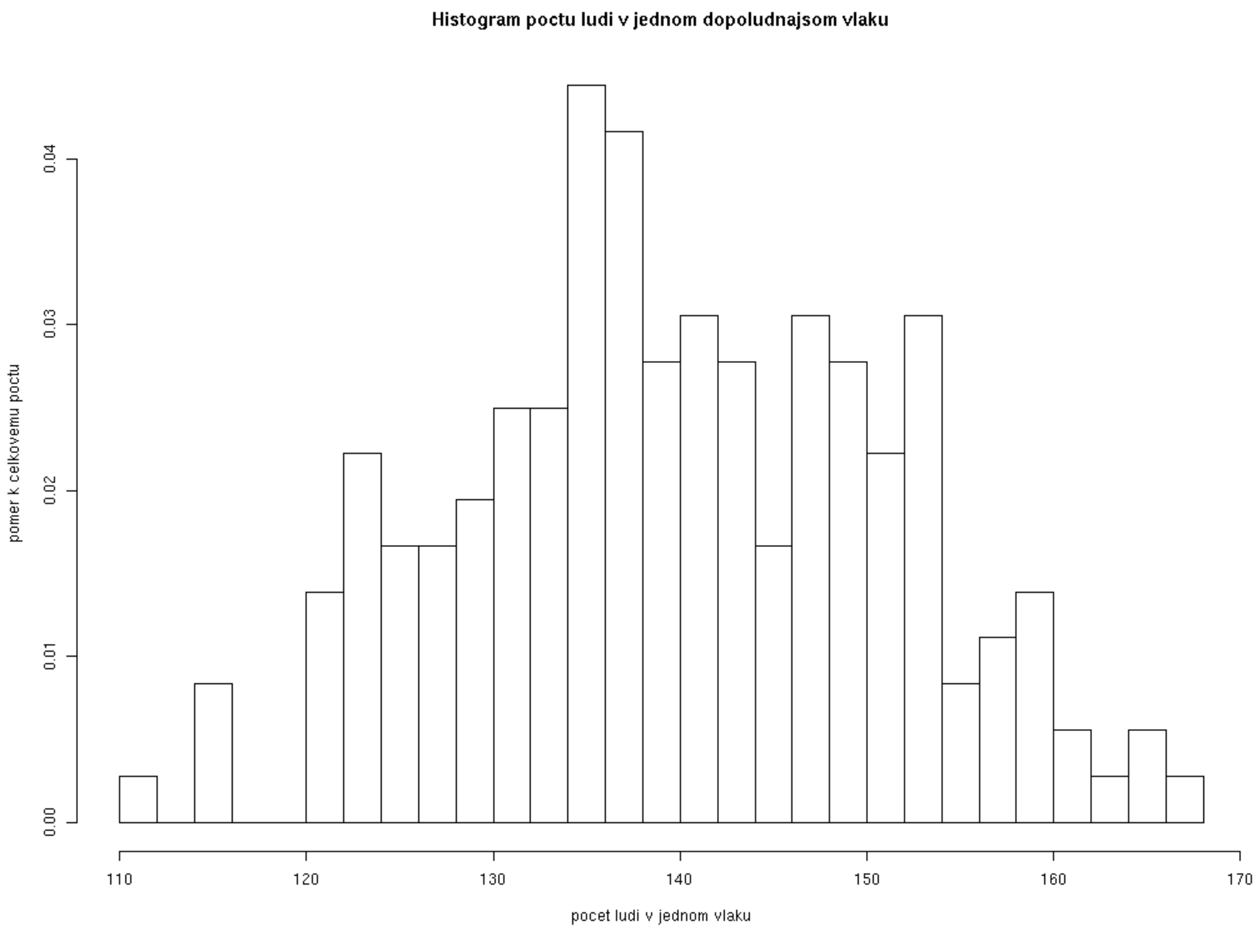
- strednú hodnotu spočítam ako aritmetický priemer, tj. mean(vektor_poctu_ludi_vo_vlaku)
- smerodatnú odchylku spočítam ako sd(vektor_poctu_ludi_vo_vlaku)
- pre výpočet kvantilu najprv uvažme nasledujúcu aproximáciu:

  $$\lambda \to \infty \text{ platí } Po(\lambda) \approx N(\mu = \lambda; \sigma^2 = \lambda)$$

- takže to znamená, že Poissonove rozdelenie môžem v našom prípade aproximovať normálnym rozdelením
- pre výpočet kvantilu použijem qnorm(0.975)

95% intervalový odhad počtu ľudí v jednom dopoludnajšom vlaku vyšiel 138,6206 – 141,9572.

4. Histogram pre počty cestujúcich v dopoludnajších vlakoch



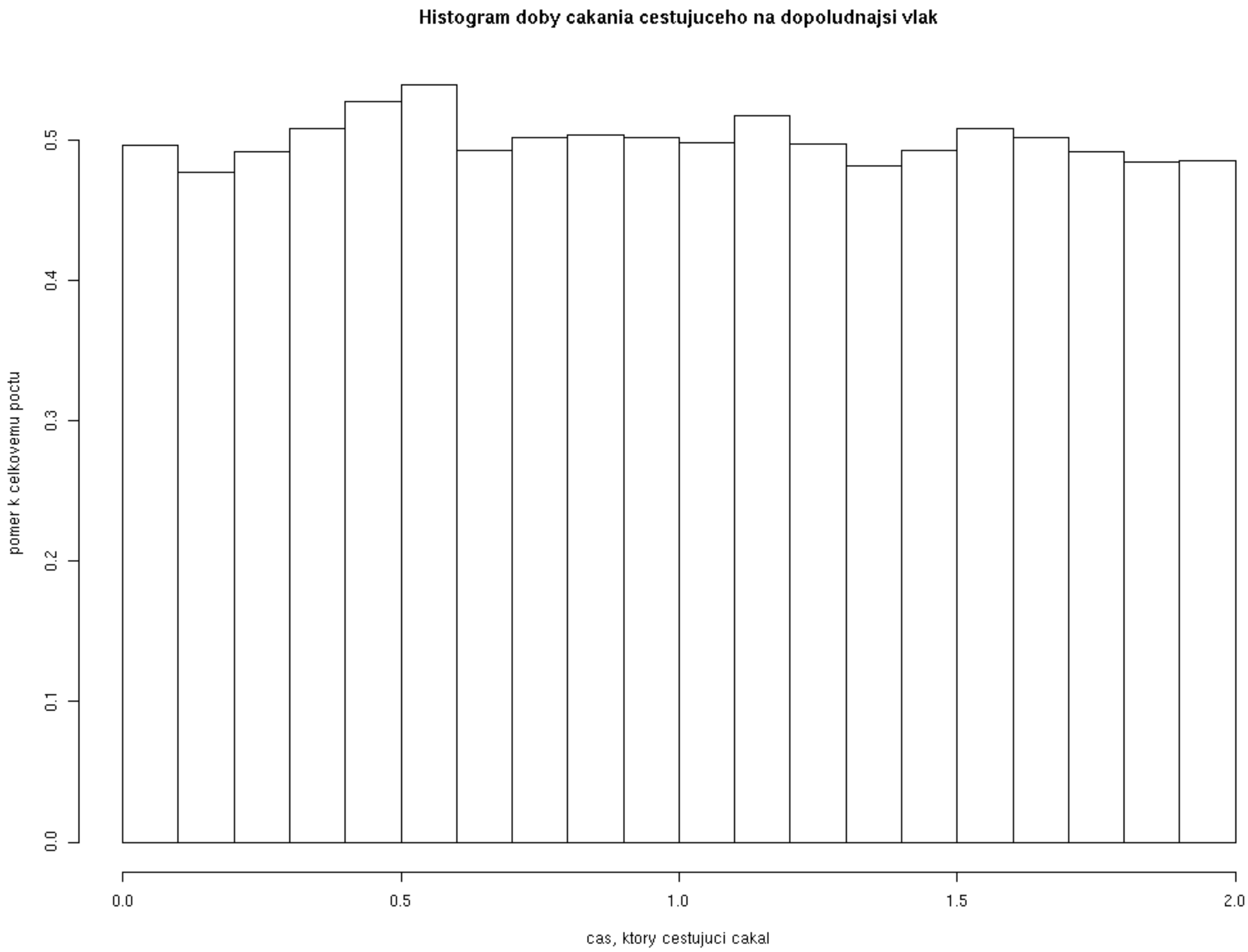
5. Aké má rozdelenie počet cestujúcich medzi B a C v jednom dopoludnajšom vlaku?

Pozrime sa najprv na definíciu Poissonovho rozdelenia:
U tohto rozdelenia nadobúda náhodná veličina X hodnot {0, 1, 2, 3, …}. Pokiaľ je počet výskytov nejakej udalosti behom časového intervalu priamo úmerný dlžke časového intervalu a priemerný počet výskytov udalosti za konštantnú časovú jednotku je Labmbda (Lambda > 0), potom náhodná veličina X ~ Po(Lambda) charakterizuje počet výskytov udalosti za konštantnú časovú jednotku.

Tato definícia presne odpovedá nášmu prípadu, počtu cestujúcich v jednotlivých vlakoch. Preto má tento počet Poissonove rozdelenie.
Poznámka: Poissonove rozdelenie (aproximované na normálne rozdelenie) som použil v druhej úlohe.

7. Rozdelenie doby čakania na vlak metra

Z nasimulovaných príchodov ľudí na stanice som si pre každého cestujúceho určil dobu, ktorú musel čakať na vlak. Túto dobu vypočítam ako čas príchodu na stanicu %% 2. Jednotlivé doby čakania som si uložil do vektora s dlžkou rovnou počtu všetkých dopoludnajších cestujúcich, konkretne vektor_cakania_na_vlak. Na základe tohto si vykreslíme histogram:



Ak sa pozrieme na histogram, ľako si odvodíme, že doba čakania na vlak bude mať rovnomerné rozdelenie.

8. Navrhovaná kapacita vlakovej súpravy

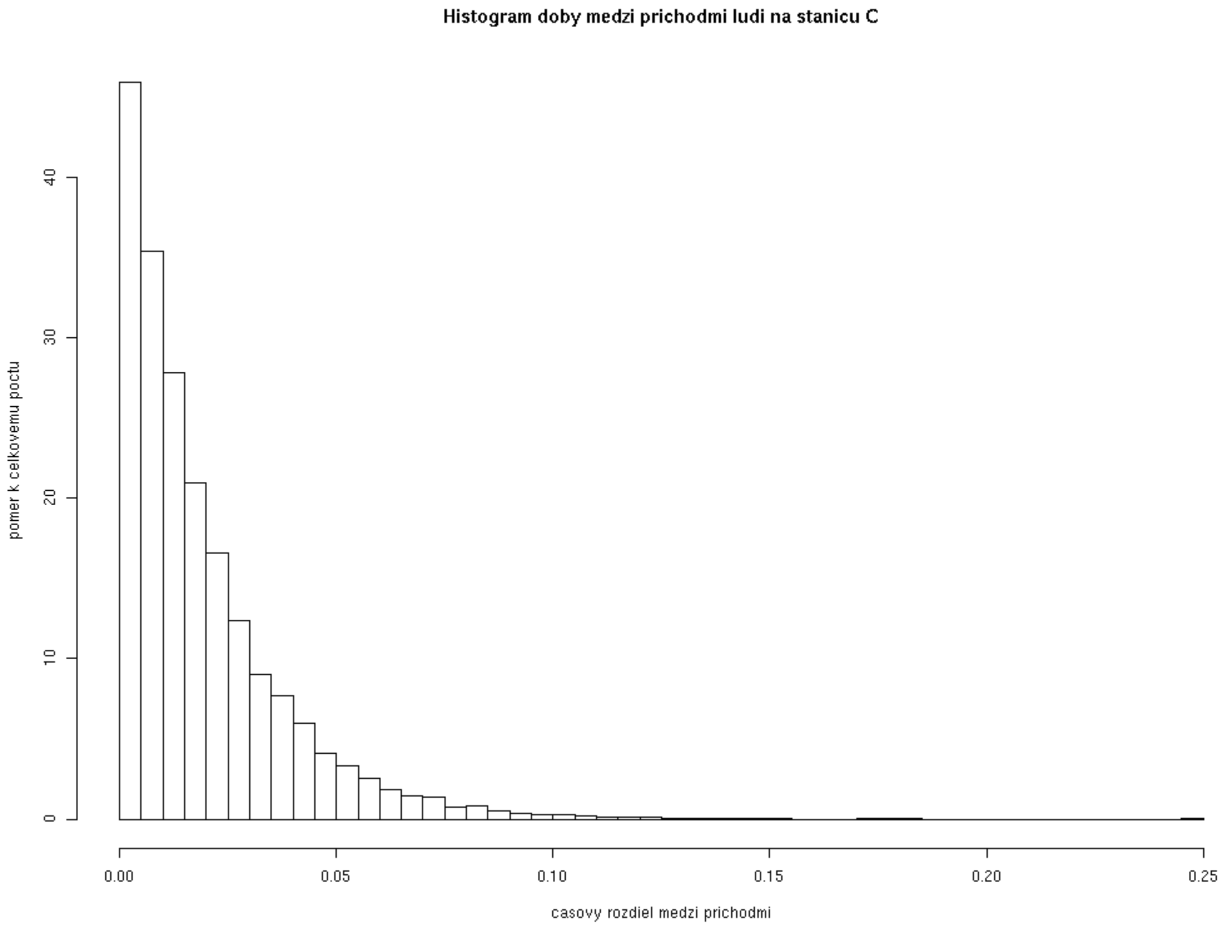
Použijeme funkciu pre kvantil Poissonovho rozdelenia, konkrétne qpois(0.98,arit_priemer).
Kde arit_priemer je bodový odhad strednej hodnoty počtu cestujúcich v jednom vlaku.
Pre nasimulované data sme dostali výsledok, že navrhovaná kapacita vlakovej súpravy je 165.

Riešenia 2. časti:

Pred riešením som si najprv nasimuloval príchody ľudí na stanicu C popoludní. Kedže čas príchodu je náhodný, použil som funkciu runif(pocet_ludi,0,480). Výsledok zotriedil a uložil do vektora, konkrétne vektor_prichodov_C.

9. Rozdelenie doby medzi príchodmi jednotlivých cestujúcich

Jedným prechodom vektora vektor_prichodov_C si vyrobím další vektor (vektor_casovych_rozdielov), ktorý obsahuje časové rozdiely medzi príchodmi jednotlivých ľudí. Na základe neho si vytvorím nasledujúci histogram.



Vidíme, že doba medzi príchodmi jednotlivých cestujúcich sa riadi exponenciálnym rozdelením.

10. Odhady strednej hodnoty počtu cestujúcich v jednom popoludnajšom vlaku

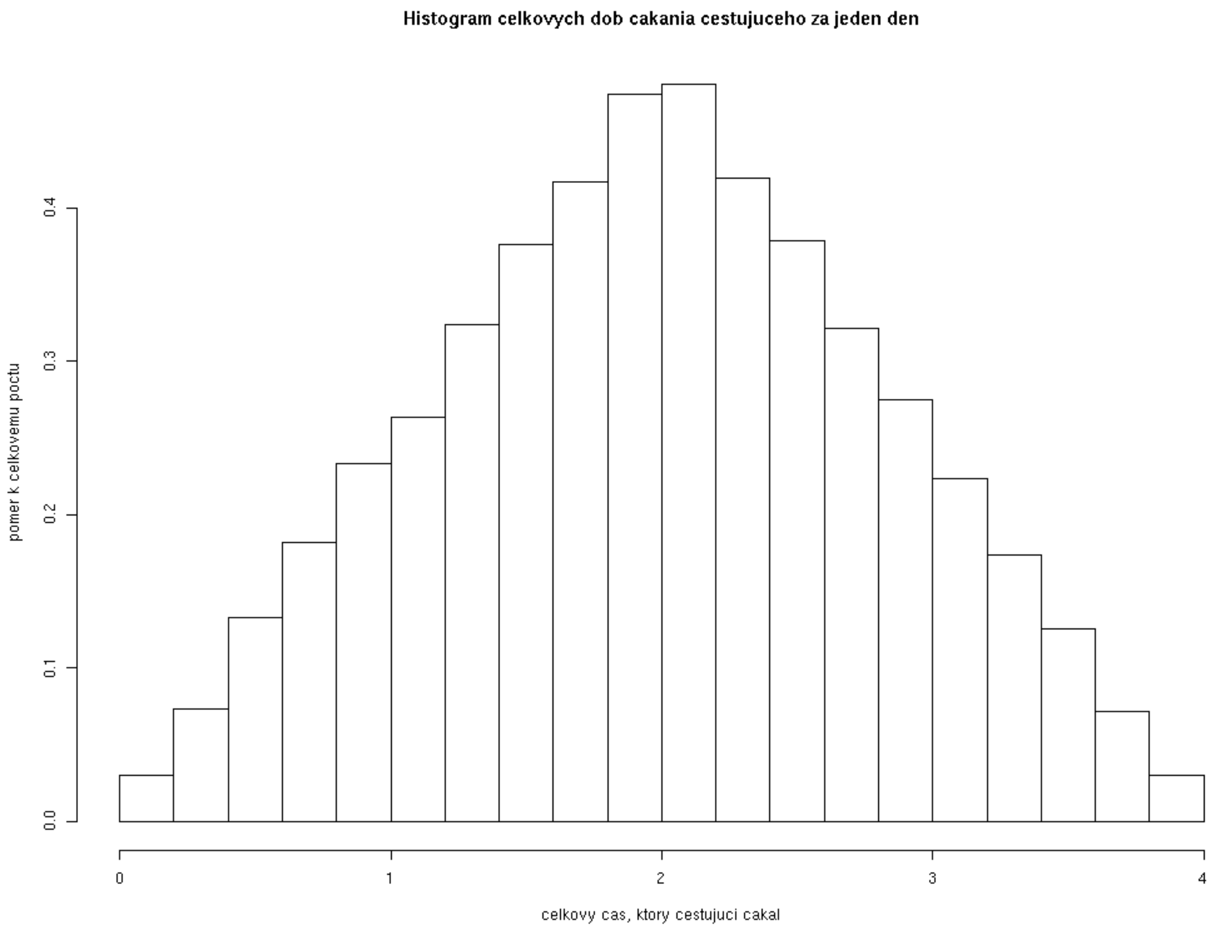
Postupujeme rovnako ako pri riešení 2. úlohy.
Bodový odhad počtu ľudí v jednom popoludnajšom vlaku je 105,2167.
95% intervalový odhad počtu ľudí v jednom popoludnajšom vlaku vyšiel 103,7970 – 106,6363.

11. Disproporcia medzi stredným počtom cestujúcich dopoludnia a popoludní

Stredný počet cestujúcich v jednom dopoludnajšom vlaku je vyšší než v popoludnajšom. Je to zapríčinené tým, že dopoludnia jazdí 180 vlakov, kdežto popoludní až 240 vlakov.
Pri predpoklade, že počet cestujúcich je rovnaký dopoludnia aj popoludní, priemerný počet cestujúcich dopoludnia musí byť väčší.

12. Rozdelenie doby čakania na vlak pre jedného človeka za jeden den

Celková doba čakania na vlak pre jedného človeka je konvolúciou 2 na sebe nezávislých čakaní dopoludnia a popoludní. Pre každého človeka si vypočítam celkový čas, ktorý strávil čakaním, a to ako súčet čakaní dopoludnia aj popoludní. Hodnoty mám uložené vo vektore vektor_celkovych_dob_cakania. Na jeho základe zostrojím histogram:



Vidíme, že graf má približne trojuholníkový tvar.
Hustota hľadaného rozdelenia bude: x/4 …....... x patrí [0, 2]
1 – x/4 ….. x patrí (2,4]