Seminarska naloga 1

Alen Kahteran

2. 11. 2020

Contents

Uvod

V nalogi bomo prek dveh različnih simulacij poskusili čim bolj natančno izračunati konstanto $e = 2.71828182845904523536028747135 27....$

e se lahko zapiše kot vsota neskončne vrste

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = \frac{1}{1} + \frac{1}{1} + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} \cdots ,$$

ali kot limita,

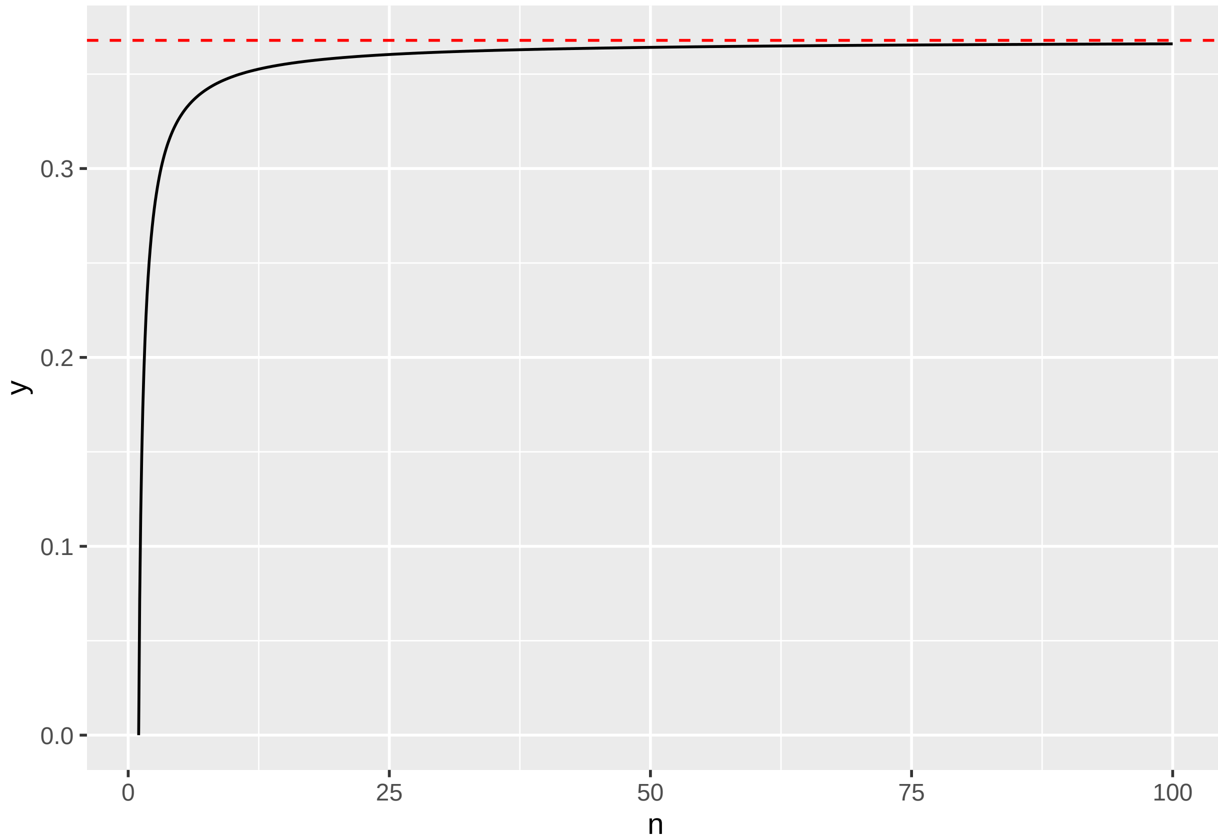
$$\lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n = e\,.$$

Podobno prejšnji enačbi velja tudi slednji zapis, ki je za nas bolj zanimiv

$$\lim_{n \to \infty} \left(1 - \frac{1}{n}\right)^n = \frac{1}{e}\,.$$

Nižje je grafični prikaz zgornje enačbe kjer je $n \in [1, 100]$, črna črta predstavlja naše podatke, črtkana rdeča črta pa $\frac{1}{e}$.

```
data <- data.frame(n=seq(1, 100, 0.1))
data$y <- (1-1/data$n)^data$n
ggplot(data, aes(n, y)) +
  geom_line() +
  geom_hline(yintercept=1/exp(1), linetype="dashed", color = "red")
```



Prvi način (a) računanja konstante e bomo opravili prek simuliranja Bernoullijevega procesa, medtem ko bomo drugega (način b) opravili prek simuliranja dveh normalno porazdeljenih spremenljivk. Načina bomo med seboj primerjali na več različnih načinov, da, med drugim, ugotovimo, kateri izmed njiju je boljši.

Način a

Bernoullijev proces predstavlja zaporedno izvajanje enakih neodvisnih Bernoullijevih poskusov, kjer sta možna samo dva izzida. V našem primeru bo to *0* in *1*. Ker lahko izzid zasede samo dve vrednosti, lahko rečemo da se podreja Bernoullijevi porazdelitvi. Funkcijo verjetnosti lahko zapišemo kot

$$f(k;p) = \begin{cases} p & \text{če je } k = 1, \\ 1-p & \text{če je } k = 0, \\ 0 & \text{v ostalih primerih} \end{cases}.$$

To lahko zapišemo tudi kot

$$\Pr(X=k) = f(k;p) = p^k(1-p)^{1-k}.$$

Lahko vidimo da je Bernoullijeva porazdelitev poseben primer binomske porazdelitve, kadar je $n = 1$.

$$\Pr(X=k) = f(k;n,p) = \binom{n}{k} p^k (1-p)^{n-k}$$

Torej, verjetnost da izberemo vrednost k iz populacije velikosti n je

$$\Pr(X = k) = \frac{1}{n}\,.$$

Če to ustavimo v funkcijo verjetnosti za Bernoullijevo porazdelitev dobimo

$$\Pr(X = k) = f(k;p) = \left(\frac{1}{n}\right)^k \left(1 - \frac{1}{n}\right)^{1-k}.$$

Iz tega lahko razberemo da je verjetnost da iz populacije velikosti n ne izvlečemo točno določene vrednosti a enaka

$$\Pr(X \neq a) = \left(1 - \frac{1}{n}\right).$$

Verjetnost da n zaporednih izvlečkov ne dobimo a lahko zapišemo kot

$$\Pr(X_1 \neq a \ in \ X_2 \neq a \ in \ \ldots \ in \ X_n \neq a) = \Pr(X_1 \neq a) \cdot \Pr(X_2 \neq a) \cdot \ \ldots \ \cdot \Pr(X_n \neq a) = \left(1 - \frac{1}{n}\right)^n$$

Kot vidimo je to zelo podobno enačbi omenjene v uvodu.

$$\lim_{n\to\infty} \left(1 - \frac{1}{n}\right)^n = \frac{1}{e}\,.$$

Se pravi, če nam uspe odsimulirati verjetnost da n zaporednih izvlečkov ne dobimo a, vemo da bo ta približno $\frac{1}{e}$ (približno zaradi naključnosti, večji kot bo n bolj bomo blizu vrednosti $\frac{1}{e}$).