

Magistrski študij Uporabna statistika
Napredni pristopi v programskem okolju R
2020/2021

2. seminarska naloga

Seminarsko nalogo naredite samostojno.

V tej seminarski nalogi čim bolj pohitrite kodo za naslednja dva problema. Pri tem ne uporabiti paralelizacije! Ni mišljeno, da bi povezali R s C++ (knjižnica Rcpp).

1. V raziskavi smo merili 1000 spremenljivk na 100 enotah. Polovica enot je pripadala prvi skupini, polovica pa drugi. Za vsako spremenljivko naredimo test t glede na skupino. Izid je vrednost testne statistike. Podatki naj bodo generirani iz standardne normalne porazdelitve, enote pa naključno razporejene v dve skupini.

Generiranje podatkov:†

```
set.seed(1234)
m <- 1000
n <- 100
X <- matrix(rnorm(m * n, mean = 0, sd = 1), nrow = n)
skupina <- rep(1:2, each = n / 2)
```

Osnovna koda za ta problem:‡

```
system.time(for(i in 1:m) t.test(X[, i] ~ skupina)$stat)
```

2. Enak problem kot prej, le da namesto testa t izvedete test χ^2 in tokrat vrednosti spremenljivk generate iz Bernoullijeve porazdelitve z verjetnostjo (parametrom) 0.5 (namesto normalne porazdelitve). Izid naj bo zopet vrednost testne statistike.

Na učilnici oddajte **Rmd dokument in pdf/html** (ob oddaji html ne uporabljajte šumnikov), ki naj za vsakega izmed problemov vsebuje naslednje:

- Vključite osnovno kodo problema in določitev časa izvedbe s funkcijo `system.time()`. V pdf/html naj bo viden izpis potrebnega časa na vašem računalniku.
- Vaša hitra koda naj bo že pripravljena za določitev časa izvedbe s funkcijo `system.time()` in naj vsebuje ukaze za nalaganje vseh morebitnih potrebnih paketov. V pdf/html naj bo viden izpis potrebnega časa na vašem računalniku.
- Pokažite, da je rezultat vaše hitre kode enak tistemu, ki ga dobite z osnovno kodo.
- Opišite, kako ste pohitrili kodo, s katerimi spremembami ste pridobili največ časa in zakaj.
- Kako si lahko s to kodo pomagamo pri oceni verjetnosti napake prve vrste testa s simulacijami? Naredite to oceno.
- Celotna koda mora delovati na operacijskem sistemu Windows.

† Pri hitrejših računalnikih morda kakšne pohitritve kode ne bodo dobro merljive. V tem primeru povečajte m in n .

‡ Kodo je na nekaterih računalnikih mogoče pohitrili tudi do 1000-krat, na drugih računalnikih pa na primer okoli 100-krat pri enakih izboljšavah, zato splošno vodilo o željeni hitrosti kode ni mogoče. Skušajte jo čim bolj pohitrili z orodji, predstavljenimi na vajah.