

Računarska statistika

Snježana Lubura Strunjak

Zagreb, 13. svibnja 2021.

Metode ponovnog uzorkovanja (re-sampling)

Jackknife metoda

Koristi se za eliminaciju pristranosti (biasa) procjenitelja nekog parametra.

Razvila se sredinom prošlog stoljeća (Quenouille, Miller, Tuckey), i pojavom bootstrap metode se prestala koristiti.

Opis metode:

- Neka je $\bar{\theta}$ procjenitelj parametra θ definiran na slučajnom uzorku X_1, \dots, X_n .
- Podijelimo taj uzorak na N poduzoraka iste duljine M tako da vrijedi $n = NM$.
- Za svaki $i = 1, \dots, N$ definirajmo novi uzorak na način da iz početnog uzorka izbacimo i -ti po redu poduzorak duljine M . Na osnovu novog uzorka definirajmo procjenitelj $\bar{\theta}_{-i}$ parametra θ .
- Tada je Jackknife procjena parametra u oznaci $J(\bar{\theta})$ definirana s

$$J_{-i}(\bar{\theta}) = N\bar{\theta} - (N-1)\bar{\theta}_{-i}, \quad i = 1, \dots, N$$

$$J(\bar{\theta}) = \sum_{i=1}^N \frac{J_{-i}(\bar{\theta})}{N}.$$

Kada je $M = 1$, onda se metoda zove complete Jackknife (važan poseban slučaj). Svojstva Jackknife procjenitelja:

- Ako $\bar{\theta}$ ima pristranost (bias) reda $\frac{1}{n}$, onda $J(\bar{\theta})$ ima pristranost reda $\frac{1}{n^2}$.
- Omogućava kreiranje robusnih intervala pouzdanosti koji se baziraju na Jackknife procjeni standardne pogreške:

$$se_{J(\bar{\theta})} = \sqrt{\frac{\sum_{i=1}^N (J(\bar{\theta}) - J_{-i}(\bar{\theta}))^2}{N(N-1)}}.$$

- Ne ponaša se dobro na podacima s outlierima (transformacije podataka možda mogu pomoći), niti za procjenitelje koji se ne mogu linearno aproksimirati (npr. medijan).

Primjer (Jackknife procjena varijance)

Koristiti programe:

CHAPTER1_3_JACKKNIFE.SAS, *JACKBOOT.SAS* i data set *law.sas7bdat* iz foldera Data. Prije izvođenja programa pomoću naredbe *LIBNAME* definirati library reference *lib*.

Primjer (Jackknife procjena korelac. koef. Bivarijatnog normalnog uzorka)

Program *CHAPTER1_3_NORMAL2_BOOT.SAS* i
CHAPTER1_3_JACKKNIFE CORR.SAS

Generirajte 15 parova točaka (X_i, Y_i) po bivarijatnoj normalnoj distribuciji ($n=15$) sa vrijednosti korelacijskog koeficijenta (Pearsonovog) uzorka = 0.562.

Izračunajte Jackknife procjene za

- Koeficijent korelacije,
- Standardnu pogrešku,
- Pristranost (bias),
- 95% interval pouzdanosti.

Zadaća

7. zadaća: rok za predaju 20.05.

Zadaća se nalazi u folderu Zadaće na MERLINU.

UPUTE: Svaki zadatak iz zadaće mora biti u svom .sas programu. Sve .sas programe nazovite na način *prezime_ime_zad1.sas*, ako je npr. 1.zadatak u pitanju, itd. Sve što radite u zadaćama mora biti u obliku koda (možete koristiti sve dostupne materijale da dobijete tražene rezultate, ali sve mora biti napisano u obliku koda).