

MC meetodid, praktikum 11

Teema 1: Parameetiline vs mitteparameetiline bootstrap

Parameetrilise bootstrap meetodi korral sobitatakse andmetele mingi jaotus (või mudel) ning seejärel genereeritakse valimeid selle jaotuse või mudeli kohaselt. Muus osas käib huvipakkuva parameetri hindamine ja usaldusintervalli leidmine samamoodi, nagu tavalise bootstrap meetodi korral.

Harjutus 1

Olgu meil juhusliku X 20st väärtusest koosnev valim

```
## [1] 39044 7619 17811 1086 12133 7522 44146
## [8] 14856 2208 13030 9005 8118 49708 2474
## [15] 2563 17895 8391 21620 7244 4227
```

Hinda bootstrap meetodil vastava jaotuse 0,8-kvantiili (kasutades statistikuna empiirilise jaotuse vastavat kvantiili) koos usaldusnivoole 0,9 vastava usaldusintervalliga, kasutades 10000 bootstrap valimit. Seejärel kontrolli Shapiro-Wilk testi abil, et logaritmitud andmed võivad vastata normaaljaotusele ning leia vastavad parameetrid $\hat{\mu}, \hat{\sigma}$ (st sobitame andmetele lognormaaljaotuse). Seejärel leia bootstrap hinnang ja bootstrap usaldusvahemik juhul, kui bootstrap valim genereeritakse vastavalt jaotusele $LN(\hat{\mu}, \hat{\sigma})$. Kui suured on erinevused hinnangutes? Kumba tulemust praegusel juhul kasutaksite?

```
valim <- c(39044, 7619, 17811, 1086, 12133, 7522, 44146, 14856, 2208, 13030,
           9005, 8118, 49708, 2474, 2563, 17895, 8391, 21620, 7244, 4227)
```

tulemused peaks välja nägema ligikaudselt sellised:

```
## [1] "Tavaline bootstrap"
## [1] "Punkthinnang: 14663"
## [1] "Usaldusintervall: (-3897,24250)"
## [1] "Parameetiline bootstrap"
## [1] "Punkthinnang: 21482"
## [1] "Usaldusintervall: (10182,30086)"
```

Harjutus 2

Uuri katseliselt 30-se valimi põhjal tavalise bootstrap meetodi abil leitud dispersiooni hinnangu 90% usaldusvahemiku õigsust juhul, kui tegelik lähtejaotus on $Po(5)$. Selleks võta $m = 100$ korda 30-elementiline valim, leia selle jaoks bootstrap usaldusintervall ja kontrolli, kas õige dispersioon kuulub sinna sisse või mitte. Kas tulemus on kooskõlas valitud usaldusnivooga?

Teema 2: Jackknife meetodi kasutamine hinnangu nihke vähendaiseks ja hinnangu dispersiooni hindamiseks

Bootstrap meetod võib osutuda väga ajakulukaks, kuna statistiku jaotuse piisava täpsusega leidmiseks tuleb sageli kasutada väga suurt arvu bootstrap valimeid.

Jackknife meetodi arvutuseeskiri on järgmine. Olgu hinnangu leidmiseks kasutatav statistik S . Tähistame kujul $S(i)$ sama statistikut, mis on hinnatud ilma i -nda vaatluseta. Defineerime

$$S_{(i)}^* = nS - (n-1)S(i)$$

Jackknife hinnang nihkele on

$$(n-1) \left(\frac{1}{n} \sum_{i=1}^n S(i) - S \right)$$

Jackknife hinnang parameetrile on

$$S^* = \frac{1}{n} \sum_{i=1}^n S_{(i)}^* = nS - \frac{n-1}{n} \sum_{i=1}^n S(i)$$

Jackknife hinnangu standardhälbe hinnang on

$$s_{jack}^* = \sqrt{\frac{n-1}{n} \sum_{i=1}^n \left(S(i) - \frac{1}{n} \sum_{j=1}^n S(j) \right)^2} = \sqrt{\frac{1}{n(n-1)} \sum_{i=1}^n (S_{(i)}^* - S^*)^2}$$

Turkey (1958) valem jackknife hinnangu usaldusintervallile olulisusnivool α :

$$(S^* - t_{n-1, 1-\frac{\alpha}{2}} s_{jack}^*, S^* + t_{n-1, 1-\frac{\alpha}{2}} s_{jack}^*),$$

kus $t_{k,\alpha}$ tähistab vabadusastmete arvuga k t-jaotuse α -kvantiili.

Harjutus 3

Leida esimeses ülesandes toodud andmete korral üldkogumi kesväärtuse ruudu tavapärase valemi abil hindamisel tehtava nihke hinnang, selle suuruse jackknife hinnang ja viimase 90% usaldusintervall. Vihje: kui $\mathbf{X}$ on vektor, siis $\mathbf{X}[-i]$ annab $\mathbf{R}$ -is sama vektori ilma i -nda komponendita. Leia vastavad hinnangud ka tavalise bootstrap meetodi abil.

Kodutöö nr. 8 (VPL harjutus, 1,5 punkti, tähtaeg 20. november 2022)