

Kodutöö

Kui ülesande lahendamisel kasutad arvuti abi, siis lisa tööle ka arvutuste tegemiseks kasutatud R'i programm!

Ülesanne 1

Teame Tallinna börsi tootluseid (protsentides) aastatel 2009-2020:

47,2 72,6 -23,9 38,22 11,38 -7,66 19,06 19,63 15,49 -6,38 10,05 5

Üheks olulisemaks investeeringu tootlust iseloomustavaks näitajaks on geomeetriline keskmine tootlus, vaata näiteks:

<https://www.investopedia.com/terms/g/geometricmean.asp>

<https://www.seb.ee/foorum/investeeringu/kuidas-investeeringu-tootlust-arvutada>

Hinda Tallinna börsi keskmist tootlust kasutades geomeetrilist keskmist tootlust ja leia ka saadud hinnangu standardviga (geomeetrilise keskmise tootluse standardhälve). Arvutuste tegemisel eelda, et aastased tootlused on teineteisest sõltumatud (tegelikult see eeldus pole päriselt korrektne).

Standardvea hindamisel kasuta

a) Bootstrap-meetodil

b) Delta meetodil.

Kommentaar: Geomeetriline keskmine on eksponent aastaste juurdekasvude logaritmitud väärtuste keskmisest. Aga keskmise dispersiooni oskad Sa ju juba leida... ja transformeeritud juhusliku suuruse dispersiooni leidmiseks on võimalik kasutada delta-meetodit.

Ülesanne 2

Meie käsutuses on järgmine valim:

3 2 0 1 3 2 1 0 2 1

Teame, et uuritava tunnuse jaotus on järgmine:

x	0	1	2	3
$P(X=x)$	$2p/3$	$p/3$	$2(1-p)/3$	$(1-p)/3$

Leia hinnang parameetrile p . Millist meetodit kasutad, milline on lõpptulemus, pane kirja ka arvutuskäik! Leia ka parameetri p hinnangu standardhälve (võib ka ligikaudne, arvutiga). Kui kasutad ülesande lahendamiseks arvuti abi siis lisa ka kasutatud programm lahendusele!

Suurima tõepära hinnangu leidmine numbrilise maksimiseerimise abil

Vahel on tõepära või log-tõepärafunktsioon piisavalt keeruline, et tõepärafunktsiooni maksimumi leidmine osutub tülikaks ülesandeks. Sellisel juhul võime hinnangu (ligikaudseks) leidmiseks kasutada numbrilisi meetodeid. Vaatame alljärgnevalt näidet, kuidas saab R-is kasutada numbrilist maksimiseerimist.

Oletame antud juhul, et vaatlused peaksid olema normaaljaotusega (sellisel juhul oskame parameetrite hinnanguid suurima tõepära meetodil leida ka täpselt ja seega saame hiljem numbriliste meetodite abil leitud hinnanguid võrrelda täpsete tulemustega):

```
# Vaatlused
x=c(12, 14, 16)

# Suurima tõepära meetodil leitud hinnangud keskväärtusele ja
# standardhälbele:
mean(x)
sqrt(mean((x-mean(x))**2))

# Hinnangud numbriliste meetodite abil

# Variant 1

# defineerime log-tõepära
l=function(arg, andmed){
  mu=arg[1]
  sigma=arg[2]
  l=sum(log( dnorm(andmed, mean=mu, sd=sigma ) ))
  return(l)
}

# Maksimiseerime log-tõepära (funktsiooni l):
tul=optim(c(2,2), l, andmed=x, control=list(fnscale=-1))
tul

# Leitud parameetrite hinnangud
tul$par
# Funktsiooni maksimum:
tul$value
```

Eelnev programm võib anda välja ka hoiatusi, sest numbrilised meetodid võivad proovida ka negatiivseid sigma väärtuseid – me pole antud programmile ette öelnud, et teise parameetri väärtused peavad olema mittenegatiivsed. Üheks lahenduseks oleks tagada, et sigma väärtus on alati mittenegatiivne – ükskõik mis siis teise parameetri väärtuseks ka poleks. Seda saab teha näiteks nii:

```
l2=function(arg, andmed){
  mu=arg[1]
  sigma=exp(arg[2])
  l=sum(log( dnorm(andmed, mean=mu, sd=sigma ) ))
  return(l)
}

x=c(10,12,14)
tul=optim(c(2,2), l2, andmed=x, control=list(fnscale=-1))
# Hinnatud keskvaartus
tul$par[1]
# Hinnatud standardhälve
exp(tul$par[2])
```

Ülesanne 3

Möödeti 120 mehe ja 80 naise pikkused, saadi järgmised tulemused:

```
set.seed(1)
naised=rnorm(120, mean=168, sd=6)
mehed=rnorm(80, mean=182, sd=7)
par(mfrow=c(2,1))
hist(mehed, xlim=c(150, 210), col="skyblue")
hist(naised, xlim=c(150, 210), col="pink")
```

Paraku läks kaduma informatsioon selle kohta, kes on mees, kes naine. Teame vaid nende 200 inimese pikkuseid (ja meeste/naiste proportsiooni), aga millised mõõtmised on naiste, millised meeste omad – seda me ei tea:

```
pikkused=c(mehed, naised)
par(mfrow=c(1,1))
hist(pikkused, col="gray70")
rm(naised, mehed)
```

Teie ülesanne on hinnata pikkuste jaotuse parameetrid (meeste pikkuste keskmine, meeste pikkuste standardhälve, naiste pikkuste keskmine, naiste pikkuste standardhälve) kasutades olemasolevaid andmeid (st teate teate pikkuseid, aga mitte inimese sugu) suurima tõepära meetodit kasutades. Pange tähele, et pikkuste jaotuseks on jaotuste segu ehk täistõenäosuse valemit kasutades:

$$F_{PIKKUS}(x) = \text{meeste_osakaal} F_{PIKKUS|Mees}(x) + \text{naiste_osakaal} F_{PIKKUS|Naine}(x)$$

Kus siis nii naiste kui ka meeste pikkuste jaotusteks on (erinevad) normaaljaotused.