

Biostatistika praktikum

Nakkushaigused. SIR mudeli hindamine.

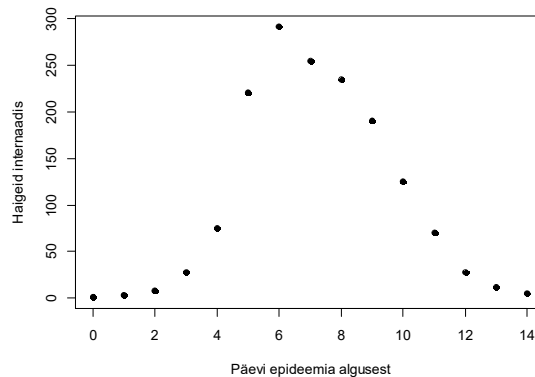
SIR-mudel sõltub kahest parameetrist, β ja γ . Kui parameetri γ kohta ka meil sageli on mingit informatsiooni ($\gamma \approx 1/\text{nakkusliku perioodi pikkus}$), siis parameetri β (mis iseloomustas mitmele inimesele päevas haigusekandja oma haigust jagab) väärtus on sageli populatsiooniti või inimgrupiti väga erinev. Sestap tekib sageli vajadus hinnata parameetri β või parameetrite β ja γ väärtust vaatlusandmete pealt.

Näide 1

Alustame ühest suhteliselt lihtsast ja klassikalisest näitest. Ühes Inglismaa poistekoolis (internaatkool) leidis 1978.a aset gripipuhang. Gripilaine koolis sai alguse ühestainsast õpilasest. Väidetavalt 1-2 päeva peale haigussümptomite ilmumist saadeti õpilane internadist ära – seega ei saanud ta peale seda aega enam uusi õpilasi nakatada (seega võiks oletada, et $\gamma \approx 1/2$). Koolis õppis gripiepideemia algusajal kokku 763 poissi. Haiguspuhangu andmed näevad välja järgmised (Näide võetud: Sulsky, D. (2015). Using real data in an SIR model):

```
# Tegelikud haigusjuhud
I_tegelik= c(1, 3, 8, 28, 75, 221, 291, 255, 235, 190, 125, 70, 28, 12, 5)
aeg=0:(length(I_tegelik)-1)
```

```
plot(aeg, I_tegelik, pch=20,
      xlab="Päevi epideemia algusest",
      ylab="Haigeid internaadis")
```



Paneme järgnevalt kirja oma SIR-mudeli. Vajame funktsiooni mis mistahes vastuvõtlike, haigete ja resistentsete inimeste arvude korral arvutaks nende arvude muutumiskiiruse (tuletised):

```
sir_mudel = function (aeg, seisund, parameetrid)
{
  S = seisund[1]      # susceptibles / vastuvõtlikud
  I = seisund[2]      # infectious / haigestunud
  R = seisund[3]      # recovered / resistentsete

  N=S+I+R

  with ( parameetrid,
        {
          # Arvuta tuletised
          dS = (-beeta * S * I/N)
          dI = ( beeta * S * I/N) - (gamma * I)
          dR = (gamma * I)

          # kombineerime tulemused
          tuletised = c(dS, dI, dR)
          return(list(tuletised))
        }
  )
}
```

Kui teaksime parameetrite väärtuseid, siis saaksime lisamooduli `deSolve` funktsiooni `lsoda` abil leida vastuvõtlike (S), nakatunute (I) ja resistentsete (R, siin kontekstis tähendab internaatkoolist ärasaadetud) õpilaste arvud mistahes ajahetke jaoks:

```
require(deSolve)

parameetrid = list(beeta = 2, gamma = 0.5)
algvaartused=c(S=762, I=1, R=0)
ajad=0:20

tulem0 = lsoda (algvaartused, ajad, sir_mudel, parameetrid)
tulem0
```

Eeldame siin arvutusi tehes, et noored koolipoisid polnud varem gripiga kokku puutunud (või gripi läbipõdenud õpilaste arv on tühiselt väike). Parameetrite väärtused on valitud siia näitese järgmisel viisil: eeldame, et kaks päeva peale haigestumist saadetakse laps internaadist ära ($\text{gamma}=1/2$) ja beeta-parameetri väärtuse oletasime selle põhjal, et järgmiseks päevaks oli üks haige enda kõrvale kaks uut haiget tekitanud, seega kokku oli haigeid $1+2=3$. Lisame neid naiivseid hinnanguid kasutava prognoosi kõvera varem `plot`-käsu abil joonistatud algandmeid kujutavale graafikule:

```
lines(ajad, tulem0[,3])
```

Proovime nüüd mudeli parameetreid beetat ja gammata hinnata kõigi vaatlusandmete pealt. Hindamismeetodina kasutame praegu vähimruutude meetodit – minimiseerime ruutkauguse tegelike ja mudeli arvates esineda võivate haigusjuhtude vahel. Kuna beeta ja gamma väärtused peavad olema mittenegatiivsed siis otsime numbriliselt andmetega kõige paremini sobivaid $\log(\text{beeta})$ ja $\log(\text{gamma})$ väärtuseid. Ükskõik millise hinnangu me $\log(\text{beeta})$ -le ka ei saaks, $\exp(\text{hinnang})$ on igal juhul mittenegatiivne.

```
f=function(par, tegelik){
  parameetrid = list(beeta = exp(par[1]), gamma = exp(par[2]))
  algvaartused=c(S=762, I=1, R=0)
  ajad=0:(length(tegelik)-1)

  tulem0 = lsoda (algvaartused, ajad, sir_mudel, parameetrid)

  return(sum((tulem0[,3]-tegelik)**2))
}

tulem=optim(log(c(2, 0.5)), f, tegelik=I_tegelik)

# Parameetrite hinnangud:
exp(tulem$par)

# R0 hinnanguline väärtus:
exp(tulem$par[1]-tulem$par[2])
```

Reproduktsiooni alusarvu hinnang tuleb mõnevõrra kõrgem kui tavaliselt gripi puhul väidetud – võimalik, et poiste internaatkoolis on gripil tavapärasest mugavam levida.

Saadud hinnangu ja vaatlusandmed võime kanda ka joonisele:

```
ajad=seq(0, 18, length=200)

parameetrid = list(beeta = exp(tulem$par[1]), gamma = exp(tulem$par[2]))
tulemused = lsoda (c(S=762,I=1,R=0), ajad, sir_mudel, parameetrid)

plot(ajad, tulemused[,3], type="l")
points(aeg, I_tegelik, pch=20)
```

Ülesanne 1

Tagantjärgi tarku on ikka jagunud. Epideemia edasist käiku tagantjärgi prognoosida – selleks pole matemaatilist mudelit vaja. Küsimus on pigem selles, kas paari päeva või nädala andmetele tuginedes on võimalik öelda, mis tulevikus võiks juhtuda.

Proovi prognoosida gripiepideemia edasist käiku kasutades 4, 5 ja 7 päeva andmeid (nelja päeva andmed: 1, 3, 8, 28). Kanna kõigi kolme hinnatud mudeli prognoosid haigete arvule graafikule.

Üks neist prognoosidest on imelik. Vaata, mis saadi imeliku prognoosi korral parameetri gamma hinnanguks. Miks võiks parameetri gamma väärtust olla raske antud juhul hinnata?

Mis juhtuks selle prognoosiga, kui kasutaksime varasema arutelu põhjal tehtud oletust, et $\gamma=0.5$ ja hindaksime vaid parameetri beeta väärtuse?

Näide 2. COVID-19 ja Eesti.

SIR-mudeli parameetrite hindamisel praegu käimasoleva haiguspuhangu kirjeldamiseks seisame silmitsi mitmete raskustega. Kõigepealt pole meil teada hetkel haigete (täpsemalt: nakkusohtlike) inimeste arv. Teame vaid uute tuvastatud haigete arvu ja diagnoosi saanute koguarvu, aga haigete ametliku tervekslugemisega on tõsiseid probleeme ja segadusi (kuna me kedagi terveks saanuks loeme – selles on palju segadust. Sageli pole haiged juba pikka aega tegelikult nakkusohtlikud, aga püsivad ikka igaks juhuks haigetena andmebaasis kirjas).

Mudeli hindamiseks võiksime seega kasutada kas ühel päeval diagnoositud uute haigusjuhtude arvu või mingiks päevaks kokku diagnoositud haigusjuhtude koguarvu. Valime praeguse näite jaoks selle viimase variandi (haigusjuhtude koguarvu).

Lisaks tuleb arvestada, et diagnoos pannakse inimesele üldjuhul mitte kohe peale nakatumist/nakkusohtlikuks saamist, vaid mitmeid päevi hiljem. Kui palju hiljem? Vihje selle kohta saab leida kasutades kurikuulsat võrkpallimängu Saaremaal. Kui võrkpallilahing Milano koondise mängijatega toimus 4. märtsil, siis COVID-19 diagnooside hüppeline kasv leidis aset 12. märtsil ehk 8 päeva hiljem (12. märtsil diagnoositi Eestis 40 uut haiget – neist 17 olid saarlased. Varasematel päevadel leidis kogu Eestis kinnitust 2-3 uut haigusjuhtu päevas).

Eeldame, et peale haigusdiagnoosi saamist inimesed enam haigust ei levita – seega nakkuslikkuse perioodi keskmiseks pikkuseks võtame 8 päeva ja seega saame esialgse hinnangu parameetritele $\gamma=1/8$. Kuna kumulatiivsete haigusjuhtude pealt on suhteliselt raske hinnata terveksaamise kiirust (parameetrit γ) siis kasutamegi edaspidi seda taustainformatsiooni pealt oletatud γ näitajat (hindame ainult parameetrit β).

Alates 13. märtsist keelas Vabariigi Valitsus eriolukorra raames avalikud kogunemised, kontserdid jne. Seatud piirangud arvatavasti mõjutasid ka haiguse võimet levida. Sestap üritame hinnata reproduktsiooni alusarvu R_0 eraldi piirangute eelsele ajale ja piirangute järgsele ajale. Kuna aga haiguse avastamine leiab aset umbes 8 päeva peale nakkuse saamist, siis loeme piiriks 21. märtsi – sellel kuupäeval diagnoosi saanud inimesed võisid juba saada nakkuse eriolukorra kehtimise ajal, enne seda kuupäeva diagnoosi saanud inimesed said nakkuse arvatavasti eriolukorrale eelnenud ajal. Alguspunktiks valime aga 1. märtsi.

Haigusjuhud TEHIK-u andmetel.

```
# Algandmed: positiivsete COVID-19 diagnooside koguarv Eestis
# 1. märts - 21. märts
I_kokku = c(1, 2, 2, 3, 10, 10, 10, 10, 13, 15, 18, 58, 102, 125, 201,
            226, 248, 276, 304, 316, 357)

# SIR model
sir_mudel = function (aeg, seisund, parameetrid)
{
  S = seisund[1]      # susceptibles / vastuvõtlikud
  I = seisund[2]      # infectious / haigestunud
  R = seisund[3]      # recovered / resistentsed

  N=S+I+R

  with ( parameetrid,
        {
          gamma = 1/8

          # Arvuta tuletised
          dS = (-beeta * S * I/N)
          dI = (beeta * S * I/N) - (gamma * I)
          dR = (gamma * I)

          # kombineerime tulemused
          results = c(dS, dI, dR)
          return(list(results))
        }
      )
}

# Mudeli hindamiseks peame teadma algseisundit.
# Kui palju aga oli 1. märtsil Eestis neid viirusekandjaid, kes olid saanud
# nakkuse aga kellel veel COVID-19 diagnoosi polnud pandud?
# Võtame nende diagnoosimata nakatanute arvuks järgmise 8 päeva jooksul
# lisandunud uute haigusjuhtude arvu (13-1):

f=function(par, tegelik){
  parameetrid = list(beeta = exp(par[1]))
  algvaartused=c(S=1328000, I=13-1, R=13)
  ajad=0:(length(tegelik)-1)
  tulem0 = lsoda (algvaartused, ajad, sir_mudel, parameetrid)
  SSE=sum((tulem0[,4]-tegelik)**2)
  return(SSE)
}
```

```
tulem=optim(c(log(0.3)), f, tegelik=I_kokku)
# Hinnatud parameetri väärtus. Näitab mitu uut inimest päevas nakatab
# üks SARS-CoV-2 viirusega nakatanud inimene (kuni ta isoleeritakse)
exp(tulem$par)
```

```
# Arvutame ka reproduktsiooni alusarvu
# R0 = beeta/gamma
exp(tulem$par)/(1/8)
```

Reproduktsiooni alusarvu tuleb sarnane kirjanduses varem avaldatuga. Võiks siiski arvata, et Eestis tegelikult on R0 väärtus veidi väiksem kui miljonilinnas Wuhanis – aga meie mudel ei arvesta väljast sissetoodud „ülemääraste“ nakkusallikatega (Milaano võrkpallurid).

Võime ka proovida oletada, kuidas haigestumine oleks jätkunud, kui eriolukorda poleks kehtestatud:

Arvutame mudeli prognoosid:

```
parameetrid = list(beta = exp(tulem$par))
algvaartused=c(S=1328000, I=13-1, R=13)
ajad=seq(0, 25, length=200)
tulem0 = lsoda (algvaartused, ajad, sir_model, parameetrid)
ajad2=seq(0, 20, length=200)
tulem2 = lsoda (algvaartused, ajad2, sir_model, parameetrid)
```

joonistame mudeli prognoosi

```
windows(width=8, height=6)
par(mar=c(6,4,2,2))
plot(ajad, tulem0[,4], col="skyblue", lwd=2, type="l", xlab="",
      ylab="COVID-19 diagnoosiga inimesi", xaxt="n",
      main="Oletus eriolukorrata tulevikust")
lines(ajad2, tulem2[,4],col="blue", lwd=2)
```

Lisame tegelikud vaatlused

```
points(0:20, I_kokku, pch=20)
```

Lisame ajatelje

```
axis(1, at=0:25, format(seq(as.Date("2020/03/1"), by=1, length=26), "%d. %b"),
      las=2)
```

Lisame ka haigestunute koguarvud hilisemaks perioodiks – ajaperioodiks

mida kehtestatud eriolukord juba võiks mõjutada

```
I_kokku_parast = c(370,394,469,524,560,625,664,696,730,788,867,960,
                  1034,1096,1107,1149,1185,1206,1246, 1291,1309)
points(21:(21+length(I_kokku_parast)-1), I_kokku_parast, col="red2", pch=21)
```

Ülesanne 2

Üheks mureks COVID-19 puhul on see, et haigus võib kulgeda ilma sümptomiteta. Seepärast muretsetakse, et kõigist nakatanud inimestest võidakse avastada vaid väike osa. Oletame, et diagnoosi saab (avastatakse) kõigest 10% nakatanud inimestest. Kuidas muutuks R0-i hinnang kui tegelikult oleks haigeid 10 korda rohkem? Muuda nii tuvastatud haigete arvu kui ka 13. märtsil diagnoosimata haigete arvu 10 korda suuremaks. Mis juhtub reproduktsiooni alusarvu hinnanguga?

Ülesanne 3

Kõigile pakub huvi see, kas peale piirangute kehtestamist õnnestus R_0 muuta 1-st väiksemaks. Leia milline on hinnanguline R_0 -i väärtus kasutades andmeid alates 25.märtsist (jätame igaks juhuks paar lisapäeva, sest paar päeva pärast eriolukorda võis olla veel ebatavaliselt palju segadust). Võid kasutada vektoris `I_kokku_parast` olevaid diagnooside arve, soovi korral võid vektorit täiendada lähipäevadel avaldatavate numbritega.

Raporteeri enda saadud hinnang R_0 -le ja lisa ka programm, mida sa arvutusteks kasutasid.

Hoiatus: saadud tulemustega ei tasu kohe ajakirjandusse tormata – meie analüüs näiteks ei arvesta sellega, et diagnoosimise põhimõtted on muutunud. Täna võib peaaegu iga end halvasti tundev inimene pääseda koroonatesti tegema, 25. märtsil oli testimisele pääsemine keerulisem ja võimalik vaid kõige tõenäolisematele COVID-19 haigetele. Tänu sellele võidakse viimasel ajal tuvastada rohkem SARS-CoV-2 viirusega nakatunud inimesi kui varem. Samuti ei arvesta meie analüüs vahepealsete piirangute karmistamisega. Samuti ei arvesta me praegu latentsiperioodiga (nakkuse saanud inimene ei ole päris koheselt teistele nakkusohtlik).