Matematinė statistika (užduotys naudojant R) 2020/ 2021 m.m. pavasario sem.

1 užduotis Atsitiktinių dydžių modeliavimas. Atvirkštinės transformacijos metodo taikymas duomenų modeliavimui. Parametrų ir jų funkcijų taškinių įverčių radimas naudojant didžiausiojo tikėtinumo (DT; maximum likelihood) metodą.

Atsakykite į klausimus: 1) Kaip parinkote pradinį artinį? 2) Koks sprendimo būdas tinkamas, kai parametras yra neneigiamas, bet R optimizavimo funkcijoje nenumatyta galimybė nurodyti galimą parametrų kitimo sritį? 3) Kaip apibrėžiamas skirstinys, kuris priklauso tik nuo postūmio (location) ir mastelio (scale) parametrų? 4) Kaip apibrėžiamas skirstinys, kuris priklauso tik nuo mastelio (scale) ir formos parametrų (shape)?

1. Naudodami atvirkštinės transformacijos metodą, sumodeliuokite atsitiktinio dydžio T, turinčio eksponentinį skirstinį su parametru λ, t. y. a. d. T pasiskirstymo funkcija yra

$$F(t;\lambda) = 1 - e^{-\lambda t}, \quad t \geq 0, \lambda > 0,$$

didumo $n = 100$ paprastąją atsitiktinę imtį. Atlikite užduotis:

a) Naudodami R programos funkcijas maxLik ir optim, didžiausiojo tikėtinumo metodu įvertinkite parametrą λ. Palyginkite su tikra parametro reikšme.

b) Raskite parametrinį išgyvenimo funkcijos ($S(t) = 1 - F(t) = \mathbf{P}(T > t)$) įvertį. Nubraižykite grafiką, kuriame būtų pavaizduotas šis įvertis ir teorinė išgyvenimo funkcija, suformuluokite išvadą.

d) Raskite taškinį medianos įvertį. Palyginkite su tikra reikšme.

2. Naudodami atvirkštinės transformacijos metodą, sumodeliuokite atsitiktinio dydžio T, turinčio Veibulo skirstinį su parametrais η ir ν, t. y. a. d. T pasiskirstymo funkcija yra

$$F(t;\eta,\nu) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\nu}},$$

didumo $n = 100$ paprastąją atsitiktinę imtį. Atlikite užduotis:

a) Įrodykite, jeigu X skirstinys yra Veibulo, t.y. $X \sim W(\eta,\nu)$; $F(x;\eta,\nu) = 1 - e^{-\left(\frac{x}{\eta}\right)^{\nu}}$, tai $Y = \ln X$ skirstinys yra $F(y) = 1 - exp\left\{-e^{(y-\mu)/\sigma}\right\}$; $\sigma = 1/\nu$; $\ln\eta = \mu$.

b) Parinkite pradinį artinį. Pastaba. $X \sim W(\eta,\nu)$; $Y = \ln X$; $\mathbf{E}Y = \mu - \gamma\sigma$, $\gamma = 0.5772156$; $\boldsymbol{V}Y = (\pi\sigma)^2/6$.

c) Naudodami R programos funkciją maxLik, optim arba kitą R optimizavimo funkciją didžiausiojo tikėtinumo metodu įvertinkite parametrus η ir ν. Palyginkite su tikromis parametrų reikšmėmis.

d) Tarkime, kad modeliuoti duomenys žymi gaminio darbo laiką. Įvertinkite tikimybę, kad gaminys dirbs ilgiau negu laiką t. Raskite laiko momentą iki kurio sugenda 30% gaminių.

3. Modeliuokite a. d. X, turinčio binominį skirstinį su parametrais k ir p, t. y.

$$\mathbf{P}(X = m \,|\, k, p) = C_k^m p^m (1-p)^{k-m},$$

dydžio $n = 50$ paprastąją atsitiktinę imtį. Naudodami R programos funkciją maxLik arba optim DT metodu įvertinkite parametrą p. Palyginkite su tikra parametro reikšme.

4. Duomenys (duom1.xlsx):

i	$(a_{i-1}, a_i]$	U_i	i	$(a_{i-1}, a_i]$	U_i
1	$(0, 100]$	8	7	$(600, 700]$	25
2	$(100, 200]$	12	8	$(700, 800]$	18
3	$(200, 300]$	19	9	$(800, 900]$	15
4	$(300, 400]$	23	10	$(900, 1000]$	14
5	$(400, 500]$	29	11	$(1000, \infty)$	18
6	$(500, 600]$	30			

Tarę, kad buvo stebėtas a.d., turintis Veibulo skirstinį, įvertinkite nežinomus parametrus.

5. Duomenys iš 1) ir 2). Įvertinkite parametrus naudodami R funkciją fitdistr(duomenys, "skirstinys").

6. Tegu laikas iki įvykio turi loglogistinį skirstinį, t.y. išgyvenimo funkcija yra $S(t) = \frac{1}{1+\alpha t^{\gamma}}$. Gauti tokie įvykio momentai (duom2.txt):

```
0.151 0.182 0.203 0.204 0.222 0.226 0.235 0.239 0.242 0.242 0.248 0.250 0.275
0.279 0.282 0.298 0.299 0.303 0.309 0.310 0.333 0.339 0.339 0.344 0.349 0.351
0.353 0.360 0.360 0.366 0.376 0.385 0.385 0.392 0.397 0.397 0.400 0.404 0.405
0.408 0.411 0.414 0.415 0.431 0.431 0.440 0.442 0.447 0.451 0.453 0.459 0.459
0.461 0.467 0.469 0.469 0.49  0.491 0.492 0.494 0.510 0.526 0.527 0.542 0.551
```

Įvertinkite parametrus α ir γ: a) naudodami pasirinktą R optimizavimo funkciją; b) funkciją fitdistr.

Raskite medianą, 70-ąjį kvartilį.

Papildoma medžiaga ir pavyzdžiai

1. Atvirkštinės transformacijos metodas

Tarkime, kad X yra tolydusis atsitiktinis dydis su pasiskirstymo funkcija $F(x)$. Tada atsitiktinio dydžio $U = F(X)$ skirstinys yra tolygusis intervale (0, 1). Taigi, jei $U_1, U_2, \ldots, U_n$ yra atsitiktinio dydžio U imtis, tai $X_i = F^{-1}(U_i)$, $i = 1, 2, \ldots, n$ yra atsitiktinio dydžio X imtis.

Pavyzdys (eksponentinis skirstinys). Pasiskirstymo funkcija: $F(x) = 1 - e^{-\lambda x}$. Pažymėkime $U = F(X)$. Tada $X_i = F^{-1}(U_i) = -\frac{1}{\lambda}\ln(1 - U_i)$.

2.1 Skirstiniai, priklausantys tik nuo mastelio ir formos parametrų

Skirstinys priklauso tik nuo mastelio (scale) η ir formos (shape) ν parametrų, jeigu jo pasiskirstymo funkciją galima užrašyti:

$$F(t; \eta, \nu) = F_0\left(\left(\frac{t}{\eta}\right)^{\nu}\right),$$

čia F_0 - žinoma pasiskirstymo funkcija, nepriklausanti nuo nežinomų parametrų.

Pavyzdžiui, Veibulo skirstinys priklauso tik nuo mastelio ir formos parametrų:

$$F(t; \eta, \nu) = F(t; \eta, \nu) = 1 - e^{-\left(\frac{t}{\eta}\right)^{\nu}} = F_0\left(\left(\frac{t}{\eta}\right)^{\nu}\right), \quad F_0(t) = 1 - e^{-t}$$

2.2 Skirstiniai, priklausantys tik nuo postūmio ir mastelio parametrų

Skirstinys priklauso tik nuo postūmio (location) ir mastelio (scale) parametrų, jeigu jo pasiskirstymo funkciją galima užrašyti:

$$F(t; \mu, \sigma) = F_0\left(\frac{t-\mu}{\sigma}\right),$$

čia F_0 - žinoma pasiskirstymo funkcija, nepriklausanti nuo nežinomų parametrų.

Pavyzdžiui, normalusis skirstinys priklauso tik nuo mastelio ir postūmio parametrų:

$$F(t; \mu, \sigma) = \Phi\left(\frac{t-\mu}{\sigma}\right),$$

čia Φ - standartinio normaliojo skirstinio pasiskirstymo funkcija.

3. R pavyzdžiai

DT įverčiai naudojant R paketo stats funkciją nlm. Pastaba. Funkcija nlm randa minimumą; reikia nurodyti: minimizuojamą funkciją, pradines parametrų reikšmes; jeigu nurodome hessian=TRUE apskaičiuojama Hessian matrica, t.y. antrųjų funkcijos išvestinių matrica.

Pavyzdys. Puasono skirstinys.

$$\mathbf{P}(X = k \,|\, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \quad \Rightarrow \quad \ln L = \sum X_i \log(\lambda) + log(X_i!) - \lambda$$

```
duom <- c(3, 0, 2, 1, 0, 4, 3, 2, 1, 2)
ln_l <- function(par) {
 (-1)* sum ( duom *log(par[1])-par[1])
}

nlm(ln_l, theta <- c(1), hessian=TRUE)

rez <- nlm(ln_l, theta <- c(1), hessian=TRUE)
rez$estimate
rez$hessian
```

DT įverčiai naudojant R paketo maxLik funkciją maxLik. Pastaba. Funkcija randa maksimumą; reikia nurodyti: funkciją, pradines parametrų reikšmes.

Pavyzdys. Puasono skirstinys.

```
library(maxLik)
ln_l <- function(par) {
   lambda <- par[1]
   sum ( duom *log(par[1])-par[1])   }

mle <- maxLik(logLik = ln_l, start = c(lambda = 1))

coef(mle)
hessian(mle)
```

Kitos R optimizavimo funkcijos.

Paketo stats funkcijos nlminb, optim, optimx. Reikia nurodyti pradines reikšmes, funkciją, galima nurodyti parametrų reikšmių ribas.

```
nlminb(pradinių_reikšmių_vektorius, funkcija ,
       lower=parametrų_reikšmių_vektorius, upper=parametrų_reikšmių_vektorius)

rez_optim <- optim(pradinių_reikšmių_vektorius, funkcija, method="L-BFGS-B",
lower = c(1.0e-13, 1.0e-13), upper=c(Inf, Inf),
      control=list(fnscale=-1), hessian=TRUE)
rez_optim$par
rez_optim$hessian
```

Pastaba. control=list(fnscale=-1) nurodome, jeigu maksimizavimo uždavinys, pagal nutylėjimą sprendžiamas minimizavimo uždavinį.