Iterativne numerične metode v posplošenih linearnih modelih

Mitja Mandić
Mentor: izr. prof. dr. Jaka Smrekar

5. september 2021

Eksponentna družina

- Porazdelitev pripada enoparametrični eksponentni družini z disperzijskim parametrom, če ima gostoto oblike
$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right),$$
za neke funkcije $a(\cdot), b(\cdot)$ in $c(\cdot)$. Parametru θ pravimo *kanonični* oziroma *naravni* parameter, ϕ pa imenujemo *disperzijski parameter*
- Velja $\mathbb{E}(Y) = b'(\theta)$ in $\mathrm{Var}(Y) = b''(\theta)a(\phi)$

Eksponentna družina

- Porazdelitev pripada enoparametrični eksponentni družini z disperzijskim parametrom, če ima gostoto oblike
$$f_Y(y;\theta,\phi)=\exp\left(\frac{y\theta-b(\theta)}{a(\phi)}+c(y,\phi)\right),$$
za neke funkcije $a(\cdot), b(\cdot)$ in $c(\cdot)$. Parametru θ pravimo *kanonični* oziroma *naravni* parameter, ϕ pa imenujemo *disperzijski parameter*
- Velja $\mathbb{E}(Y)=b'(\theta)$ in $\mathrm{Var}(Y)=b''(\theta)a(\phi)$
- Za $Y\sim N(\mu,\sigma^2)$ z gostoto $\frac{1}{\sqrt{2\pi\sigma^2}}\exp-\frac{(y-\mu)^2}{2\sigma^2}$ izrazimo
$$\theta=\mu,\ \phi=\sigma^2,\ a(\sigma^2)=\sigma^2,\ b(\mu)=\frac{\mu^2}{2}$$

Eksponentna družina

- Porazdelitev pripada enoparametrični eksponentni družini z disperzijskim parametrom, če ima gostoto oblike
$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right),$$
za neke funkcije $a(\cdot), b(\cdot)$ in $c(\cdot)$. Parametru θ pravimo *kanonični* oziroma *naravni* parameter, ϕ pa imenujemo *disperzijski parameter*
- Velja $\mathbb{E}(Y) = b'(\theta)$ in $\mathrm{Var}(Y) = b''(\theta)a(\phi)$
- Za $Y \sim N(\mu, \sigma^2)$ z gostoto $\frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(y-\mu)^2}{2\sigma^2}$ izrazimo
$$\theta = \mu,\ \phi = \sigma^2,\ a(\sigma^2) = \sigma^2,\ b(\mu) = \frac{\mu^2}{2}$$
- Za $Y \sim Bin(m,p)$ je $\mathrm{P}(Y = y) = \binom{m}{y} p^y (1-p)^{m-y}$, kar preoblikujemo v
$$\exp\left(y \log\left(\frac{p}{1-p}\right) + m\log(1-p) + \log\binom{m}{y}\right)$$
in preberemo $\theta = \log\frac{p}{1-p},\ b(\theta) = m\log(1+e^\theta),\ a(\phi) = 1.$

Posplošeni linearni modeli

Vsak posplošeni linearni model sestavljajo trije deli:

- slučajni del – porazdelitev proučevane slučajne spremenljivke

Posplošeni linearni modeli

Vsak posplošeni linearni model sestavljajo trije deli:

- slučajni del – porazdelitev proučevane slučajne spremenljivke
- sistematični del – linearna relacija med pojasnjevalnimi spremenljivkami

Posplošeni linearni modeli

Vsak posplošeni linearni model sestavljajo trije deli:

- slučajni del – porazdelitev proučevane slučajne spremenljivke
- sistematični del – linearna relacija med pojasnjevalnimi spremenljivkami
- povezovalna funkcija – povezava med sistematičnim delom in pričakovano vrednostjo. Če je enaka naravnemu parametru porazdelitve ji rečemo *kanonična* povezovalna funkcja

Posplošeni linearni modeli

Vsak posplošeni linearni model sestavljajo trije deli:

- slučajni del – porazdelitev proučevane slučajne spremenljivke
- sistematični del – linearna relacija med pojasnjevalnimi spremenljivkami
- povezovalna funkcija – povezava med sistematičnim delom in pričakovano vrednostjo. Če je enaka naravnemu parametru porazdelitve ji rečemo *kanonična* povezovalna funkcjia

Skupaj:

$$g(\mu) = X\beta$$

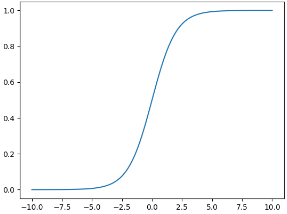
Logistični model

- Za računanje verjetnosti iz binarnih podatkov, predpostavimo model oblike

$$\mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right) = X\beta$$

Logistični model

- Za računanje verjetnosti iz binarnih podatkov, predpostavimo model oblike
$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = X\beta$$
- Izrazimo
$$p_i = \frac{\exp(x_i^\top \beta)}{1 + \exp(x_i^\top \beta)}$$



Logistični model

- Za računanje verjetnosti iz binarnih podatkov, predpostavimo model oblike
$$\operatorname{logit}(p) = \log\left(\frac{p}{1-p}\right) = X\beta$$
- Izrazimo
$$p_i = \frac{\exp(x_i^\top \beta)}{1 + \exp(x_i^\top \beta)}$$
- Izračunamo še prvi in drugi odvod logaritemske funkcije verjetja

Metoda največjega verjetja

- Najprej privzemimo gostote oblike $f_X(x;\theta) = f(x;\theta_1,\ldots,\theta_r)$ za nek $\theta \in \Theta^r$
- Pri fiksni realizaciji poskusa definiramo *funkcijo verjetja*

$$F(X_1,\ldots,X_n;\underbrace{\theta_1,\ldots,\theta_r}_{\theta}) = f(X_1,\theta)\cdots f(X_n,\theta)$$

Metoda največjega verjetja

- Najprej privzemimo gostote oblike $f_X(x;\theta) = f(x;\theta_1,\ldots,\theta_r)$ za nek $\theta \in \Theta^r$
- Pri fiksni realizaciji poskusa definiramo *funkcijo verjetja*

$$F(X_1,\ldots,X_n;\underbrace{\theta_1,\ldots,\theta_r}_{\theta}) = f(X_1,\theta)\cdots f(X_n,\theta)$$

- Iščemo njen maksimum - to bo hkrati tudi maksimum $\log(F)$
- Sistemu

$$\frac{\partial}{\partial\theta_j}\log F(X,\theta) = 0, \;\; j = 1,\ldots,r$$

pravimo sistem enačb verjetja, njegova rešitev je *cenilka največjega verjetja*

Metoda največjega verjetja

- Najprej privzemimo gostote oblike $f_X(x;\theta) = f(x;\theta_1,\ldots,\theta_r)$ za nek $\theta \in \Theta^r$
- Pri fiksni realizaciji poskusa definiramo *funkcijo verjetja*

$$F(X_1,\ldots,X_n;\underbrace{\theta_1,\ldots,\theta_r}_{\theta}) = f(X_1,\theta)\cdots f(X_n,\theta)$$

- Iščemo njen maksimum - to bo hkrati tudi maksimum $\log(F)$
- Sistemu

$$\frac{\partial}{\partial\theta_j}\log F(X,\theta) = 0,\ \ j = 1,\ldots,r$$

  pravimo sistem enačb verjetja, njegova rešitev je *cenilka največjega verjetja*
- Niso nujno nepristranske, so pa dosledne, če je rešitev enolična
- **Običajno niso eksplicitno rešljive**

Numerične metode

- Iterativne metode, ki temeljijo na Newtonovi metodi za iskanje ničel funkcije

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$

Numerične metode

- Iterativne metode, ki temeljijo na Newtonovi metodi za iskanje ničel funkcije
$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)}$$
- Za iskanje ekstremov logaritemske funkcije verjetja $\ell(\theta)$ enačimo njen gradient z nič in dobimo
$$\theta_{i+1} = \theta_i - \frac{\frac{\partial}{\partial\theta}\ell(\theta)}{\frac{\partial^2}{\partial\theta^2}\ell(\theta)},$$
računanje in invertiranje Hessiana pa je lahko problematično. Kako se lahko temu izognemo?

Fisherjeva zbirna metoda

$$\theta_{i+1} = \theta_i + \mathrm{FI}(\theta)^{-1}\left(\frac{\partial}{\partial\theta}\ell\right),$$

kjer je $\mathrm{FI}(\theta) = \mathbb{E}\left((\frac{\partial}{\partial\theta}\ell)(\frac{\partial}{\partial\theta}\ell)^\top\right)$

Fisherjeva zbirna metoda

$$\theta_{i+1} = \theta_i + \mathrm{FI}(\theta)^{-1} \left(\frac{\partial}{\partial \theta} \ell \right),$$

kjer je $\mathrm{FI}(\theta) = \mathbb{E}\left((\frac{\partial}{\partial \theta}\ell)(\frac{\partial}{\partial \theta}\ell)^\top \right)$ Informacijska enakost pravi

$$\mathrm{FI}(\theta) = \mathbb{E}\left((\frac{\partial}{\partial \theta}\ell)(\frac{\partial}{\partial \theta}\ell)^\top \right) = -\mathbb{E}\left(\frac{\partial^2}{\partial \theta^2} \ell \right),$$

Fisherjeva zbirna metoda

$$\theta_{i+1} = \theta_i + \mathrm{FI}(\theta)^{-1}\left(\frac{\partial}{\partial\theta}\ell\right),$$

kjer je $\mathrm{FI}(\theta) = \mathbb{E}\left((\frac{\partial}{\partial\theta}\ell)(\frac{\partial}{\partial\theta}\ell)^\top\right)$ Informacijska enakost pravi

$$\mathrm{FI}(\theta) = \mathbb{E}\left((\frac{\partial}{\partial\theta}\ell)(\frac{\partial}{\partial\theta}\ell)^\top\right) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\ell\right),$$

Za kanonične povezovalne smo dokazali

$$\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\ell\right) = \frac{\partial^2}{\partial\theta^2}\ell$$

in sledi

$$\mathrm{FI}(\theta) = -\frac{\partial^2}{\partial\theta^2}\ell$$

Fisherjeva zbirna metoda

$$\theta_{i+1} = \theta_i + \mathrm{FI}(\theta)^{-1}\left(\frac{\partial}{\partial\theta}\ell\right),$$

kjer je $\mathrm{FI}(\theta) = \mathbb{E}\left((\frac{\partial}{\partial\theta}\ell)(\frac{\partial}{\partial\theta}\ell)^\top\right)$ Informacijska enakost pravi

$$\mathrm{FI}(\theta) = \mathbb{E}\left((\frac{\partial}{\partial\theta}\ell)(\frac{\partial}{\partial\theta}\ell)^\top\right) = -\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\ell\right),$$

Za kanonične povezovalne smo dokazali

$$\mathbb{E}\left(\frac{\partial^2}{\partial\theta^2}\ell\right) = \frac{\partial^2}{\partial\theta^2}\ell$$

in sledi

$$\mathrm{FI}(\theta) = -\frac{\partial^2}{\partial\theta^2}\ell$$

$\rightarrow$ za kanonične povezovalne funkcije Fisherjeva zbirna metoda in Newtonov algoritem sovpadata!

Fisherjeva zbirna metoda

Taylorjev polinom za funkcijo zbira (ob upoštevanju $\ell(\theta^*) = 0$) je

$$\ell(\theta^* + s) = \ell(\theta^*) + \frac{1}{2} s^\top \frac{\partial^2}{\partial \theta^2} \ell(\theta^*) s.$$

Poleg tega pa velja še:

$$-\frac{\partial^2}{\partial \theta^2} \ell = \mathrm{FI}(\theta) = \mathrm{Var}\left(\frac{\partial}{\partial \theta} \ell \right)$$

Fisherjeva zbirna metoda

Taylorjev polinom za funkcijo zbira (ob upoštevanju $\ell(\theta^*) = 0$) je

$$\ell(\theta^* + s) = \ell(\theta^*) + \frac{1}{2} s^\top \frac{\partial^2}{\partial \theta^2} \ell(\theta^*) s.$$

Poleg tega pa velja še:

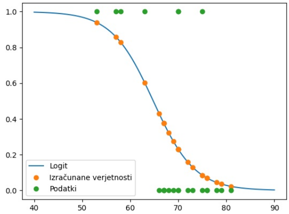
$$-\frac{\partial^2}{\partial \theta^2} \ell = \mathrm{FI}(\theta) = \mathrm{Var}\left(\frac{\partial}{\partial \theta} \ell \right)$$

torej je v θ^* maksimum in imamo konstanten (naraščajoč) alogirtem!

Rezultati

Iz interneta smo pridobili podatke o tesnilih iz vesoljskih poletov pred dobo *Challengerja* in izračunali:

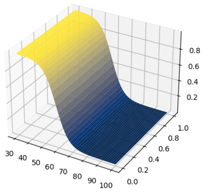
- verjetnost odpovedi z logistično regresijo le s podatki o temperaturi



Rezultati

Iz interneta smo pridobili podatke o tesnilih iz vesoljskih poletov pred dobo *Challengerja* in izračunali:

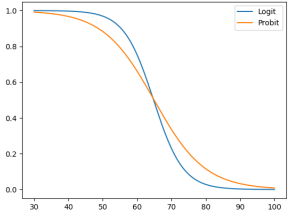
- verjetnost odpovedi z logistično regresijo le s podatki o temperaturi
- verjetnost odpovedi s temperaturo in pritiskom



Rezultati

Iz interneta smo pridobili podatke o tesnilih iz vesoljskih poletov pred dobo *Challengerja* in izračunali:

- verjetnost odpovedi z logistično regresijo le s podatki o temperaturi
- verjetnost odpovedi s temperaturo in pritiskom
- primerjava logistične in probit regresije le s podatki o temperaturi



Literatura