Iterativne numerične metode v posplošenih linearnih modelih

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

9. september 2021

Eksponentna družina

- Porazdelitev pripada enoparametrični eksponentni družini z disperzijskim parametrom, če ima gostoto oblike

$$f_Y(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right),$$

θ imenujemo *naravni* ali *kanonični* parameter, ϕ pa *disperzijski* parameter

Eksponentna družina

- Porazdelitev pripada enoparametrični eksponentni družini z disperzijskim parametrom, če ima gostoto oblike
$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right),$$
θ imenujemo *naravni* ali *kanonični* parameter, ϕ pa *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
$$f_Y(y; \theta, \phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right),$$
θ imenujemo *naravni* ali *kanonični* parameter, ϕ pa *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 $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$ 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

$$f_Y(y;\theta,\phi) = \exp\left(\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right),$$

  θ imenujemo *naravni* ali *kanonični* parameter, ϕ pa *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 $\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right)$ 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 funkcjia

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

Skupaj:

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

Logistični model

- Predpostavimo binomsko porazdelitev Y. Uporablja se za računanje verjetnosti iz binarnih podatkov, model je oblike

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

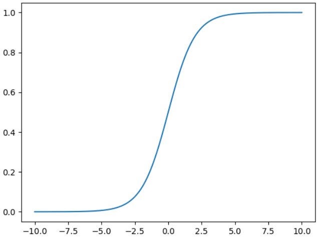
Logistični model

- Predpostavimo binomsko porazdelitev Y. Uporablja se za računanje verjetnosti iz binarnih podatkov, model je oblike

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

- Izrazimo

$$p_i = \frac{\exp(x_i^\top \beta)}{1 + \exp(x_i^\top \beta)}$$



Logistični model

- Predpostavimo binomsko porazdelitev Y. Uporablja se za računanje verjetnosti iz binarnih podatkov, model je oblike

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

- Izrazimo

$$p_i = \frac{\exp(x_i^\top \beta)}{1 + \exp(x_i^\top \beta)}$$

- Logaritemska funkcija verjetja, funkcija zbira in njen odvod

$$\ell(\beta) = \sum_{i=1}^{n} \left(y_i(x_i^\top \beta) - m_i \log(1 + \exp(x_i^\top \beta)) \right)$$

$$\frac{\partial}{\partial \beta_j} \ell(\beta) = \sum_{i=1}^{n} \left(x_{ij}(y_i - \mu_i) \right), \quad j = 0, \ldots, r$$

$$\frac{\partial^2}{\partial \beta_j \partial \beta_k} \ell(\beta) = -\sum_{i=1}^{n} \left(x_{ij} v_i(\beta) x_{ik} \right), \quad j, k = 0, \ldots, r$$

Probit model

$$p = \Phi(X\beta)$$

- Razvit v enake namene kot logistični model, znova se za Y predpostavi binomsko porazdelitev
- Vidimo, da povezovalna funkcija $\Phi^{-1}(p)$ ni kanonična

Probit model

$$p = \Phi(X\beta)$$

- Razvit v enake namene kot logistični model, znova se za Y predpostavi binomsko porazdelitev
- Vidimo, da povezovalna funkcija $\Phi^{-1}(p)$ <u>ni</u> kanonična
- Logaritemska funkcija verjetja, funkcija zbira in njen odvod

$$\ell(\beta) = \sum_{i=1}^{n} \left(y_i \log \Phi(x_i^\top \beta) + (m_i - y_i) \log(1 - \Phi(x_i^\top \beta)) \right)$$

$$\frac{\partial}{\partial \beta_j} \ell(\beta) = \sum_{i=1}^{n} \frac{y - m_i \Phi(x_i^\top \beta)}{\Phi(x_i^\top \beta)(1 - \Phi(x_i^\top \beta))} \phi(x_i^\top \beta) x_{ij}, \quad j = 0, \ldots, r$$

$$\frac{\partial^2}{\partial \beta_j \partial \beta_k} \ell(\beta) = \sum_{i=1}^{n} x_{ij} \frac{(y_i - m_i \Phi(x_i^\top \beta))(-\phi(x_i^\top \beta) x_i^\top \beta) - m_i \phi(x_i^\top \beta)^2}{\Phi(x_i^\top \beta)(1 - \Phi(x_i^\top \beta))} x_{ik} -$$

$$- x_{ij} \frac{\phi(x_i^\top \beta)^2 (1 - 2\Phi(x_i^\top \beta))}{(\Phi(x_i^\top \beta)(1 - \Phi(x_i^\top \beta)))^2} x_{ik}, \quad j, k = 0, \ldots, r$$

Metoda največjega verjetja

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

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

- Iščemo njen maksimum, kar bo hkrati tudi maksimum $\log(F)$

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, kar bo hkrati tudi maksimum $\log(F)$
- Sistemu

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

pravimo sistem enačb verjetja, njegova rešitev je kandidatka za *cenilko 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*

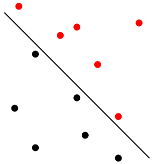
$$F(X_1,\ldots,X_n;\underbrace{\theta_1,\ldots,\theta_r}_{\theta}) = f(X_1,\theta)\cdots f(X_n,\theta)$$

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

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

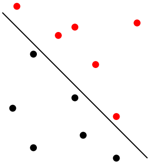
pravimo sistem enačb verjetja, njegova rešitev je kandidatka za *cenilko največjega verjetja*
- Niso nujno nepristranske, so pa dosledne, če je rešitev enolična
- **Običajno niso eksplicitno rešljive**

Obstoj cenilke največjega verjetja v logističnem modelu

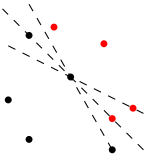


Slika: Popolna separacija. Rešitev ne obstaja

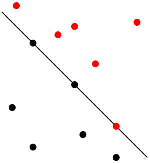
Obstoj cenilke največjega verjetja v logističnem modelu



Slika: Popolna separacija. Rešitev ne obstaja

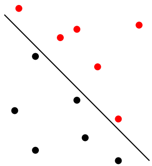


Slika: Nepopolna separacija, drugi primer. Rešitev ne obstaja

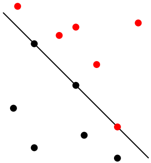


Slika: Nepopolna separacija, prvi primer. Rešitev ne obstaja

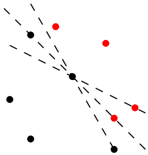
Obstoj cenilke največjega verjetja v logističnem modelu



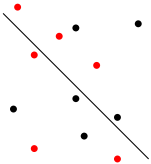
Slika: Popolna separacija. Rešitev ne obstaja



Slika: Nepopolna separacija, prvi primer. Rešitev ne obstaja



Slika: Nepopolna separacija, drugi primer. Rešitev ne obstaja



Slika: Prekrivanje. Rešitev obstaja in je enolična

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 metodo prilagodimo. Enačimo gradient $\ell(\theta)$ z nič in dobimo
$$\theta_{i+1} = \theta_i - \left(\frac{\partial^2}{\partial \theta^2} \ell(\theta_i) \right)^{-1} \frac{\partial}{\partial \theta} \ell(\theta_i).$$

  Računanje in invertiranje Hessiana pa je lahko problematično. Kako se temu izognemo?

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 metodo prilagodimo. Enačimo gradient $\ell(\theta)$ z nič in dobimo
$$\theta_{i+1} = \theta_i - \left(\frac{\partial^2}{\partial \theta^2} \ell(\theta_i) \right)^{-1} \frac{\partial}{\partial \theta} \ell(\theta_i).$$
Računanje in invertiranje Hessiana pa je lahko problematično. Kako se temu izognemo?
- Namesto inverza uvedemo računanje premikov. Iteracijski korak je tako
$$\left(\frac{\partial^2}{\partial \theta^2} \ell(\theta_i) \right) h = -\frac{\partial}{\partial \theta} \ell(\theta_i)$$
$$\theta_{i+1} = \theta_i + h$$

Fisherjeva zbirna metoda

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

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

Fisherjeva zbirna metoda

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

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

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

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

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

V nalogi smo za kanonične povezovalne funkcije 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

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

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

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

V nalogi smo za kanonične povezovalne funkcije 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

Velja še:

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

torej je Hessejeva matrika log-verjetja negativno semidefinitna.

Fisherjeva zbirna metoda

Velja še:

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

torej je Hessejeva matrika log-verjetja negativno semidefinitna.
Taylorjev polinom za funkcijo zbira (ob upoštevanju $\frac{\partial}{\partial\theta}\ell(\theta^*) = 0$) je

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

Fisherjeva zbirna metoda

Velja še:

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

torej je Hessejeva matrika log-verjetja negativno semidefinitna. Taylorjev polinom za funkcijo zbira (ob upoštevanju $\frac{\partial}{\partial\theta}\ell(\theta^*) = 0$) je

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

- $\Rightarrow$ imamo konveksen optimizacijski problem
- $\Rightarrow$ reševanje sistema enačb za premik je enostavnejše

Iteracijski korak logističnega modela:

$$\underbrace{(X^\top v(\hat{\beta}_i)X)}_{\text{Fisher info}}\, h = \underbrace{X^\top (y - \mu(\hat{\beta}_i))}_{\text{funkcija zbira}}$$

$$\hat{\beta}_{i+1} = \hat{\beta}_i + h$$

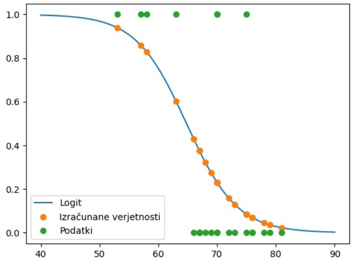
function oceniParametre(iteracije, X, Y, $\beta_{zacetni}$, ϵ)

$p = \frac{\exp(X^\top \beta_{star})}{1+\exp(X^\top \beta_{star})}$
$V = p(1-p)$
$Score = X^\top (Y - p)$
$Info = X^\top V X$
Reši sistem na h: $Info * h = Score$
$\beta_{star} = \beta_{zacetni}$
$\beta_{nov} = \beta_{star} + h$
while $i \leq iteracije$ **do**
  if $\beta_{nov} - \beta_{star} \geq \epsilon$ **then**
    $p = \frac{\exp(X^\top \beta_{nov})}{1+\exp(X^\top \beta_{nov})}$
    $V = p(1-p)$
    $Score = X^\top (Y - p)$
    $Info = X^\top V X$
    Razreši na h: $Info * h = Score$
    $\beta_{star} = \beta_{nov}$
    $\beta_{nov} = \beta_{star} + h$
  else
    Dosegli smo želeno natančnost v zadostnem številu korakov
return β_{nov}
  end if
end while

Rezultati

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

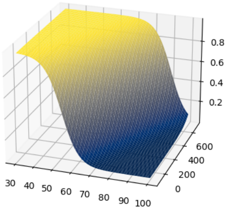
- verjetnost odpovedi z logistično regresijo le s podatki o temperaturi



Rezultati

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

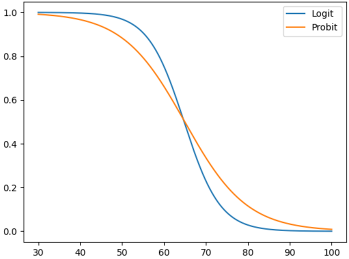
- verjetnost odpovedi z logistično regresijo le s podatki o temperaturi
- verjetnost odpovedi s temperaturo in pritiskom



Rezultati

Iz interneta smo pridobili podatke o odpovedi tesnil 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



Povzetek

- Želeli smo bolje razumeti ozadje računanja parametrov v posplošenih linearnih modelih
- Za logistični model smo izpeljali enačbe verjetja in raziskali kdaj rešitve obstajajo
- Formule smo posplošili za vse modele s porazdelitvami iz eksponentne družine in videli prednosti uporabe kanoničnih povezovalnih funkcij
- Newtonovo metodo smo prilagodili za iskanje ekstremov in dokazali njeno sovpadanje z Fisherjevo zbirno metodo za kanonične modele
- Vso teorijo smo na koncu povezali v praktičnih primerih

Literatura I

A. Agresti.
An introduction to categorical data analysis.
John Wiley & Sons, 2007.

A. Albert and J. A. Anderson.
On the existence of maximum likelihood estimates in logistic regression models.
Biometrika, 71(1):1–10, 1984.

E. B. Erhard.
Logistic regression and Newton-Raphson.
https://statacumen.com/teach/SC1/SC1_11_LogisticRegression.pdf.

K. Lange.
Numerical analysis for statisticians.
Springer Science & Business Media, 2010.

Literatura II

P. McCullagh and J. Nelder.
Generalized linear models.
Springer US, 1989.

B. Plestenjak.
Numerične metode.
https://www.fmf.uni-lj.si/~kozak/PedagoskoDelo/Gradiva/NumericneMetodeI_praktiki/Skripta/BorPlestenjakKnjigaNM.pdf, 2010.

G. Rodríguez.
Lecture notes on Generalized Linear Models.
https://data.princeton.edu/wws509/notes/, 2007.