

Израчунавање оцена максималне веродостојности

6.час

19. април 2016.

У многим ситуацијама приликом оцењивања тешко је наћи аналитички облик за оцену максималне веродостојности. У тим ситуацијама, обично је потребно израчунати оцену максималне веродостојности нумеричким путем.

Постоје многе нумеричке методе за израчунавање ове оцене. Неке од њих су:

- Њутн-Рафсон алгоритам
- Фишеров алгоритам
- ЕМ алгоритам

Њутн-Рафсон алгоритам

Алгоритам опште намене за проналажење решења нелинеарне једначине или решење нелинеарног система једначина. Оцену максималне веродостојности добијамо као решење једне такве једначине.

Претпоставимо да желимо да нађемо решење једначине $g(x_0) = 0$, где је g диференцијабилна једначина.

Како је број x близак броју x_0 Тејлоровим развојем добијамо:

$$0 = g(x_0) \approx g(x) + g'(x)(x_0 - x)$$

$$x_0 \approx x - \frac{g(x)}{g'(x)}.$$

Њутн-Рафсон алгоритам

На овај начин ако нам је дата оцена x_k можемо одредити оцену x_{k+1} по формули

$$x_{k+1} = x_k - \frac{g(x_k)}{g'(x_k)}$$

и ова процедура се понавља за $k = 0, 1, 2, \dots$ док $|g(x_k)|$ или $|g(x_k)/g'(x_k)|$ не буде довољно мало.

Овај алгоритам се може користити за оцену максималне веродостојности или неке друге оцене која се може дефинисати као решење једне једначине или система једначина.

Њутн-Рафсон алгоритам

Претпоставимо да имамо заједничку расподелу $f(\mathbf{x}; \theta)$ за вектор $\mathbf{X} = (X_1, \dots, X_n)$ и нека је $\mathcal{L}(\theta)$ функција максималне веродостојности.

Претпоставимо да је $\hat{\theta}$ оцена максималне веродостојности која задовољава једначину $S(\hat{\theta}) = 0$, где је $S(\theta)$ извод $\ln \mathcal{L}(\theta)$.

Нека је $\hat{\theta}^{(k)}$ оцена за θ након k итерација, тада је

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \frac{S(\hat{\theta}^{(k)})}{H(\hat{\theta}^{(k)})}$$

$$H(\theta) = -\frac{\partial^2}{\partial \theta^2} \ln \mathcal{L}(\theta)$$

Поступак се понавља док $|S(\hat{\theta})|$ или разлика $|\hat{\theta}^{(k)} - \hat{\theta}^{(k+1)}|$ не постане довољно мало.

Њутн-Рафсон алгоритам

Да би користили Њутн-Рафсон алгоритам, потребно је задати почетну (иницијалну) оцелу $\hat{\theta}^{(0)}$.

Ако је $\hat{\theta}^{(0)}$ довољно добра оцелу параметра θ , тада оцелу након једног корака

$$\hat{\theta}^{(1)} = \hat{\theta}^{(0)} + \frac{S(\hat{\theta}^{(0)})}{H(\hat{\theta}^{(0)})}$$

има практично исте особине као MLE .

$$\begin{aligned}\sqrt{n}(\hat{\theta}_n^{(0)} - \theta) &\rightarrow_p 0 \\ \sqrt{n}(\hat{\theta}_n^{(1)} - \hat{\theta}_n) &\rightarrow_p 0\end{aligned}$$

Оцелу $\hat{\theta}^{(0)}$ можемо добити методом момената или неком другом методом.

Њутн-Рафсон алгоритам - пример

Нека су $(X_1, \dots, X_n)$ независне једнако расподељене случајне величине са Кошијевом расподелом чија је функција густине

$$f(x; \theta) = \frac{1}{\pi(1 + (x - \theta)^2)}.$$

Да би илустровали поступак добијања оцене претпоставимо да је $\theta = 10$ и да имамо узорак обима 100. Оцена максималне веродостојности задовољава једначину

$$S(\hat{\theta}) = \sum_{i=1}^n \frac{2(x_i - \hat{\theta})}{1 + (x_i - \hat{\theta})^2} = 0.$$

Приметимо да је густина расподеле симетрична око θ што нам наговештава да узорачка средина или узорачка медијана могу бити оцене θ .

Њутн-Рафсон алгоритам - пример

Даље рачунамо по формули

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \frac{S(\hat{\theta}^{(k)})}{H(\hat{\theta}^{(k)})}$$

k	$\hat{\theta}^{(k)}$	$\ln \mathcal{L}(\hat{\theta}^{(k)})$
0	10.04490	-239.6569
1	10.06934	-239.6433
2	10.06947	-239.6433
3	10.06947	-239.6433

где је

$$H(\theta) = 2 \sum_{i=1}^n \frac{1 - (x_i - \theta)^2}{(1 + (x_i - \theta)^2)^2}.$$

Фишеров алгоритам

Једноставном модификацијом Њутн-Рафсоновог алгоритма добијамо Фишеров алгоритам. У овом алгоритму мењамо H са H где је

$$H_{ij}^*(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}[H_{ij}(\boldsymbol{\theta})] = -E_{\boldsymbol{\theta}} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(\mathbf{X}; \boldsymbol{\theta}) \right]$$

Ако је $\hat{\boldsymbol{\theta}}^{(k)}$ оцена параметра $\boldsymbol{\theta}$ након k итерација, тада важи

$$\hat{\boldsymbol{\theta}}^{(k+1)} = \hat{\boldsymbol{\theta}}^{(k)} + \left[H \left(\hat{\boldsymbol{\theta}}^{(k)} \right) \right]^{-1} S \left(\hat{\boldsymbol{\theta}}^{(k)} \right)$$

Фишеров алгоритам - пример

Нека су $(X_1, \dots, X_n)$ независне једнако расподељене случајне величине са Кошијевом расподелом.

$$H(\theta) = 2 \sum_{i=1}^n \frac{1 - (x_i - \theta)^2}{(1 + (x_i - \theta)^2)^2}$$

$$H^*(\theta) = \frac{n}{\pi} \int_{-\infty}^{\infty} \frac{1 - (x - \theta)^2}{(1 + (x - \theta)^2)^3} dx = \frac{n}{2}.$$

k	$\hat{\theta}^{(k)}$	$\ln \mathcal{L}(\hat{\theta}^{(k)})$
0	10.04490	-239.6569
1	10.06710	-239.6434
2	10.06923	-239.6433
3	10.06945	-239.6433
4	10.06947	-239.6433

$$\hat{\theta}^{(k+1)} = \hat{\theta}^{(k)} + \frac{4}{n} \sum_{i=1}^n \frac{x_i - \hat{\theta}^{(k)}}{1 + (x_i - \hat{\theta}^{(k)})^2}.$$

ЕМ алгоритам

Претпоставимо да имамо заједничку расподелу $f(\mathbf{x}; \theta)$ за вектор $\mathbf{X} = (X_1, \dots, X_n)$ и нека је $\mathbf{Y} = (Y_1, \dots, Y_n)$ случајан вектор такав да је $Y_i = g_i(\mathbf{X})$, где су $g_1, \dots, g_m$ познате функције. Густина вектора $\mathbf{Y}$ је

$$f_Y(\mathbf{y}; \theta) = \int_{A(\mathbf{y})} f_X(\mathbf{x}; \theta) d\mathbf{x}$$

$$A(\mathbf{y}) = \{\mathbf{x} : y_i = g_i(\mathbf{x}) \text{ for } i = 1, \dots, m\}.$$

ЕМ алгоритам прво одреди оцену функције веродостојности, па затим максимизира функцију веродостојности да би добио нове оцене параметра. Ова два корака се понављају док се не постигне конвергенција оцена.

Е корак

Овај корак подразумева проналажење функције веродостојности за θ .
Дата оцена функције максималне веродостојности после k корака дефинисана је

$$\ln \mathcal{L}^{(k)}(\theta) = E[\ln f_X(\mathbf{X}; \theta) | \mathbf{Y} = \mathbf{y}; \hat{\theta}^{(k)}]$$

где је се претпоставља даје права вредност параметра $\hat{\theta}^{(k)}$

M корак

Оцену одређујемо тако што тражимо максимум функције из корака Е. Израчунавање $\hat{\theta}^{(k+1)}$ често захтева коришћење неких нумеричких метода као што је Њутн-Рафсон алгоритам. Након израчунате оцене $\hat{\theta}^{(k+1)}$ враћамо се на корак Е.

Ови кораци се понављају док се не постигне конвергенција.

ЕМ алгоритам

У многим ситуацијама (на пример код експоненцијалне фамилије расподела) $f(\mathbf{x}|\theta)$ је производ два неконстантна фактора, где један од њих не зависи од θ , што се може записати као

$$\ln f_X(\mathbf{x}; \theta) = h_1(\mathbf{x}; \theta) + h_2(\mathbf{x})$$

Функција $\ln f(\mathbf{x}; \theta) = h_1(\mathbf{x}; \theta) + h_2(\mathbf{x})$ у кораку E је

$$\ln \mathcal{L}^{(k)}(\theta) = E[h_1(\mathbf{X}; \theta) | \mathbf{Y} = \mathbf{y}; \hat{\theta}^{(k)}]$$

Ова модификација може значајно поједноставити израчунавања. Корак M остаје исти као раније: бирамо $\hat{\theta}^{(k+1)}$ које максимизира $\ln \mathcal{L}^k(\theta)$

ЕМ алгоритам - пример

Нека су независне једнакорасподељене случајне величине са расподелом која је комбинација две Пуасонове расподеле:

$$f_X(x; \lambda, \theta) = \theta \frac{\exp(-\lambda)\lambda^x}{x!} + (1 - \theta) \frac{\exp(-\mu)\mu^x}{x!}$$

Нека су $(X_1, Y_1), \dots, (X_n, Y_n)$ *iid* парови случајних величина за које важи

$$\begin{aligned} P(Y_i = y) &= \theta^y (1 - \theta)^{1-y} \quad \text{for } y = 0, 1 \\ P(X_i = x | Y = 0) &= \frac{\exp(-\mu)\mu^x}{x!} \quad \text{for } x = 0, 1, 2, \dots \\ P(X_i = x | Y = 1) &= \frac{\exp(-\lambda)\lambda^x}{x!} \quad \text{for } x = 0, 1, 2, \dots \end{aligned}$$

Функција $\ln \mathcal{L}^k(\lambda, \mu, \theta)$ је једнака:

$$\begin{aligned} \ln \mathcal{L}_c(\theta, \lambda, \mu) &= \sum_{i=1}^n y_i [\ln(\theta) + x_i \ln(\lambda) - \lambda] \\ &\quad + \sum_{i=1}^n (1 - y_i) [\ln(1 - \theta) + x_i \ln(\mu) - \mu] - \sum_{i=1}^n \ln(x_i!) \end{aligned}$$

ЕМ алгоритам - пример

Даље треба да рачунамо очекивану вредност логаритма функције веродостојности. Довољно нам је да нађемо очекивану вредност за Y_i ако је $X_i = x_i$ познато за све вредности параметара θ, λ, μ .

$$E(Y_i|X_i = x; \theta, \lambda, \mu) = \frac{\theta \exp(-\lambda)\lambda^x}{\theta \exp(-\lambda)\lambda^x + (1 - \theta) \exp(-\mu)\mu^x}$$

У наставку за дате оцене $\hat{\theta}^{(k)}, \hat{\lambda}^{(k)}$ и $\hat{\mu}^{(k)}$ добијамо оцене $\hat{\theta}^{(k+1)}, \hat{\lambda}^{(k+1)}$ и $\hat{\mu}^{(k+1)}$ тако што максимизирамо

$$\begin{aligned} \ln \mathcal{L}^{(k)}(\theta, \lambda, \mu) &= \sum_{i=1}^n \hat{y}_i^{(k)} [\ln(\theta) + x_i \ln(\lambda) - \lambda] \\ &\quad + \sum_{i=1}^n (1 - \hat{y}_i^{(k)}) [\ln(1 - \theta) + x_i \ln(\mu) - \mu] \end{aligned}$$

ЕМ алгоритам - пример

где је

$$\hat{y}_i^{(k)} = E \left(Y_i | X_i = x; \hat{\theta}^{(k)}, \hat{\lambda}^{(k)}, \hat{\mu}^{(k)} \right)$$

Може се показати да је

$$\begin{aligned}\hat{\theta}^{(k+1)} &= \frac{1}{n} \sum_{i=1}^n \hat{y}_i^{(k)} \\ \hat{\lambda}^{(k+1)} &= \frac{\sum_{i=1}^n x_i \hat{y}_i^{(k)}}{\sum_{i=1}^n \hat{y}_i^{(k)}} \\ \hat{\mu}^{(k+1)} &= \frac{\sum_{i=1}^n x_i (1 - \hat{y}_i^{(k)})}{\sum_{i=1}^n (1 - \hat{y}_i^{(k)})}\end{aligned}$$