Föreläsning 13: Logistisk regression T.K.

TK

05.12.2017

Your Learning Outcomes

- Odds, Odds Ratio, Logit function, Logistic function
- Logistic regression
  - definition
  - likelihood function:
  - maximum likelihood estimate

The setting: *Y is a binary r.v.. x is its covariant, predictor or explanatory variable. x can be continuous, categorical.*

We cannot model the association of Y to x by a direct linear regression,

$$Y = \alpha + \beta x + e$$

where e is, e.g., $\mathcal{N}(0, \sigma^2)$.

KTH Matematik

A Case

Y = *Bacterial Meningitis or Acute Viral Meningitis.*
x =*cerebrospinal fluid total protein count.*

A Case

Diaz, Armando A and Tomba, Emanuele and Lennarson, Reese and Richard, Rex and Bagajewicz, Miguel J and Harrison, Roger G: *Prediction of protein solubility in* Escherichia coli *using logistic regression*, **Biotechnology and bioengineering**, 105, 2 pp. 374–383, 2010.

The authors develop a model for the prediction of the solubility of proteins overexpressed in the bacterium *Escherichia coli*. The model uses the statistical technique of logistic regression. To build this model, 32 covariates x_i that could potentially correlate well with solubility were used.

Logistic regression provides the probability p of a certain protein to belong ($= Y$) to one set or another.

Part I: Concepts

- *Odds, Odds Ratio*
- *Logit function*
- *Logistic function*

Odds

The **odds** of a statement (event e.t.c) A is calculated as the probability $p(A)$ of observing A divided by the probability of not observing A:

$$\text{odds of A} = \frac{p(A)}{1 - p(A)}$$

E.g., in humans an average of 51 boys are born in every 100 births, the odds of a randomly chosen delivery being boy are:

$$\text{odds of a boy} = \frac{0.51}{0.49} = 1.04$$

The odds of a certain thing happening are infinite.

An Aside: Posterior Odds

The posterior probability $p(H \mid E)$ of H given the evidence E:

$$\text{posterior odds of H} = \frac{p(H|E)}{1 - p(H|E)}$$

If there are only two alternative hypotheses, H and not-H, then

$$\text{posterior odds of H} = \frac{p(H|E)}{p(\text{not-}H|E)}$$

$$= \frac{p(E|H)}{p(E \mid \text{not-}H)} \frac{p(H)}{p(\text{not-}H)}.$$

by Bayes formula

$$= \text{likelihood ratio} \times \text{prior odds}$$

Matematik

Odds ratio

The **odds ratio** ψ is the ratio of odds from two different conditions or populations.

$$\psi = \frac{\text{odds of } A_1}{\text{odds of } A_2} = \frac{\frac{p(A_1)}{1-p(A_1)}}{\frac{p(A_2)}{1-p(A_2)}}$$

KTH Matematik

The function logit(p)

The logarithmic odds of success is called the logit of p

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right)$$

The function logit(p) and its inverse

$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right)$$

If $\theta = \text{logit}(p)$, then the inverse function is

$$p = \text{logit}^{-1}(\theta) = \frac{e^{\theta}}{1+e^{\theta}}$$

KTH VETENSKAP OCH KONST
KTH Matematik

The logit(p) and its inverse

$$\mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right), \quad 0 < p < 1$$

$$p = \mathrm{logit}^{-1}(\theta) = \frac{e^{\theta}}{1+e^{\theta}} = \frac{1}{1+e^{-\theta}}$$

The function

$$\sigma(\theta) = \frac{1}{1+e^{-\theta}}, \quad -\infty < \theta < \infty,$$

is called the **logistic function** or **sigmoid**.

Note that $\sigma(0) = \frac{1}{2}$.

Logistic function

In biology the logistic function refers to change in size of a species population.

The logit(p) and the logistic function

Sats

The logit function

$$\theta = \text{logit}(p) = \log\left(\frac{p}{1-p}\right), \quad 0 < p < 1$$

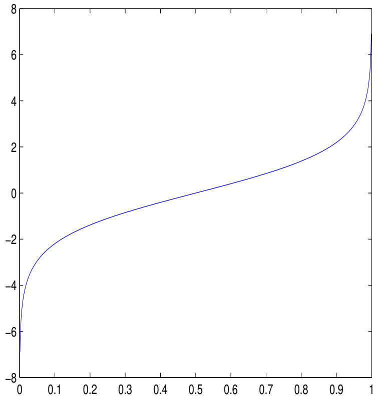
and the logistic function

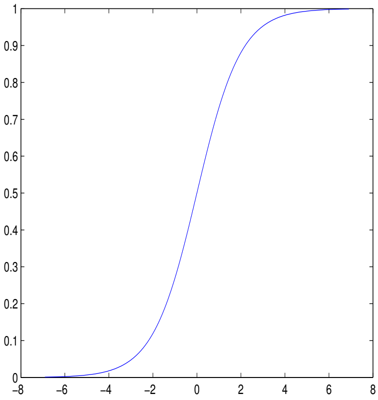
$$p = \sigma(\theta) = \frac{1}{1+e^{-\theta}}, \quad -\infty < \theta < \infty,$$

are inverse functions to each other.

The logit(p) and the logistic function

$$\theta = \mathrm{logit}(p) = \log\left(\frac{p}{1-p}\right), \quad 0 < p < 1 \quad p = \sigma(\theta) = \frac{1}{1+e^{-\theta}}, \quad -\infty < \theta < \infty,$$





The logistic distribution

We see that

- $0 \leq \sigma(x) \leq 1$.
- $\sigma(x) \uparrow 1$, as $x \to +\infty$.
- $\sigma(x) \downarrow 0$, as $x \to -\infty$.
- $\sigma(x)$ is strictly increasing.

Hence $\sigma(x)$ is the distribution function of a random variable.

The logistic distribution

If for every x

$$P(\epsilon \leq x) = \sigma(x) = \frac{1}{1+e^{-x}}, \quad -\infty < x < \infty$$

then we say that ϵ has the **logistic distribution**

$$\epsilon \sim \text{logistic}(0, 1)$$

The logistic distribution: an exercise

Let $P \sim U(0,1)$ and

$$\epsilon = ln\frac{P}{1-P}.$$

Then

$$\epsilon \sim \text{logistic}(0,1)$$

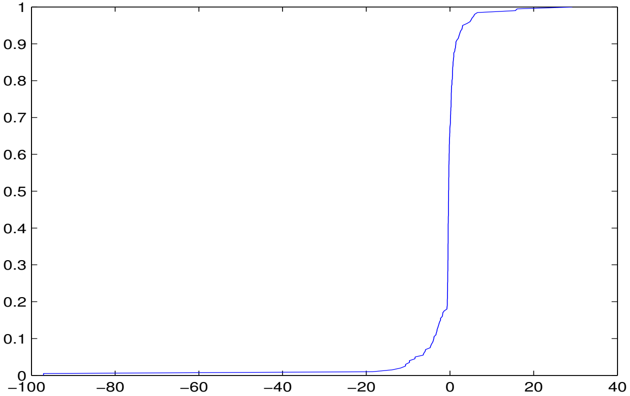
$$\epsilon \sim \text{logistic}(0,1)$$

Then

$$P(-\epsilon \leq x) = P(\epsilon \leq x).$$

Simulating $\epsilon \sim \text{Logistic}(0,1)$

Simulate $p_1, \ldots, p_n$ from the uniform distribution on $(0,1)$ and then do $\epsilon_i = \text{logit}(p_i)$, $i = 1, \ldots, n$. In the figure we plot the empirical distribution function of ϵ_i for $n = 200$.



Part II: Logistic Regression

- *A regression function*
- *log odds of* $Y \leftarrow$ *a regression function*
- *How to generate* Y

KTH Matematik

Let $\beta = (\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$ be $(p+1) \times 1$ vector and $\mathbf{X} = (1, X_1, X_2, \ldots, X_p)$ be a $(p+1) \times 1$ -vector of (predictor) variables. We set, as in multiple regression,

$$\boldsymbol{\beta}^T \mathbf{X} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$$

Then

$$G(\mathbf{X}) = \sigma(\boldsymbol{\beta}^T \mathbf{X}) = \frac{1}{1 + e^{-\beta^T \mathbf{X}}} = \frac{e^{\beta^T \mathbf{X}}}{1 + e^{\beta^T \mathbf{X}}}$$

Logistic regression

The predictor variables $(X_1, X_2, \ldots, X_p)$ can be binary, ordinal, categorical or continuous.

$$G(\mathbf{X}) = \sigma(\boldsymbol{\beta}^T \mathbf{X}) = \frac{e^{\boldsymbol{\beta}^T \mathbf{X}}}{1 + e^{\boldsymbol{\beta}^T \mathbf{X}}}$$

By construction $0 < G(\mathbf{X}) < 1$. Then logit is well defined and

$$\mathrm{logit}(G(\mathbf{X})) = \ln \frac{G(\mathbf{X})}{1 - G(\mathbf{X})} = \boldsymbol{\beta}^T \mathbf{X} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p.$$

Logistic regression

Let now Y be a binary random variable such that

$$Y = \begin{cases} 1 & \text{with probability } G(\mathbf{X}) \\ -1 & \text{with probability } 1 - G(\mathbf{X}) \end{cases}$$

Definition

If the logit of $G(\mathbf{X})$ (or log odds of Y) is

$$\text{logit}(G(\mathbf{X})) = \ln \frac{G(\mathbf{X})}{1 - G(\mathbf{X})} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p,$$

then we say that Y follows a logistic regression w.r.t. the predictor variables $\mathbf{X} = (1, X_1, X_2, \ldots, X_p)$.

Logistic regression

$$Y = \begin{cases} \text{success} & \text{with probability } G(\mathbf{X}) \\ \text{failure} & \text{with probability } 1 - G(\mathbf{X}) \end{cases}$$

$$\mathrm{logit}(G(\mathbf{X})) = \ln \frac{G(\mathbf{X})}{1 - G(\mathbf{X})} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p.$$

is extensively applied in biomedical research.

Logistic regression

$$P(Y = 1 \mid \mathbf{X}) = G(\mathbf{X}) = \sigma(\boldsymbol{\beta}^T \mathbf{X}) = \frac{e^{\beta^T \mathbf{x}}}{1 + e^{\beta^T \mathbf{x}}}$$

$$P(Y = 0 \mid \mathbf{X}) = 1 - G(\mathbf{X}) = 1 - \frac{1}{1 + e^{-\beta^T \mathbf{x}}} = \frac{e^{-\beta^T \mathbf{x}}}{1 + e^{-\beta^T \mathbf{x}}}$$

$$= \frac{1}{1 + e^{\beta^T \mathbf{x}}}.$$

KTH Matematik

Logistic regression: genetic epidemiology

Suppose we have two populations, where $X_i = x_1$ in first population and $X_i = x_2$ in the second population, all other predictors are equal in the two populations. Then a medical geneticist finds it useful to calculate the logarithm of the odds ratio

$$\ln \psi = \ln \frac{p_1}{1-p_1} - \ln \frac{p_2}{1-p_2}$$

$$= \beta_i \left(x_1 - x_2 \right)$$

or

$$\psi = e^{\beta_i (x_1 - x_2)}$$

Hence a unit change in X_i corresponds to e^{β_i} change in odds and β_i change in logodds.

KTH VETENSKAP OCH KONST
KTH Matematik

We need the following regression model

$$Y^* = \boldsymbol{\beta}^T \mathbf{X} + \epsilon$$

where $\boldsymbol{\beta}^T \mathbf{X} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$ and

$$\epsilon \sim \text{Logistic}(0, 1),$$

i.e. the variable Y^* can be written directly in terms of the linear predictor function and an additive random error variable. The logistic distribution is the probability distribution the random error.

Generating model and/or how to simulate

Take a continuous latent variable Y^* (latent= an unobserved random variable) that is given as follows:

$$Y^* = \boldsymbol{\beta}^T \mathbf{X} + \epsilon$$

where $\boldsymbol{\beta}^T \mathbf{X} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p$ and

$$\epsilon \sim \text{Logistic}(0, 1).$$

Define the response Y as the indicator for whether the latent variable is positive:

$$Y = \begin{cases} 1 & \text{if } Y^* > 0 \;\text{ i.e. } \; -\varepsilon < \boldsymbol{\beta}^T \cdot \mathbf{X}, \\ 0 & \text{otherwise.} \end{cases}$$

Then Y follows a logistic regression w.r.t. $\mathbf{X}$. We need only to verify that

$$P(Y = 1 \mid \mathbf{X}) = \frac{1}{1 + e^{-\beta^T \mathbf{X}}}.$$

$$
\begin{aligned}
P(Y = 1 \mid \mathbf{X}) &= P(Y^* > 0 \mid \mathbf{X}) && (1)\\
&= P(\boldsymbol{\beta}^T \mathbf{X} + \varepsilon > 0) && (2)\\
&= P(\varepsilon > -\boldsymbol{\beta}^T \mathbf{X}) && (3)\\
&= P(-\varepsilon < \boldsymbol{\beta}^T \mathbf{X}) && (4)\\
&= P(\varepsilon < \boldsymbol{\beta}^T \ \mathbf{X}) && (5)\\
&= \sigma(\boldsymbol{\beta}^T \mathbf{X}) && (6)\\
&= \frac{1}{1 + e^{-\boldsymbol{\beta}^T \mathbf{X}}} && (7)
\end{aligned}
$$

where we used that the logistic distribution is symmetric (and continuous), as found in the exercise classes, or

$$\Pr(-\varepsilon \le x) = \Pr(\varepsilon \le x).$$

KTH VETENSKAP OCH KONST

KTH Matematik

Part III: Logistic Regression

- *Special case:* $\boldsymbol{\beta}^T \mathbf{X} = \beta_0 + \beta_1 x$.
- *Likelihood*
- *Maximum Likelihood*
- `logisticmle.m`

KTH Matematik

We consider the model:

$$\boldsymbol{\beta}^T \mathbf{X} = \beta_0 + \beta_1 x.$$

$$P(Y = 1 \mid x) = \sigma(\boldsymbol{\beta}^T \mathbf{X}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

Notation:

$$P(Y = y \mid x) = \sigma(\boldsymbol{\beta}^T \mathbf{X})^y (1 - \sigma(\boldsymbol{\beta}^T \mathbf{X})^{1-y}$$

$$= \begin{cases} \sigma(\boldsymbol{\beta}^T \mathbf{X}) & \text{if } y = 1 \\ 1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}) & \text{if } y = 0 \end{cases}$$

Data $(x_i, y_i)_{i=1}^{n}$, likelihood function with the notation above

$$L(\beta_0, \beta_1) = P(Y = y_1 \mid x_1) \cdot P(Y = y_2 \mid x_2) \cdots P(Y = y_n \mid x_n)$$

$$= \sigma(\boldsymbol{\beta}^T \mathbf{X}_1)^{y_1}(1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_1)^{1-y_1} \cdots \sigma(\boldsymbol{\beta}^T \mathbf{X}_n)^{y_n}(1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_n)^{1-y_n}$$

$$= A \cdot B$$

$$A = \sigma(\boldsymbol{\beta}^T \mathbf{X}_1)^{y_1} \cdot \sigma(\boldsymbol{\beta}^T \mathbf{X}_2)^{y_2} \cdots \sigma(\boldsymbol{\beta}^T \mathbf{X}_n)^{y_n}$$

$$B = (1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_1)^{1-y_1}(1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_2)^{1-y_2} \cdots (1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_n)^{1-y_n}$$

$$\ln L\left(\beta_0, \beta_1\right) = \ln A + \ln B$$

$$= \sum_{i=1}^{n} y_l \ln \sigma(\boldsymbol{\beta}^T \mathbf{X}_i)$$

$$+ \sum_{i=1}^{n} (1 - y_i) \ln(1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_i))$$

$$= \sum_{i=1}^{n} \ln(1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_i)) + \sum_{i=1}^{n} y_l \ln \frac{\sigma(\boldsymbol{\beta}^T \mathbf{X}_i)}{1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_i)}$$

$$\ln(1-\sigma(\boldsymbol{\beta}^T \mathbf{X}_i)) = \ln\left(1 - \frac{1}{1+e^{-(\beta_0+\beta_1 x_i)}}\right)$$

$$= \ln\left(\frac{e^{-(\beta_0+\beta_1 x)}}{1+e^{-(\beta_0+\beta_1 x_i)}}\right)$$

$$= \ln\left(\frac{1}{1+e^{\beta_0+\beta_1 x_i}}\right) =$$

$$= -\ln\left(e^{\beta_0+\beta_1 x_i}\right)$$

logodds

$$\ln \frac{\sigma(\boldsymbol{\beta}^T \mathbf{X}_i)}{1 - \sigma(\boldsymbol{\beta}^T \mathbf{X}_i)}$$

$$= \beta_0 + \beta_1 x_i$$

KTH VETENSKAP OCH KONST

KTH Matematik

In summary:

$$\ln L(\beta_0, \beta_1) = \sum_{i=1}^{n} y_i (\beta_0 + \beta_1 x_i) - \sum_{i=1}^{n} \ln \left(e^{\beta_0 + \beta_1 x_i} \right)$$

KTH Matematik

$$\frac{\partial}{\partial \beta_1} \ln L(\beta_0, \beta_1)$$

$$= \sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \frac{\partial}{\partial \beta_1} \ln\left(e^{\beta_0 + \beta_1 x_i}\right)$$

$$\frac{\partial}{\partial \beta_1} \ln\left(e^{\beta_0 + \beta_1 x_i}\right) = \frac{e^{\beta_0 + \beta_1 x_i} x_i}{1 + e^{\beta_0 + \beta_1 x_i}}$$

$$= P(Y = 1 \mid x_i) \cdot x_i.$$

KTH Matematik

$$\frac{\partial}{\partial \beta_1} \ln L(\beta_0, \beta_1) = 0$$

$$\Leftrightarrow$$

$$\sum_{i=1}^{n} \left(y_i x_i - P(Y = 1 \mid x_i) \cdot x_i \right) = 0.$$

In the same manner we can also find

$$\frac{\partial}{\partial \beta_0} \ln L(\beta_0, \beta_1) = \sum_{i=1}^{n} \left(y_i - P(Y = 1 \mid x_i) \right) = 0$$

These two equations have no closed form solution w.r.t. β_0 and β_1.

KTH Matematik

Newton-Raphson for MLE: a one-parameter case

For one parameter θ, set $f(\theta) = \ln L(\theta)$. We are searching for the solution of

$$f^{'}(\theta) = \frac{d}{d\theta} f(\theta) = 0.$$

Newton-Raphson method

$$\theta^{new} = \theta^{old} + \frac{f^{'}\left(\theta^{old}\right)}{f^{''}\left(\theta^{old}\right)},$$

where a good initial value is desired.

Newton-Raphson for logistic MLE

$$\begin{pmatrix} \beta_0^{new} \\ \beta_1^{new} \end{pmatrix} = \begin{pmatrix} \beta_0^{old} \\ \beta_1^{old} \end{pmatrix} + H^{-1}(\beta_0^{old}, \beta_1^{old}) \begin{pmatrix} \frac{\partial}{\partial \beta_0} \ln L\left(\beta_0^{old}, \beta_1^{old}\right) \\ \frac{\partial}{\partial \beta_1} \ln L\left(\beta_0^{old}, \beta_1^{old}\right) \end{pmatrix}.$$

where $H^{-1}(\beta_0^{old}, \beta_1^{old})$ is the matrix inverse of the 2×2 matrix (next slide)

Newton-Raphson for logistic MLE

$$H(\beta_0^{old}, \beta_1^{old}) = \begin{pmatrix} \frac{\partial^2}{\partial \beta_0^2} \ln L\left(\beta_0^{old}, \beta_1^{old}\right) & \frac{\partial^2}{\partial \beta_0 \beta_1} \ln L\left(\beta_0^{old}, \beta_1^{old}\right) \\ \frac{\partial^2}{\partial \beta_1 \beta_0} \ln L\left(\beta_0^{old}, \beta_1^{old}\right) & \frac{\partial^2}{\partial \beta_1^2} \ln L\left(\beta_0^{old}, \beta_1^{old}\right) \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^n P\left(Y=1 \mid x_i\right)\left(1 - P(Y=1 \mid x_i\right) \text{ right})\text{cdot}x_i & \sum_{i=1}^n P\left(Y=1 \mid x_i\right)\left(1 - P(Y=1 \mid x_i\right) \\ \sum_{i=1}^n P\left(Y=1 \mid x_i\right)\left(1 - P\left(Y=1 \mid x_i\right)\right) \cdot x_i & \sum_{i=1}^n P\left(Y=1 \mid x_i\right)\left(1 - P(Y=1 \mid x_i)\right) x_i^2 \end{pmatrix}$$

[bs,stderr,phat,deviance] = logisticmle(y,x)

Input:

y - responses, a binary vector, values 0 and 1

x - the covariate, as a vector

Output:

bs - estimators of beta0 and beta1

stderr - standard error of the estimate = square roots of the diagonal elements of H^{-1}.

phat - estimator of p= P(Y=1)

deviance - deviance

Confidence Interval with degree of confidence $1 - \alpha$

$$\widehat{\beta}_{MLE,i} \pm \lambda_{\alpha/2} \cdot stderr(\widehat{\beta}_{MLE,i})$$

KTH Matematik