TMA4315 Generalized linear models H2018

Module 5: Generalized linear models - common core

Mette Langaas, Department of Mathematical Sciences, NTNU - with contributions from Ingeborg Hem

11.10.2017 [PL], 12.10.2017 [IL]

(Latest changes: 07.10.2018 first version)

Overview

Learning material

- Textbook: Fahrmeir et al (2013): Chapter 5.4, 5.8.2.
- Classnotes 27.09.2018

Topics

- random component: exponential family
  - elements: θ, ϕ, w, $b(\theta)$
  - elements for normal, binomial, Poisson and gamma
  - properties: $\mathrm{E}(Y) = b'(\theta)$ and $\mathrm{Var}(Y) = b''(\theta)\frac{\phi}{w}$ (and proof)
- systematic component= linear predictor
  - requirements: full rank of design matrix
- link function and response function
  - link examples for normal, binomial, Poisson and gamma
  - requirements: one-to-one and twice differentiable
  - canonical link

- likelihood inference set-up: $\theta_i \leftrightarrow \mu_i \leftrightarrow \eta_i \leftrightarrow \beta$
- the loglikelihood
- the score function
- expected Fisher information matrix for the GLM and covariance for $\hat{\beta}$
  - what about covariance of $\hat{\beta}$ when ϕ needs to be estimated?
  - estimator for dispersion parameter
- Fisher scoring and iterated reweighted least squares (IRWLS)
- Pearson and deviance statistic
- AIC

– so, for the first time: no practical examples or data sets to be analysed!

Jump to interactive.

GLM — three ingredients

Random component - exponential family

In Module 1 we introduced distributions of the Y_i, that could be written in the form of a *univariate exponential family*

$$f(y_i \mid \theta_i) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{\phi} \cdot w_i + c(y_i, \phi, w_i)\right)$$

where we said that

- θ_i is called the canonical parameter and is a parameter of interest
- ϕ is called a nuisance parameter (and is not of interest to us=therefore a nuisance (plage))
- w_i is a weight function, in most cases $w_i = 1$ (NB: can not contain any unknown parameters)
- b and c are known functions.

Elements - for normal, Bernoulli, Poisson and gamma

We have seen:

Distribution	θ	$b(\theta)$	ϕ	w	$\mathrm{E}(Y) = b'(\theta)$	$b''(\theta)$
normal	μ	$\frac{1}{2}\theta^2$	σ^2	1	$\mu = \theta$	1
Bernoulli	$\ln\left(\frac{p}{1-p}\right)$	$\ln(1+\exp(\theta))$	1	1	$p = \frac{\exp(\theta)}{1+\exp(\theta)}$	$p(1 -$
Poisson	$\ln\mu$	$\exp(\theta)$	1	1	$\lambda = \exp(\theta)$	λ
gamma	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\frac{1}{\nu}$	1	$\mu = -1/\theta$	μ^2

Properties

$\mathrm{E}(Y_i) = b'(\theta_i)$ and $\mathrm{Var}(Y_i) = b''(\theta_i)\frac{\phi}{w_i}$

In class we study the handwritten proof together: Proof

$b''(\theta_i)$ is often called the variance function $v(\mu_i)$.

Systematic component - linear predictor

Nothing new - as always in this course: $\eta_i = \mathbf{x}_i^T \beta$, and we require that the $n \times p$ design matrix $\mathbf{X} = (\mathbf{x}_1^T, \mathbf{x}_2^T, \ldots, \mathbf{x}_n^T)$ has full rank (which is p).

Remark: in this course we always assume that $n >> p$.

Link function - and response function

Link function

$$\eta_i = g(\mu_i)$$

Response function

$$\mu_i = h(\eta_i)$$

Examples for normal, binomial, Poisson and gamma

random component	response function and link function
normal	$h(\eta_i) = \eta_i$ and $g(\mu_i) = \mu_i$, "identity link".
binomial	$h(\eta_i) = \frac{e^{\eta_i}}{1+e^{\eta_i}}$ and $g(\mu_i) = \ln\left(\frac{\mu_i}{1-\mu_i}\right) = \text{logit}(p_i)$. NB
Poisson	$h(\eta_i) = \exp(\eta_i)$ and $g(\mu_i) = \ln(\mu_i)$, log-link.
gamma	$h(\eta_i) = -\frac{1}{\eta_i}$ and $g(\mu_i) = -\frac{1}{\mu_i}$, negative inverse, or h

Requirements

- one-to-one (inverse exists)
- twice differential (for score function and expected Fisher information matrix)

Canonical link

$$\eta_i = \theta_i$$

so

$$g(\mu_i) = \theta_i$$

When the canonical link is used some of the results for the GLM (to be studied in the next sections) are simplified.

Likelihood inference set-up

$\theta_i \leftrightarrow \mu_i \leftrightarrow \eta_i \leftrightarrow \beta$

A more informative drawing made in class.

See class notes or Fahrmeir et al (2015), Section 5.8.2 for the derivation of the loglikelihood, score and expected Fisher information matrix.

Loglikelihood

$$l(\beta) = \sum_{i=1}^{n} l_i(\beta) = \sum_{i=1}^{n} \frac{1}{\phi}(y_i\theta_i - b(\theta_i))w_i + \sum_{i=1}^{n} c(y_i, \phi, w_i)$$

Remark: the part of the loglikelihood involving both the data and the parameter of interest is for a *canonical link* equal to

$$\sum_{i=1}^{n} y_i\theta_i = \sum_{i=1}^{n} y_i\mathbf{x}_i^T\beta = \sum_{i=1}^{n} y_i \sum_{j=1}^{p} x_{ij}\beta_j = \sum_{j=1}^{p} \beta_j \sum_{i=1}^{n} y_i x_{ij}$$

Score function

$$s(\beta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)\mathbf{x}_i h'(\eta_i)}{\mathrm{Var}(Y_i)} = \mathbf{X}^T \mathbf{D} \Sigma^{-1} (\mathbf{y} - \mu)$$

where $\Sigma = \mathrm{diag}(\mathrm{Var}(Y_i))$ and $\mathbf{D} = \mathrm{diag}(h'(\eta_i))$ (derivative wrt η_i).

Remark: observe that $s(\beta) = 0$ only depends on the distribution of Y_i through μ_i and $\mathrm{Var}(Y_i)$.

Canonical link:

$$s(\beta) = \sum_{i=1}^{n} \frac{(y_i - \mu_i)\mathbf{x}_i w_i}{\phi}$$

since $\frac{\partial \mu_i}{\partial \eta_i} = b''(\theta_i)$.

CHECK last result!

Expected Fisher information matrix for the GLM and covariance for $\hat{\beta}$

$$F_{[h,l]}(\beta) = \sum_{i=1}^{n} \frac{x_{ih}x_{il}(h'(\eta_i))^2}{\mathrm{Var}(Y_i)}$$

$$F(\beta) = \mathbf{X}^T\mathbf{W}\mathbf{X}$$

where $\mathbf{W} = \mathrm{diag}(\frac{h'(\eta_i)^2}{\mathrm{Var}(Y_i)})$.

Canonical link:

$$\frac{\partial^2 l_i}{\partial \beta_j \partial \beta_l} = -\frac{x_{ij} w_i}{\phi} (\frac{\partial \mu_i}{\partial \beta_l})$$

which do not contain any random variables, so the observed must be equal to the expected Fisher information matrix.

CHECK last result!

Fisher scoring and iterated reweighted least squares (IRWLS)

$$\beta^{(t+1)} = \beta^{(t)} + F(\beta^{(t)})^{-1} s(\beta^{(t)})$$

Insert formulas for expected Fisher information and score function.

$$\beta^{(t+1)} = (\mathbf{X}^T \mathbf{W}(\beta^{(t)})\mathbf{X})^{-1}\mathbf{X}^T\mathbf{W}(\beta^{(t)})\tilde{\mathbf{y}}_i^{(t)}$$

where $\mathbf{W}$ is as before $\mathbf{W} = \text{diag}(\frac{h'(\eta_i)^2}{\text{Var}(Y_i)})$ - but now the current version of $\beta^{(t)}$ is used. The diagonal elements are called the *working weights*. The $\tilde{\mathbf{y}}_i^{(t)}$ is called the *working response vector* and has element i given as

$$\tilde{\mathbf{y}}_i^{(t)} = \mathbf{x}_i^T \beta^{(t)} + \frac{y_i - h(\mathbf{x}_i^T \beta^{(t)})}{h'(\mathbf{x}_i^T \beta^{(t)})}.$$

Remark: Convergens? With full rank of $\mathbf{X}$ and positive diagonal elements of $\mathbf{W}$ we are certain that the inverse will exist, but there might be that the temporary version of $\mathbf{W}$ can cause problems.

See what is output from glm- observe working weights as weights..

```
fitgrouped = glm(cbind(y, n - y) ~ ldose, family = "binomia
names(fitgrouped)
fitgrouped$weights
fitgrouped$residuals
```

```
##  [1] "coefficients"      "residuals"         "fitted.val
##  [4] "effects"           "R"                 "rank"
##  [7] "qr"                "family"            "linear.pre
## [10] "deviance"          "aic"               "null.devia
## [13] "iter"              "weights"           "prior.weig
## [16] "df.residual"       "df.null"           "y"
## [19] "converged"         "boundary"          "model"
## [22] "call"              "formula"           "terms"
## [25] "data"              "offset"            "control"
## [28] "method"            "contrasts"         "xlevels"
##          1         2         3         4         5
##  3.254867  8.227383 14.321313 13.378893 10.261055  5.156
```

Estimator for dispersion parameter

FORMULA

Distribution of the MLE

As before we have that maximum likelihood estimator $\hat{\beta}$ asymptotically follows the multivariate normal distribution with mean β and covariance matrix equal to the inverse of the expected Fisher information matrix. This is also true when we replace the unknown β with the estimated $\hat{\beta}$ for the expected Fisher information matrix.

$$\hat{\beta} \approx N_p(\beta, F^{-1}(\hat{\beta}))$$

and with

$$F(\hat{\beta}) = \mathbf{X}^T\hat{\mathbf{W}}\mathbf{X}$$

where $\hat{\mathbf{W}}$ denotes that $\hat{\beta}$ is used then calculating $\mathbf{W} = \text{diag}(\frac{h'(\eta_i)^2}{\text{Var}(Y_i)})$.

What about the distribution of $\hat{\beta}, \hat{\phi}$?

The concept of orthogonal parameters.

Hypothesis testing

Same as before - for the Wald we insert the formula for the covariance matrix of $\hat{\beta}$, for the LRT we insert the loglikelihoods and for the score test we insert formulas for the score function and expected Fisher information matrix.

Model assessment and model choice

Pearson and deviance statistic

Group observations together in groups of maximal size (covariate patterns? interval versions thereof?). Group i has n_i observations, and there are G groups. Asymptotic distribution correct if all groups have big n_i. For individual data asymptotic results can not be trusted.

Deviance

$$D = -2[\sum_{i=1}^{g}(l_i(\hat{\mu}_i) - l_i(\bar{y}_i))]$$

with approximate χ^2-distribution with $G - p$ degrees of freedom.

Pearson:

$$X_P^2 = \sum_{i=1}^{G} \frac{(y_i - \hat{\mu}_i)^2}{v(\hat{\mu}_i)/w_i}$$

with approximate $\phi \cdot \chi^2$-distribution with $G - p$ degrees of freedom.

Remember that the variance function $v(\hat{\mu}_i) = b''(\theta_i)$ (this is a function of μ_i because $\mu_i = b'(\theta_i)$).

AIC

Let p be the number of regression parameters in our model.

$$\text{AIC} = -2 \cdot l(\hat{\beta}) + 2p$$

If the dispersion parameter is estimated use $(p + 1)$ in place of p.

Interactive session

Problem 1: Exam 2011, problem 3

a) Define the class of generalized linear models (GLMs), and explicitly list all requirements for each part of the model.

b) Below are three likelihoods, three link functions, and three linear components listed. Explain which *combinations* that give valid GLMs (8 in total), and also comment on these models (you do not have to mathematically prove which are valid).

Likelihoods:

1. Gaussian, $Y \sim N(\mu, \sigma^2)$
2. Binomial, $Y \sim Bin(n, \pi)$, where n is not fixed (hence is unknown and be estimated)
3. Poisson, $Y \sim Poisson(\lambda)$

Link functions:

1. $\eta = \cos(\mu)$

Problem 2: December 2005, Problem 2 (modified)

1. Derive the formula for the (scaled) deviance for the binomial distribution.

2. The covariance matix for the estimated coefficents are given as $\text{Cov}(\hat{\beta}) = (\mathbf{X}^T\mathbf{W}\mathbf{X})^{-1}$ where $\mathbf{X}$ is the design matrix.

1. (New) The matrix $\mathbf{W}$ is a diagonal matrix. What is on the diagonal?
2. Calulate the elements of $\mathbf{W}$ for a Poisson regression- both with log and identity link. Compare.
3. Calulate the elements of $\mathbf{W}$ for a binary regression - both with logit and identity link. Compare.
4. (New) Which insight did this give you into the role of the link function and its effect on the covariance for the parameter estimates?

Problem 3: Exam 2006, problem 2 (a, b, d)

Let $Y_1, Y_2, \ldots, Y_N$ be independent and exponentially distributed random variables, where Y_i has the density

$$f(y_i; \alpha_i) = \alpha_i e^{-\alpha_i y_i} \text{ for } y_i > 0, \alpha_i > 0, i = 1, 2, \ldots, N.$$

a) Show that the distribution of Y_i comes from the exponential family. Use the general formulas to find $\mathrm{E}(Y_i)$ and $\mathrm{Var}(Y_i)$ as functions of α_i.

b) Show that the log-likelihood for the data $y_1, \ldots, y_n$ can be written as

$$l = \sum_{i=1}^{N} \{-\alpha_i y_i + \ln \alpha_i\}$$

Use this to show that the deviance for a generalized linear model with estimated expectations $\hat{\mu}_i = \hat{y}_i$ is

Problem 4: Exam 2007, problem 2 a, b, c

Assume that $Y_1, \ldots, Y_N$ are independent continuous distributed random variables, where the density of Y_i is given by

$$f(y_i; \gamma_i) = \begin{cases} \frac{\gamma_i^2}{2} y_i e^{-\gamma_i y_i} & \text{for } y_i \geq 0 \\ 0 & \text{else} \end{cases}$$

where γ_i is a scalar parameter.

a) Show that the distribution of Y_i comes from the exponential family. Hint: usually we choose to let $\phi = \frac{1}{2}$.

Use the general formulas to show that $\mathrm{E}(Y_i) = 2/\gamma_i$ and $\mathrm{Var}(Y_i) = 2/\gamma_i^2$.

Assume a GLM for $Y_1, \ldots, Y_N$ where the distribution of Y_i is as above for all i, with the following link function:

$$\eta = g(\mu) = \ln(\mu) = x^T \beta$$

Problem 5: Exam UiO December 2017, Problem 2

We assume that the random variable Λ is gamma distributed with pdf

$$f(\lambda; \nu, \mu) = \frac{(\nu/\mu)^\nu}{\Gamma(\nu)} \lambda^{\nu-1} e^{-\lambda/\mu};\ \ \lambda > 0$$

and further that given $\Lambda = \lambda$, the random variable Y is Poisson distributed with parameter λ. Thus the conditional pmf of Y given $\Lambda = \lambda$ takes the form

$$\mathrm{P}(Y = y|\lambda) = \frac{\lambda^y}{y!} \exp(-\lambda),\ \ y = 0, 1, 2, \ldots.$$

a) Show that the marginal pmf of Y is given by

$$p(y; \mu, \nu) = \frac{\Gamma(y+\nu)}{\Gamma(\nu)\Gamma(y+1)} \left(\frac{\mu}{\mu+\nu}\right)^y \left(\frac{\nu}{\mu+\nu}\right)^\nu;\ \ y = 0, 1, 2, \ldots$$

This is the negative binomial distribution.

Exam questions

December 2015

One of the important concepts we have discussed in this course, is deviance (for Gaussian regression, Poisson regression and logistic regression).

1. Explain what deviance is, and how it relates to residual sum of squares (RSS) for Gaussian regression. Remark 2017/2018: we have called this “sums of squares of errors - SSE”
2. Discuss how it relates to a likelihood ratio test (LRT) for comparing two nested regression models.
3. Discuss how deviance can be used to construct “ANOVA” tables for Poisson regression and logistic regression. Remark 2017/2018: these are called analysis of deviance tables.
4. Discuss how deviance can be used to define residuals, for Poisson regression and logistic regression.

Further reading

- A. Agresti (2015): "Foundations of Linear and Generalized Linear Models." Wiley.