GLM

Generalized Linear Models

Methods Seminar - Kai

19.04.2023

GEORG-AUGUST-UNIVERSITÄT GÖTTINGEN IN PUBLICA COMMODA SEIT 1737

Introduction/ Overview

- What is a Generalized Linear Model (GLM)?
  - What is an Ordinary Linear Model (LM)
  - How do we relax a LM to a GLM?
- When do we use GLMs?
  - Overview
  - 3 Examples

This presentation is based on the textbook *Regression - Models, Methods and Applications by Fahrmeir, Kneib, Lang and Marx (2013), ISBN 978-3-642-34332-2* incl. direct citations. The textbook can be downloaded as PDF from the SUB. The presentation also contains materials from the lecture *GLM by Prof. Thomas Kneib*.

The Ordinary Linear Model (LM) *Ch. 3 in Fahrmeir et al.*

LM: Properties

$$y = \beta_0 + \beta_1 * x_1 + \cdots + \beta_k * x_k + \varepsilon$$

- Modelling the relationship between y and $x1, \ldots x_k$
- Parameters $\beta_1 \ldots \beta_k$ are unknown and need to be estimated
- The relationship is not exact, as it is affected by random noise ε

Following fundamental assumptions

1. The functional relationship between y and $x_1 \ldots x_k$ is **linear**
2. The error ε is **additive**

Ordinary Least Squares (OLS) estimation method

The model can be rewritten as a **linear combination**

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon$$

We seek for the vector of parameters β that **minimizes the errors** in vector ε. $\mathbf{y}$ is the vector with observations of the dependent variable, $\mathbf{X}$ the matrix with observations of the covariates (design matrix).
If we add *further assumptions*, the vector β is straightforwardly solvable/ estiamable.

3. The errors are in mean (expectation) 0 (**unbiasedness**)
4. constant error variance σ^2 across observations (**homoscedastic and uncorrelated errors**) $Cov(\varepsilon) = E(\varepsilon\varepsilon') = \sigma^2\mathbf{I}$
5. ε $\mathbf{x}$ are (stochastically) **independent**

By the way: Gaussian distribution of the errors is not an assumption - it will (only) become relevant for confidence interval calculation and statistical testing.

Ordinary Least Squares (OLS) estimation method

- The unknown regression coefficients β are estimated by minimizing the sum of the squared deviations. The squared deviations are:

$$LS(\beta) = \sum_{i=1}^{n} \varepsilon_i = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \sum_{i=1}^{n} (y_i - \mathbf{x}_i' * \beta)^2$$

It can be rearranged to (page 105 in Fahrmeir et al.):

$$\begin{aligned} &= (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \\ &= \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\beta + \beta'\mathbf{X}'\mathbf{X}\beta \end{aligned}$$

Ordinary Least Squares (OLS) estimation method

which can be derived to:

$$\frac{\partial LS(\beta)}{\partial \beta} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\beta$$

We therefore obtain the least squares estimator $\hat{\beta}$ by setting the eq. to zero or equivalently by solving the so-called normal equations:

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{y}$$
$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

The principle of Maximum Likelihood

The principle of Maximum Likelihood

- Alternative philosophy of estimating functional relationships
- Seeking the set of parameters that most probably describe the observations = maximum plausibility
- How likely is it that the observations are described by a functional relationship?
- We need a function that describes the likelihood density of the regression function
- The estimate becomes also a random quantity
- This likelihood function is a density distribution function, ML therefore does not work without distributional assumptions!
- The likelihood function is maximized by calculating the 1st and 2nd derivative and setting the first derivative to zero

Estimating LM with ML

Estimating LM with ML

- Assuming normally distributed errors, the observations can be described by the density: $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2\mathbf{I})$. We obtain the likelihood function:

$$L(\beta, \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)\right)$$

- and the log likelihood function

$$l(\beta, \sigma^2) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(2\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$$

When maximizing the log-likelihood with respect to β, we can ignore the first two terms because they are independent of β. Maximizing the rest is equivalent to minimizing $(\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta)$, which is the least squares criterion.

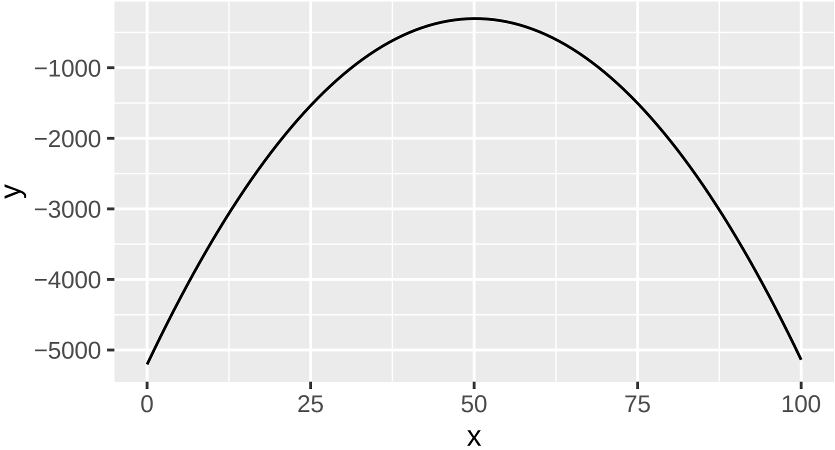
Numerical example that OLS is eqivalent to ML

On the example on an arithmetic mean

```
dat <- rnorm(100, mean = 50,  sd = 5)
std <- sd(dat)

## Likelihood Function ##
L <-  function (x) {sum((log((1 / (sqrt(2 * pi * std^2))) *
                        (exp( - ((dat - x)^2) / (2 * std^2) )))))}
```

Numerical example that OLS is eqivalent to ML



Numerical example that OLS is eqivalent to ML

```r
optimize(f = L,
  interval = c(0, 100), maximum = TRUE
)
```

```
## $maximum
## [1] 50.16894
##
## $objective
## [1] -303.6817
```

```r
mean(dat)
```

```
## [1] 50.16894
```

Generalized Linear Models *Ch. 5 in Fahrmeir et al.*

From LM to GLM

Why GLM?

- Linear models are well suited for regression analyses when the response variable is continuous and at least symmetrically distributed
- In some cases, an appropriate transformation is sufficient to ensure the assumptions (e.g. take logarithm or root of the response variable)
- However, in many applications the response is not a continuous variable, but e.g. rather binary, categorical, or a count variable
- Moreover, considerably skewed data as typical e.g. in a life span, the amount of damages, or income is not estimable with LM
- Typical problem in e.g. survey analysis

From LM to GLM

Within a broad framework, generalized linear models (GLMs) unify many regression approaches with response variables that do not necessarily follow a normal distribution.

- The assumption of linearity between the covariates and the response is relaxed
- The relationship must still be described by a linear predictor, but this linear predictor can be related with a monotonic transformation function, the **link function**
- We thus can relax the fundamental assumptions 1 and 2!

How is the Generalized Linear Model now connected with the ML philosophy?

- We can take any **distribution function** from the so called Exponential Family to define the likelihood function
- The likelihood function will be concave
- The solution will be unique
- A closed form solution often exists

*Note that GLM (Generali**sed** LM) is not to be confused with a General LM*

Example 1: Count Data Regression

- Count data often have a non-symmetric (non-gaussian thus anyway) shape
- For example dbh frequencies in a forest stand
- dbh frequencies may depend on the trees' ages, stand density, ...
- An LM would, however, lead to unsatisfactory estimates
- Count data regression is a common strategy to overcome this issue
- link function is the log, also the ordinary linear function without a link function is common
- Distributional assumption is the Poisson distribution

Example 1: Count Data Regression - Estimation

$$L(\beta) = \frac{\lambda_i^{y_i} exp(-\lambda_i)}{y_i!}$$

$$l(\beta) = \sum_{i=1}^{n} (y_i log(\lambda_i) - \lambda_i)$$

with $log(\lambda_i) = \eta_i = \mathbf{x}_i\beta$
Can be solved as (p. 296 in Fahrmeir et al.)

$$\sum_{i=1}^{n} \mathbf{x}_i(y_i - \lambda_i)$$

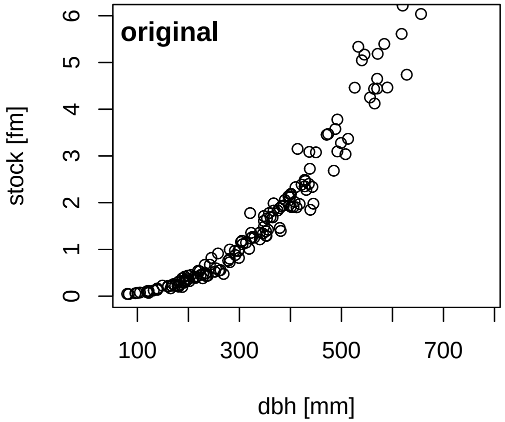
$\lambda_i = exp(\mathbf{x}_i\beta)$ leads to a nonlinear equation system. The solution must be calculated numerically.

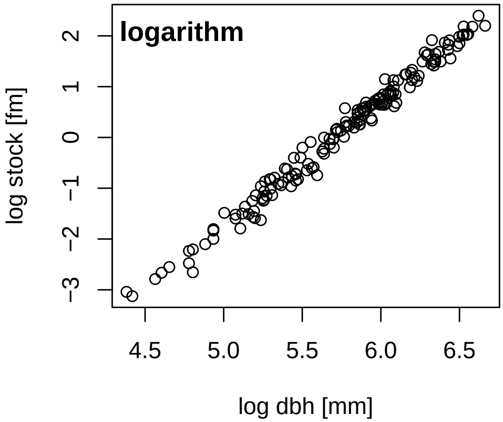
Example 2: Log-linear Regression

- Consider an exponential functional relationship $y = x^b * \varepsilon$
- Distribution of the response is gaussian
- The link function is the inverse of the functional relationship, thus

$$\eta_i = log(\mathbf{x_i}\beta)$$

Data Example

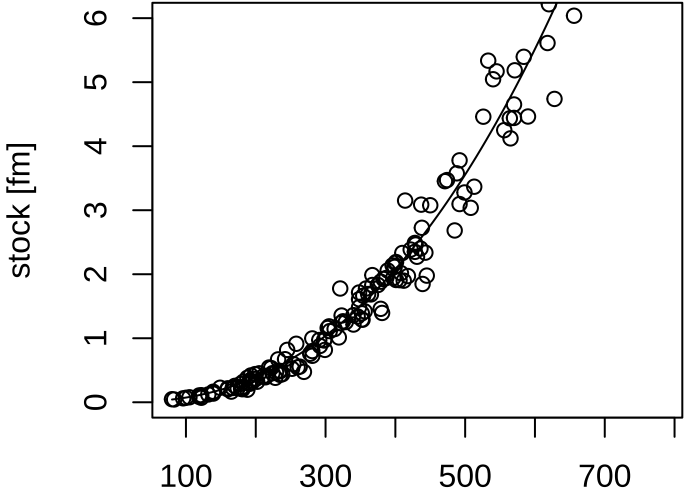




Data Example

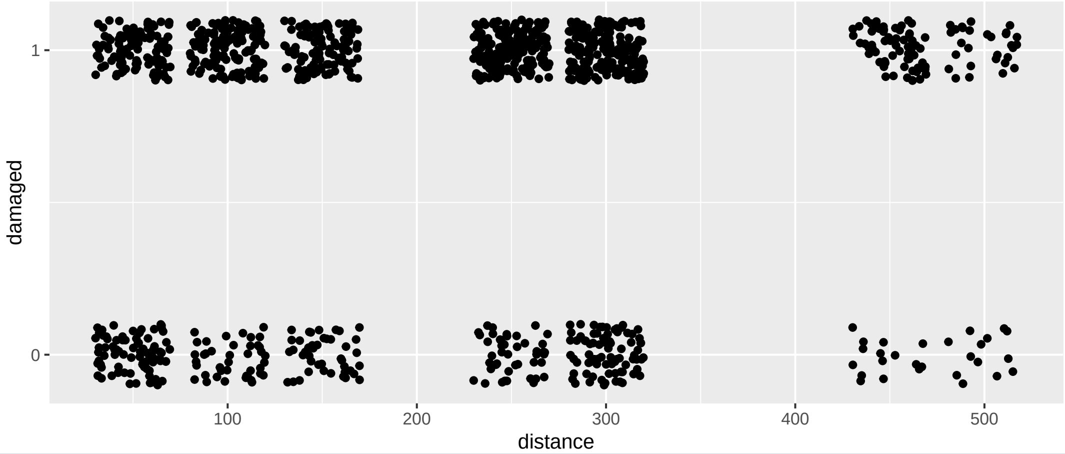
```
mod.allo <- glm(stock_mR ~ log(dbh_mm),
                data = allo,
                family = gaussian(link = "log"))
```

Data Example

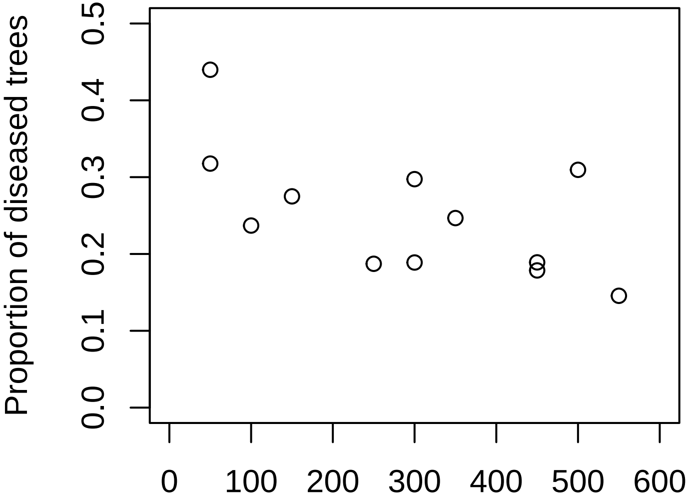


Example 2: Binary Regression (Logit Model)

- Consider a binary response (e.g. yes/ no, dead/ alive, ...)
- An ordinary LM is obviously not suitable



Example 2: Binary Regression (Logit Model)

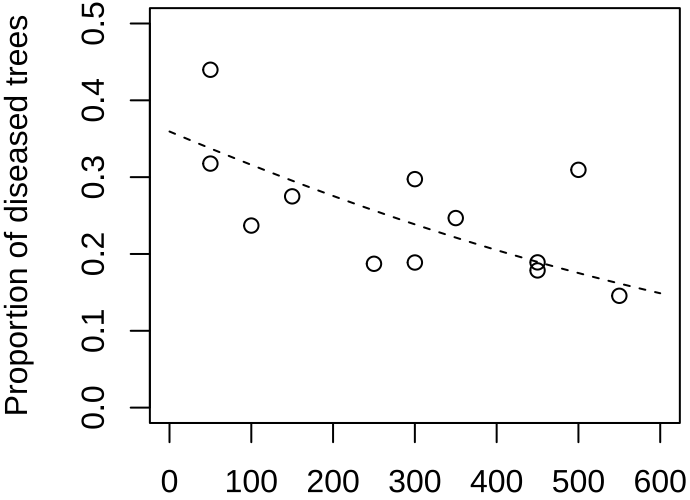


Example 3: Binary Regression (Logit Model)

- The logit function meets our requirements: It is between 0 and 1 and and s-shaped (sigmoidal)
- Thus: link $=$ logit $log(\frac{\eta}{1-\eta})$
- Distribution is usually the binomial distribution (but others are possible)

```
model.logit  <- glm(cbind(diseased, trees - diseased) ~ distance,
                    family = binomial(link = "logit"),
                    data = forest.damage.dist)
```

Example 3: Binary Regression (Logit Model)



Example 3: Binary Regression (Logit Model)

Model interpretation is straightforward

```
##
## Call:
## glm(formula = cbind(diseased, trees - diseased) ~ distance, family = binomial(link = "lo
##     data = forest.damage.dist)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.5196  -0.5696  -0.1531   1.3874   2.2457
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.5778737  0.1040218  -5.555 2.77e-08 ***
## distance    -0.0019428  0.0003488  -5.570 2.55e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
```

Concluding remarks

Concluding remarks

- A model is generalized linear as long as it
  - contains a linear predictor,
  - the link function is a monotonous transformation, and
  - the distribution of the error is described by the Exponential Family.
- GLM is an alternative to nasty non-linear regressions.
- It is lean - It requires low demands on calculation times and calculation resources.
- It opens all likelihood based inference methods (AIC, Likelihood Ratio, ...)
- Visualization, model interpretation, model comparison is straightforward.
- It also opens the other advantages of linear models. It can be connected with General Regression. Thus our assumptions 4 and 5 can also be relaxed.