

Generalized Linear Models

In this lecture, we will motivate the need for generalized linear models (GLMs) by considering limitations of the linear model. This will involve details of the exponential family of distributions, before considering the definition of GLM. Examples involving logistic and Poisson regression will be considered.

This presentation follows essentially the developments in Fahrmeir, L. and Tutz, G., *Multivariate Statistical Modelling based on Generalized Linear Models*, Springer, 2001.

This is also available via the library.

Context: one-level regression

Suppose that we have one or more predictors $\mathbf{x}^T = (x_1, \dots, x_p)$ and one univariate response variable y . These may be continuous or categorical.

Given data (\mathbf{x}_i, y_i) , $i = 1, \dots, n$, our aim is to find a model between predictors and responses.

As we have seen from lectures 1 and 2, the linear model has that

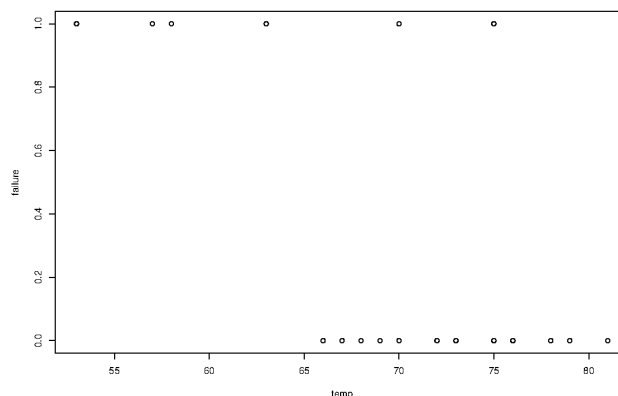
$$y_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i$$

or equivalently,

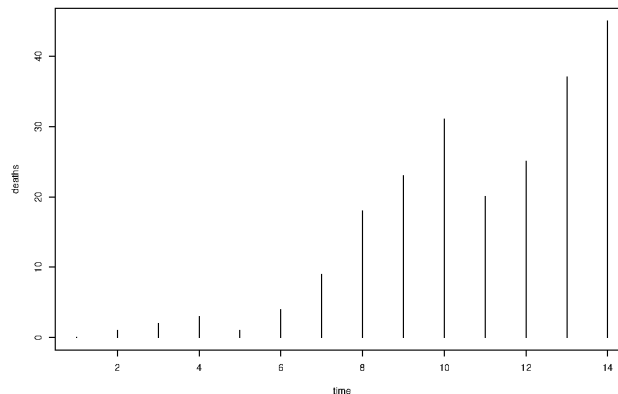
$$E(y_i | \mathbf{x}_i) = \mathbf{x}_i^T \boldsymbol{\beta}.$$

Is the linear model capable to deal with all “one-level” problems? The answer is of course “no”, as we will see with the following examples.

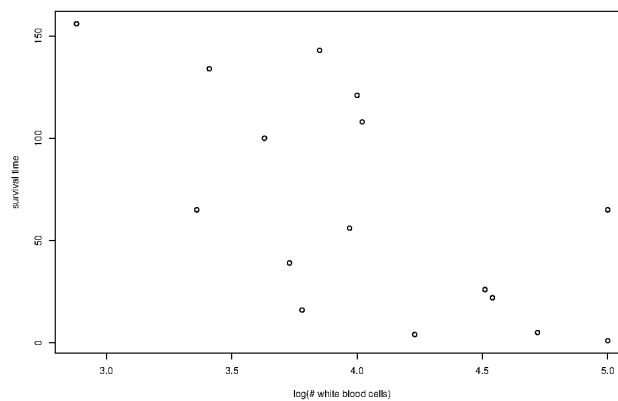
Example 1: Temperature ($^{\circ}F$) at time of take-off and occurrence/non-occurrence (1=yes, 0=no) of an “O-ring”-failure at the 23 U.S. space shuttle flights prior to the Challenger disaster of January 20, 1986.



Example 2: Cases of death due to AIDS in Australia in 14 successive quarters from Jan 1983 to Jun 1986.



Example 3: For 17 leukaemia patients, the survival time after diagnosis is provided in dependence on of the base 10 logarithm of the white blood cell count (at time of diagnosis).



Limitations of the linear model

For all three examples, the range of the response variable is restricted e.g. $\{0, 1\}$ in example 1, the natural numbers in example 2 and the positive reals in example 3. Consequently, we cannot use a normal distribution to model these responses.

Exponential family of distributions

We wish to be able to model responses from a wide range of distributions. Almost all commonly used distributions are members of the *exponential family*.

The probability density function (pdf, in the continuous case), or probability mass function (pmf, in the discrete case), of exponential family random variables takes the form

$$f(y|\theta, \phi) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} + c(y, \phi) \right\}$$

where

θ is the natural parameter of the family,
 ϕ is a scale or dispersion parameter and
 $b(\cdot)$ and $c(\cdot)$ are specific functions corresponding to the type of the family.

Example: Normal distribution

For a normal (also known as Gaussian) distribution, we have the pdf

$$\begin{aligned}
 f(y) &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2} \left(\frac{y - \mu}{\sigma} \right)^2 \right\} \\
 &= \exp \left\{ \frac{y\mu - \mu^2/2}{\sigma^2} - \frac{y^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \right\}.
 \end{aligned}$$

Therefore,

- $\theta = \mu$,
- $b(\theta) = \theta^2/2 = \mu^2/2$,
- $\phi = \sigma^2$.

Example: Bernoulli distribution

For a Bernoulli distribution, we have the probability mass function (pmf) as

$$\begin{aligned}
 f(y) &= \pi^y (1 - \pi)^{1-y} \\
 &= \exp \left\{ y \log \left(\frac{\pi}{1 - \pi} \right) + \log(1 - \pi) \right\}
 \end{aligned}$$

where $\pi = P(y = 1)$ is the probability for "success". This is an exponential family with

- $\theta = \log(\pi/(1 - \pi))$
- $b(\theta) = \log(1 + \exp(\theta)) = -\log(1 - \pi)$,
- $\phi = 1$.

Other exponential family distributions include (but are not limited to) Binomial, Poisson, Geometric, Negative Binomial, Gamma and Exponential distributions.

Page 21 of Fahrmeir and Tutz gives the table on the next page.

Table 2.1. Simple exponential families with dispersion parameter

$f(y \theta, \phi, \omega) = \exp \left\{ \frac{y\theta - b(\theta)}{\phi} \omega + c(y, \phi, \omega) \right\}$				
(a) Components of the exponential family				
Distribution		$\theta(\mu)$	$b(\theta)$	ϕ
Normal	$N(\mu, \sigma^2)$	μ	$\theta^2/2$	σ^2
Bernoulli	$B(1, \pi)$	$\log(\pi/(1-\pi))$	$\log(1 + \exp(\theta))$	1
Poisson	$P(\lambda)$	$\log \lambda$	$\exp(\theta)$	1
Gamma	$G(\mu, \nu)$	$-1/\mu$	$-\log(-\theta)$	ν^{-1}
Inverse Gaussian	$IG(\mu, \sigma^2)$	$1/\mu^2$	$-(-2\theta)^{1/2}$	σ^2
(b) Expectation and variance				
Distribution	$E(y) = b'(\theta)$	var. fct. $b''(\theta)$	$\text{var}(y) = b''(\theta)\phi/\omega$	
Normal	$\mu = \theta$	1	σ^2/ω	
Bernoulli	$\pi = \frac{\exp(\theta)}{1+\exp(\theta)}$	$\pi(1-\pi)$	$\pi(1-\pi)/\omega$	
Poisson	$\lambda = \exp(\theta)$	λ	λ/ω	
Gamma	$\mu = -1/\theta$	μ^2	$\mu^2 \nu^{-1}/\omega$	
Inverse Gaussian	$\mu = (-2\theta)^{-1/2}$	μ^3	$\mu^3 \sigma^2/\omega$	

Derivatives are denoted by $b'(\theta) = \partial b(\theta)/\partial \theta$, $b''(\theta) = \partial^2 b(\theta)/\partial \theta^2$. The weight ω is equal to 1 for individual ungrouped observations. For grouped data, y denotes the average of individual responses, the densities are scaled, and the weight ω equals the group size (i.e., the number of repeated observations in a group).

Generalized linear models (GLMs)

Let $\mu_i = E(y_i|x_i)$. A *generalized linear model* is determined by

- the type of the exponential family which specifies the distribution of $y_i|x_i$,

- the form of the linear predictor $\mathbf{x}_i^T \boldsymbol{\beta}$, i.e. the selection and coding of covariates,
- the transformation

$$\mu_i = h(\mathbf{x}_i^T \boldsymbol{\beta}), \quad \text{equivalently} \quad g(\mu_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (0.1)$$

where

h is a known one-to-one *response function*,
 g is the *link function*, i.e. the inverse of h .

The natural link (function) Several choices of the link function can be considered, but the most common choice is the *natural* or *canonical* link

$$g(\cdot) = \theta(\cdot)$$

which is uniquely determined by the respective exponential family, and can be directly read from the exponential family density. For instance, for Poisson responses one has $g(\cdot) = \log(\cdot)$.

For Gaussian responses, $\theta(\cdot)$ is the identity function and so the natural link is the identity link. In other words, the usual Gaussian linear model is a special case of the generalized linear model!

Logistic regression

For data (\mathbf{x}_i, y_i) , with a $\{0, 1\}$ -valued response, we are interested in modelling the probability of “success”, $\pi_i = P(y_i = 1 | \mathbf{x}_i)$.

One has

$$\pi_i = P(Y_i = 1 | \mathbf{x}_i) = E(y_i | \mathbf{x}_i) = \mu_i$$

so that the binary regression model can be generally formulated as

$$\pi_i = h(\mathbf{x}_i^T \boldsymbol{\beta}).$$

Now, recall that for the Bernoulli distribution, the pmf is

$$f(y) = \exp \left\{ y \log \left(\frac{\pi}{1 - \pi} \right) + \log(1 - \pi) \right\}$$

Hence, the natural link for the Bernoulli distribution is the so-called *logit link*

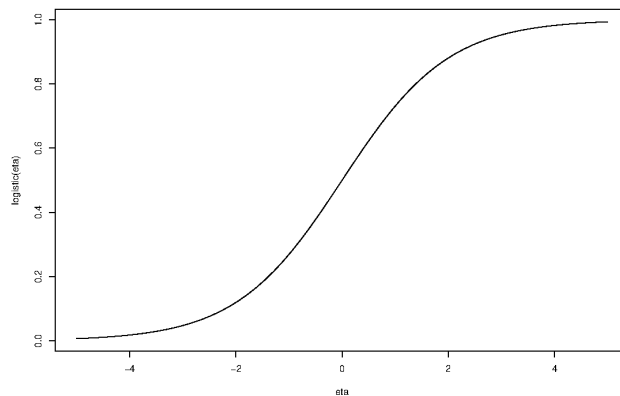
$$g(\pi) = \log \left(\frac{\pi}{1 - \pi} \right),$$

and the corresponding response function is

$$h(\cdot) = g^{-1}(\cdot) = \frac{\exp(\cdot)}{1 + \exp(\cdot)}.$$

Binary regression using the logit link is known as *logistic regression*.

The logistic response function $h(\mathbf{x}_i^T \boldsymbol{\beta}) = \frac{\exp(\mathbf{x}_i^T \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta})}$ is shown below.



The figure shows clearly how the linear predictor range $\mathbf{x}_i^T \boldsymbol{\beta} \in (-\infty, \infty)$ is mapped to $(0, 1)$.

Example 1 revisited: Here $\mathbf{x}_i^T \boldsymbol{\beta} = \beta_0 + \beta_1 \text{temp}_i$, $y_i = 1$ if failure and 0 otherwise.

```
shuttle <- read.table("https://andyclightly.github.io/teaching/MATH43515/
shuttle.asc", header=TRUE)
```

```
head(shuttle)
```

```
##   flight temp td
## 1      1   66  0
## 2      2   70  1
## 3      3   69  0
## 4      4   68  0
## 5      5   67  0
## 6      6   72  0
```

```
glm1<- glm(td~temp, family=binomial(link=logit), data = shuttle)
glm1

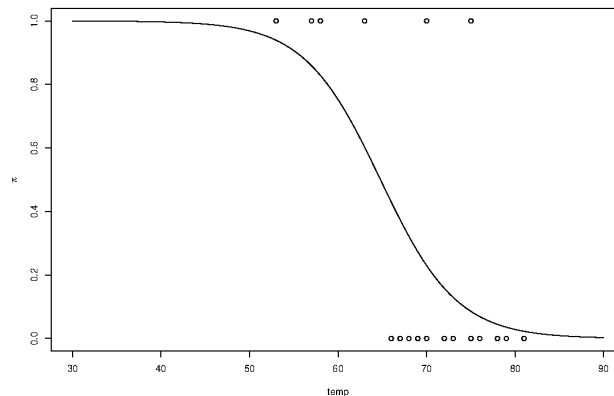
##
## Call:  glm(formula = td ~ temp, family = binomial(link = logit), data = shuttle)
##
## Coefficients:
## (Intercept)      temp
##    15.0429    -0.2322
##
## Degrees of Freedom: 22 Total (i.e. Null);  21 Residual
## Null Deviance:      28.27
## Residual Deviance: 20.32  AIC: 24.32
```

We see that $\mathbf{x}_i^T \hat{\boldsymbol{\beta}} = 15.0429 - 0.2322 \text{temp}_i$.

Hence, the probability of failure in dependence of the temperature is

$$\hat{\pi} = \frac{\exp(15.0429 - 0.2322 \text{temp})}{1 + \exp(15.0429 - 0.2322 \text{temp})}.$$

On the day of the accident, the temperature was 31F.



```
predict(glm1, newdata= data.frame(temp=31),
type="response")
##          1
## 0.9996088
```

The model fit would have told us that

$$\hat{P}(y = 1 | x = 31) = \hat{\pi} = 0.9996088 \quad (!!!)$$

The complimentary lecture slides also include an example on Poisson regression.

Summary

GLMs are an extremely powerful device to fit models to a wide range of response types. In practice, GLMs are not more difficult to fit than LMs — just use the `glm` function in R. However, a bit of knowledge is needed to understand what the fitted models actually mean, and how the parameters need to be interpreted.