Technische Universität München

Department of Mathematics

Bachelor's Thesis

Generalized linear models with parametric link families in R

Nicolai Flach

Supervisor: Prof. Claudia Czado, Ph.D.

Advisor: Matthias Killiches

Submission Date: July 18, 2014

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

Garching, July 18, 2014

_____

Nicolai Flach

Zusammenfassung

Diese Bachelorarbeit behandelt das Thema "Generalisierte lineare Modelle mit parametrischen Linkfunktionen in R". Nach einer kurzen Einleitung widmen wir uns den mathematischen Grundlagen der linearen Regression, um die Theorie der generalisierten linearen Modelle herzuleiten. Vieles der zugrundliegenden Theorie lässt sich auf die generalisierten Modelle verallgemeinern. Während die linearen Modelle bei der Verteilung für den Fehler auf die Normalverteilung beschränkt sind, kann in generalisierten linearen Modellen dafür jede Verteilung der Exponentialfamilie verwendet werden. Im dritten Kapitel wird neben der Beschreibung des Modells ein Überblick über die Schätzung mittels der Maximum-Likelihood-Methode gegeben. Hierfür wird der sogenannte Fisher scoring Algorithmus hergeleitet, welcher sich in den IWLS-Algorithmus ("iterative weighted least squares") umschreiben lässt. Dieser fußt auf der Theorie der Schätzmethode der kleinsten Quadrate, welche wir im Kapitel über lineare Modelle erklären. Ein zentraler Punkt dieser Arbeit ist die Definition eigener parametrischer Linkfunktionen und deren Implementierung in R. Daher wird in Unterkapitel 3.3 darauf eingegangen, wie die Standard-Linkfunktionen in R implementiert sind. Außerdem wird ein Überblick gegeben, welche Wahl für die Linkfunktion für die jeweilige Verteilungsfamilie zulässig ist. Zuletzt wird das theoretische Fundament gelegt um anhand einer Kennzahl, der Devianz, einzuschätzen, wie gut das Modell zu den gegebenen Daten passt.

Nach dem theoretischen Teil werden im vierten Kapitel die Daten vorgestellt, welche in dieser Arbeit immerzu Verwendung in den Beispielen finden. Im fünften Kapitel wird die R interne `glm`-Funktion auf die eben genannten Daten angewendet. Hauptaugenmerk gilt der Devianz, welche eine Maßzahl für den sogenannten "goodness of fit" ist. Sie misst wie stark die Erwartungswerte des Modells von den gegebenen Daten abweichen. Je kleiner die Abweichung ist, desto besser passt das Modell zu den Daten. Daher gilt es dasjenige Modell auszuwählen, welches eine minimale Devianz liefert. In den Beispielen der gewöhnlichen `glm`-Funktion sind die resultierenden Devianzen zum Teil nicht zufriedenstellend. Im sechsten Kapitel führen wir deshalb Transformationen für die Linkfunktionen ein. Dies ist das Kernthema der Arbeit und wurde bereits von Czado (2007) thematisiert. Mithilfe der parametrischen Linkfunktionen lässt sich in allen Fällen ein Modell erzielen, welches besser auf die Daten passt, da die resultierende Devianz geringer ist.

Jedoch kennt man den optimalen Parameter nicht, der für die Minimierung der Devianz in der parametrischen Linkfunktion verwendet werden sollte. In Unterkapitel 6.3 wird die `glmProfile`-Funktion vorgestellt, die dieses Problem lösen soll. Nach Eingabe eines Parameter-Vektors wird ein Plot der Devianz gegen den Parameter und die entsprechenden Werte ausgegeben. Außerdem wird der Parameter angegeben, der die Devianz minimiert mit entsprechender minimaler Devianz. Die `glmProfile`-Funktion wird im siebten Kapitel auf die in dieser Arbeit verwendeten Datensätze angewandt. Im achten Kapitel vergleichen wir die Ergebnisse der `glm`-Funktion für die Standard-Linkfunktionen mit den Ergebnissen der `glm`-Funktion für die modifizierten Linkfunktionen für die optimalen Parameter. Abschließend geben wir eine kurze Zusammenfassung und einen Ausblick auf interessante Weiterführungen der vorgestellten Themen und Funktionen.

Contents

List of Figures

List of Tables

1   Introduction

In today's world, further analysis of the available data is often necessary. By saying the word "data" we refer to a set of n observations or measurements made from different groups of objects or subjects (compare to Dobson (1999)(p. 1)). Regression analysis is commonly used to describe the relationship among certain variables. Thus, we focus on one *dependent* variable and try to describe it through one or more *independent* variables.

$$
\text{non-random measurement} \begin{cases} \underline{\textbf{Covariates}} \\ \text{independent variables} \\ \text{explanatory variables} \\ \text{predictor variables} \\ \text{regressor variables} \end{cases}
$$

$\Downarrow$ Examine the relationship between the observations $\Downarrow$

$$
\text{random measurement} \begin{cases} \underline{\textbf{Response}} \\ \text{dependent variable} \\ \text{outcome} \end{cases}
$$

Table 1.1: Basic notation of the thesis.

Hence, we find ourselves in a *probabilistic model* and model our response through a *linear regression model* using the following equation:

$$
y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i
$$

Here y_i is the random *response* and $x_{i1}, \ldots, x_{ik}$ is the set of known *covariates*. $\beta_0, \beta_1, \ldots, \beta_k$ are the unknown *regression parameters* and ε_i is the random *error term*. The underlying assumption is that the relationship between the response and the covariates is linear. Myers et al. (2002)(p. 1ff.) states more precisely that the mean of the response is a linear function of the unknown parameters. We will describe the *linear regression model* in detail in Chapter 2. The description in Section 2.1 provides the notional basis for the entire thesis. In addition we will focus on the *least squares estimation* and on the *maximum likelihood estimation* in the framework of the linear models (see Section 2.2). However, as Section 2.3 reveals, the linear models may be inappropriate in some situations.

For that reason we introduce the *generalized linear models* (GLMs) in Chapter 3. The advantage of GLMs compared to the linear regression models is, that one can examine a more applicable class of error distributions. The derived statistical models can handle distributions coming from the *exponential family*, for example:

- normal distribution
- binomial distribution
- Poisson distribution
- gamma distribution

After the description of the generalized linear models in Section 3.1 we will focus on the estimation of the regression parameters. Section 3.2 will address the *maximum likelihood estimation.* We will derive the *Fisher scoring algorithm* and rewrite it into the *iterative weighted least squares algorithm.* Both algorithms can be used to estimate the vector of unknown regression parameters $\boldsymbol{\beta} \in \mathbb{R}^p$. In Section 3.3 we will give an overview of important *families.* In particular, the implementation of the common link functions in R is of interest. Afterwards we will define the *deviance* as a measure for the goodness of fit of a generalized linear model (see Section 3.4). The third chapter will conclude with an overview of the link functions in R and additional comments (see Section 3.5).

For every generalized linear model we have to define a relationship between the linear predictor and the mean through the so-called *link functions.* The problem is that we have to choose the link function before even getting started with the regression analysis. By this time we often have insufficient information about an adequate choice of the link function. Consequently, we have to face cases, in which the link function seems inappropriate. In these cases we want to increase the goodness of fit by using *parametric link families* performing a transformation on the tails (compare to Czado (1992) and Czado (2007)). In Chapter 6 we will introduce the parametric link families and we will exemplify how GLMs in R can be fitted using user-defined link functions (see Section 6.2). We will present the tail modifying functions and an extension to the `glm` function in R (called `glmProfile`). Either a single tail ("left" or "right") or both tails can be modified to increase the goodness of fit. Therefore the parametric link functions can be seen as one- or two-parametric extensions of the common link functions. The parameters of these user-defined link functions can be derived by the `glmProfile` function.

We will accompany the theory by examples of the corresponding functions applied on the data sets we will introduce in Chapter 4. In the first part we will apply the ordinary `glm` function with a common link function (see Section 3.3 and Chapter 5) on the data sets. In Section 6.2 we will examine the data again using parametric link functions in the ordinary `glm` function. The optimal parameters for the tail modifiying GLMs can be derived by the `glmProfile` function, which we will define in Section 6.3. In Chapter 7 we want to present the output of the `glmProfile` function.

To clarify the improvement of the tail transformations we will present a short summary of the main results of the examples in Chapter 8. The results show that the tail modified GLMs will increase the goodness of fit compared to the ordinary generalized regression models in all cases. To conclude the work we will give a brief outlook and a summary of the thesis in Chapter 9. The calculations, which needed to be done throughout the work, can be found in the appendix (see Chapter A).

Remark 1.1 (Usage of statistical software)
Throughout this thesis we will support our theory using examples implemented in the software environment of R using the R version 3.1.0 (2014-04-10). R is mainly used for statistical computing and graphics and runs on almost every operating system. In addition the users can apply a tremendous variety of functions coming along with the common packages. A lot of functions and routines used in special analyses can be included using the corresponding packages. There is also the possibility to include R output in LaTeX-documents. For my thesis I used the package `knitr` (see Xie (2013)) which allows to write dynamic documents with R.

R is very similar to the environment of the programming language S and it is a GNU project (i.e. available as free software). A lot of code written in S runs under R as well. Apart from some important differences we can regard R as a different version of S. For further information about the R Project please visit: `http://www.r-project.org/`.

Remark 1.2 (Idea of parametric link functions in S)
In this thesis we want to modify the common link functions in R. In Czado (1992) the ordinary link families were extended by using an advantageous parametric class of link transformations. This idea was elaborated in Czado (2007), where parametric link families were used to fit GLMs in S. All functions (i.e. the `hpsi` functions and the `glmProfile` function), which we present in this thesis, were therefore implemented in the statistical environment of S.

The main task of this thesis is to implement the parametric link families in the statistical programming environment of R. In most cases we could use the framework as presented in Czado (2007), converting the code into a basis running in the R environment. However, there were also parts, in which we had to come up with new ideas (especially the definition of the link functions is different). We also want to present the underlying theory of generalized linear models in detail following the notation in Czado et al. (2013). The resulting R code and the R data frames will be presented in the corresponding chapters. Moreover, all data sets and functions described throughout this thesis were put together in the package `ParLinkFam`, which contains help files with descriptions and application examples.

2 The linear models

As the classical regression models are of great importance in statistical data analysis, we first consider the simple linear regression model. In Chapter 2 we follow Czado et al. (2013)(Sections 2.1 and 2.2) in presenting important results for the linear models. The notional basis will be introduced in Section 2.1, where we will formulate the linear model. Section 2.2 will focus on the estimation of the unknown parameters through least squares estimation and through maximum likelihood estimation. However, as Section 2.3 shows, the linear models cannot be used in all the desired applications. Hence, we will use the underlying theory to derive the generalized linear models (as presented in Chapter 3).

2.1 Model description

The content presented in Section 2.1 can be found in Czado and Schmidt (2011)(Section 7.1) and Bates and Watts (2007)(Section 1.1). The linear regression model describes the random *response* Y in dependency of the k known *predictors* (denoted by: x_1, ..., x_k). For each *observation* (i.e. $\forall i \in \{1, \cdots, n\}$) we assume that the observation y_i made from the response Y_i is a linear function of the values of the covariates for this observation, denoted by x_{i1}, ..., x_{ik}. Of course this will not fit perfectly in all cases. Thus, for each observation we will have to add a random error term (denoted by ε_i). In addition the linear model contains the so-called *intercept* β_0.

Definition 2.1 (The linear model and its assumptions)
   (i) **linearity:**
      for each observation we assume that the random response is related to the covariates in a linear way:

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i \qquad \forall i \in \{1, \ldots, n\} \qquad (2.1)$$

   with a random error term satisfying: $\mathbb{E}\left[\varepsilon_i\right] = 0 \quad \forall i \in \{1, \cdots, n\}$.

  (ii) **independence:**
      the random error terms $\varepsilon_1, \ldots, \varepsilon_n$ are independent.

 (iii) **variance homogeneity:**
      they also have a constant variance and it holds:

$$\mathbb{V}ar\left[Y_i\right] = \mathbb{V}ar\left[\varepsilon_i\right] = \sigma^2$$

 (iv) **normality:**
      lastly, the error terms $\varepsilon_1, \ldots, \varepsilon_n$ follow a normal distribution.

Remark 2.2
   • for one specific observation $i \in \{1, \ldots, n\}$ we will summarize the covariates x_{i1}, ..., x_{ik} in the covariate vector $\boldsymbol{x}_i$, also taking into account the influence of the intercept, i.e. we get:
$$\boldsymbol{x}_i := (1, x_{i1}, ..., x_{ik})^\top \in \mathbb{R}^{k+1} \qquad (2.2)$$

- we are interested in estimating the *regression parameters* $\beta_0, \ldots, \beta_k$. All in all we will thus estimate $p := k + 1$ *unknown* parameters from the n observations. The vector of regression parameters will be denoted by:

$$\boldsymbol{\beta} := (\beta_0, \beta_1, \ldots, \beta_k)^\top \in \mathbb{R}^p \tag{2.3}$$

- Y_i is a random variable and thus it can have an expectation or a variance. On the other hand, we also have non-random quantities, for example the covariate vector $\boldsymbol{x}_i$. However, sometimes random variables are also inconveniently denoted with small letters (e.g. ε_i). Further commonly used abbreviations and notation are presented in Section A.5.

- due to Definition 2.1 we can conclude for the error terms that $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)\ \forall i \in \{1, \cdots, n\}$ since they follow a normal distribution with mean $\mu_\varepsilon = 0$ and a constant variance σ^2. Thus, $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. random variables.

- since $\beta_0, \ldots, \beta_k \in \mathbb{R}$ and $\boldsymbol{x}_i \in \mathbb{R}^p$ we see that only ε_i is random in Equation (2.1). Consequently, the distribution of Y_i must be the same as for ε_i. Thus, Y_i has to be normally distributed. For the mean μ_i and the variance σ^2 we get:

$$\mu_i := \mathbb{E}\left[Y_i\right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} = \boldsymbol{x}_i^\top \boldsymbol{\beta} \tag{2.4}$$
$$\mathbb{V}ar\left[Y_i\right] = \mathbb{V}ar\left[\varepsilon_i\right] = \sigma^2$$
$$\Rightarrow Y_i \sim \mathcal{N}(\mu_i, \sigma^2) \qquad \forall i \in \{1, \ldots, n\}$$

Distribution of the random vectors

Since $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. $\mathcal{N}(0, \sigma^2)$ distributed it is appropriate to think about the distribution of the vector $\boldsymbol{\varepsilon} := (\varepsilon_1, \ldots, \varepsilon_n)^\top$. Afterwards we can derive a distribution of the following vector:

$$\boldsymbol{Y} := (Y_1, \ldots, Y_n)^\top \tag{2.5}$$

in order to transform the linear regression model of Definition 2.1 into matrix-vector notation. For this we define the *design matrix* $X \in \mathbb{R}^{n \times p}$.

Definition 2.3 (The design matrix)
For n observations and the n corresponding covariate vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ the design matrix $X \in \mathbb{R}^{n \times p}$ is given by:

$$X := \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \vdots & & & & \vdots \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{pmatrix} \overset{Eq.\ (2.2)}{=} \begin{pmatrix} \rule[0.5ex]{1em}{0.4pt} & \boldsymbol{x}_1^\top & \rule[0.5ex]{1em}{0.4pt} \\ \rule[0.5ex]{1em}{0.4pt} & \boldsymbol{x}_2^\top & \rule[0.5ex]{1em}{0.4pt} \\ & \vdots & \\ \rule[0.5ex]{1em}{0.4pt} & \boldsymbol{x}_n^\top & \rule[0.5ex]{1em}{0.4pt} \end{pmatrix} \in \mathbb{R}^{n \times p}$$

Definition 2.4 (Matrix-vector notation for linear models)
Using this definition we can rewrite the linear model in matrix-vector notation as follows:

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \varepsilon \qquad \text{with} \ \ \boldsymbol{\varepsilon} \sim \mathcal{N}_n(\boldsymbol{0}, \sigma^2 \mathbb{I}_n)$$

where $\mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$ denotes the n-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix Σ (see also Subsection 2.2.2).

Remark 2.5
Assuming normality conditions we have (according to Definition 2.4):

$$\mathbb{E}\left[\boldsymbol{Y}\right] \stackrel{\mathbb{E}[\varepsilon]=0}{=} X\boldsymbol{\beta} \tag{2.6}$$

$$\mathbb{V}ar\left[\boldsymbol{Y}\right] = \sigma^2 \mathbb{I}_n$$

$$\Rightarrow \boldsymbol{Y} \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 \mathbb{I}_n) \tag{2.7}$$

2.2   Parameter estimation in linear models

We want to derive estimates for the unknown vector of regression parameters $\boldsymbol{\beta} \in \mathbb{R}^p$, assuming that the conditions of Definition 2.1 hold. We will introduce two techniques to derive the estimates $\widehat{\boldsymbol{\beta}} = (\widehat{\beta}_0, \ldots, \widehat{\beta}_k)^\top \in \mathbb{R}^p$ out of the n observations given by the observations made from the response (i.e. $y_1, \ldots, y_n$) and the observations made from the corresponding covariate vector (i.e. $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$). Together these observations will form the data (see Myers et al. (2002)(p. 8)).

2.2.1   Least squares estimation

In this subsection we will follow the calculation in Myers et al. (2002)(p. 7ff.). We want to find values for $\widehat{\boldsymbol{\beta}}$ s.t. for every observation $i \in \{1, \ldots, n\}$ the *fitted values*

$$\widehat{y_i} := \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \cdots + \widehat{\beta}_k x_{ik}$$

do not lie too far from the observations $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$. Therefore we want to determine the values $\widehat{\beta}_0, \ldots, \widehat{\beta}_k$ of $\beta_0, \ldots, \beta_k$ that minimize the *sum of the squares of the errors*, which is given by:

Definition 2.6 (Sum of the squares of the errors)
The i-th error term ($i \in \{1, \ldots, n\}$) is given by:

$$\varepsilon_i = y_i - (\beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}) \stackrel{see\ Eq.\ (2.4)}{=} y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta}$$

Consequently, the sum of the squares of the errors is defined as:

$$Q(\boldsymbol{\beta} \mid \boldsymbol{y}) := \sum_{i=1}^{n} \varepsilon_i^2 \stackrel{see\ Eq.\ (2.4)}{=} \sum_{i=1}^{n} (y_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2$$

Again we can rewrite the quantities in vector notation:

$$\boldsymbol{y} := (y_1, \ldots, y_n)^\top \in \mathbb{R}^n$$

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top \stackrel{Def.\ 2.4}{=} \boldsymbol{y} - X\boldsymbol{\beta} \in \mathbb{R}^n$$

Therefore, we can rewrite the sum of the squares of the errors as:

$$Q(\boldsymbol{\beta} \mid \boldsymbol{y}) = \|\boldsymbol{\varepsilon}\|_2^2 = \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 \tag{2.8}$$

To minimize Equation (2.8) we must at least satisfy the necessary condition of first order (compare to Ulbrich and Ulbrich (2012)(Chapter 5)):

$$\frac{\partial Q(\boldsymbol{\beta} \mid \boldsymbol{y})}{\partial \boldsymbol{\beta}} = 0$$

$$\Leftrightarrow \frac{\partial}{\partial \boldsymbol{\beta}} \left(\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2 \right) = 0$$

$$\Leftrightarrow \frac{\partial}{\partial \boldsymbol{\beta}} \left((\boldsymbol{y} - X\boldsymbol{\beta})^\top (\boldsymbol{y} - X\boldsymbol{\beta}) \right) = 0$$

$$\Leftrightarrow \frac{\partial}{\partial \boldsymbol{\beta}} \left(\boldsymbol{y}^\top \boldsymbol{y} - \underbrace{\boldsymbol{\beta}^\top X^\top \boldsymbol{y}}_{=\left(\boldsymbol{y}^\top X\boldsymbol{\beta}\right)^\top \in \mathbb{R}} - \boldsymbol{y}^\top X\boldsymbol{\beta} + \boldsymbol{\beta}^\top X^\top X\boldsymbol{\beta} \right) = 0$$

$$\Leftrightarrow \frac{\partial}{\partial \boldsymbol{\beta}} \left(\boldsymbol{y}^\top \boldsymbol{y} - 2 \underbrace{\boldsymbol{y}^\top X}_{=(X^\top \boldsymbol{y})^\top} \boldsymbol{\beta} + \boldsymbol{\beta}^\top X^\top X\boldsymbol{\beta} \right) = 0$$

$$\Leftrightarrow \frac{\partial}{\partial \boldsymbol{\beta}} \left(\boldsymbol{y}^\top \boldsymbol{y} - 2(X^\top \boldsymbol{y})^\top \boldsymbol{\beta} + \boldsymbol{\beta}^\top X^\top X\boldsymbol{\beta} \right) = 0$$

$$\overset{X^\top X \text{ is symmetric}}{\Leftrightarrow} -2X^\top \boldsymbol{y} + 2X^\top X\boldsymbol{\beta} = 0$$

$$\Leftrightarrow X^\top X\boldsymbol{\beta} = X^\top \boldsymbol{y} \tag{2.9}$$

Remark 2.7 (Least squares normal equation)
Equation (2.9) is called the *least squares normal equation*. If the design matrix $X \in \mathbb{R}^{n \times p}$ is invertible (i.e. has full rank [rank(X) = min$\{p, n\}$ = p, assuming $n > p$]), then we can rewrite the least squares normal equation by solving it for the estimate $\widehat{\boldsymbol{\beta}}$. I.e. the minimum of $Q(\boldsymbol{\beta} \mid \boldsymbol{y})$ is attained at:

$$\widehat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top \boldsymbol{y}$$

We refer to the solution $\widehat{\boldsymbol{\beta}}$ using the term *least squares solution* or *ordinary least squares estimator for* $\widehat{\boldsymbol{\beta}}$.

2.2.2   Maximum likelihood estimation

In the following we will derive the maximum likelihood estimator for $\boldsymbol{\beta}$ following the general definitions given in Czado and Schmidt (2011)(Section 3.3). Recall that we derived Equation (2.7) and accordingly we have (assuming normality conditions hold):

$$\boldsymbol{Y} \sim \mathcal{N}_n(X\boldsymbol{\beta}, \sigma^2 \mathbb{I}_n)$$

In this case the likelihood function is given by the density of the multivariate normal distribution in n dimensions. According to Seber (1977)(p. 22ff.) for $\boldsymbol{Y} \sim \mathcal{N}_n(\boldsymbol{\mu}, \Sigma)$ (where $\Sigma \in \mathbb{R}^{n \times n}$ is a positive definite matrix) the density is given by:

$$f_{\boldsymbol{Y}}(\boldsymbol{y} \mid \boldsymbol{\mu}, \Sigma) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Sigma}} \exp \left\{ -\frac{1}{2} (\boldsymbol{y} - \boldsymbol{\mu})^\top \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \right\}$$

Now for the response vector $\boldsymbol{Y}$ in a linear model we have:

$$\Sigma = \sigma^2 \mathbb{I}_n$$
$$\Rightarrow \sqrt{\det \Sigma} = \sqrt{(\sigma^2)^n} = (\sigma^2)^{\frac{n}{2}}$$
$$\Rightarrow \Sigma^{-1} = \frac{1}{\sigma^2} \mathbb{I}_n$$
$$\Rightarrow (\boldsymbol{y} - \boldsymbol{\mu})^\top \Sigma^{-1}(\boldsymbol{y} - \boldsymbol{\mu}) = \frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{\mu})^\top(\boldsymbol{y} - \boldsymbol{\mu}) = \frac{1}{\sigma^2}\|\boldsymbol{y} - \boldsymbol{\mu}\|_2^2$$
$$\mu = X\boldsymbol{\beta}$$
$$\Rightarrow \|\boldsymbol{y} - \boldsymbol{\mu}\|_2^2 = \|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2$$

Hence, the likelihood function of $(\boldsymbol{\beta}, \sigma)$ given $\boldsymbol{y}$ equals:

$$L(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y}) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2\right\}$$

The log likelihood is thus given by:

$$\ln(L(\boldsymbol{\beta}, \sigma \mid \boldsymbol{y})) = \ln\left(\frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}}\right) - \frac{1}{2\sigma^2}\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2$$
$$= -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\|\boldsymbol{y} - X\boldsymbol{\beta}\|_2^2$$
$$\stackrel{Eq.\ (2.8)}{=} -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}Q(\boldsymbol{\beta} \mid \boldsymbol{y}) \qquad (2.10)$$

Remark 2.8 (Same estimates from both estimation methods)
Since the only part depending on $\boldsymbol{\beta}$ in Equation (2.10) is $Q(\boldsymbol{\beta} \mid \boldsymbol{y})$, the maximum likelihood estimation yields to the same estimate as the least squares estimation. In particular, the least squares solution $\widehat{\boldsymbol{\beta}}$ is both the least squares estimate and the maximum likelihood estimate of $\boldsymbol{\beta}$.

2.3   Disadvantages of the linear model

The arguments presented in this section can be found in Myers et al. (2002)(Chapter 1).

Remark 2.9 (Importance of the linear regression model)
The linear regression models are of importance because of a variety of reasons:

(i) if we have that $\mathbb{E}[Y_i] = f(x_i)$ (for a single covariate x_i) is the relationship between the response and the covariate, then a first order Taylor approximation yields to:

$$\mathbb{E}[Y_i] = f(x_0) + \frac{df(x)}{dx}\big|_{x=x_0}(x - x_0) + remainder$$

Which leads to (ignoring the remainder and the error term): $\beta_0 + \beta_1(x - x_0)$. For k covariates the first order Taylor approximation yields to Equation (2.4). Hence, the linear models approximate the response as a first order Taylor approximation does.

(ii) one can estimate the unknown p = k+1 parameters $\beta_0, \ldots, \beta_k$ by solving p linear equations simultaneously using the method of *least squares*. Many programs facilitate regression model fitting through a implementation of this method.

(iii) further the statistical theory is well-developed and implemented in statistical computer software (like R).

(iv) we can extend the theory of linear models to derive generalized linear models.

On the one hand, the linear regression model and its requirements are easy to understand. On the other hand, this implies that in many situations the linear model is considered as too restrictive and not suitable. Some reasons why it may be inappropriate are:

- we can only apply it for responses which follow a normal distribution.

- for continuous response variables the normality assumption can be unrealistic, e.g. non-negative and highly right-skewed responses.

These restrictions imply that we cannot analyze discrete responses such as injuries or patients suffering from specific diseases or the occurrence of natural phenomena (like hurricanes or earthquakes). In addition we cannot explore binary responses as many fields in science and engineering do. Often we regard responses being either a success (encoded with 1) or a failure (encoded with 0).

We thus introduce a more general regression model meeting our requirements:

- applicable to a variety of problems, e.g. by allowing distributions from the exponential family, such as:

    - normal distribution (see Subsection A.1.1)
    - binomial distribution (see Subsection A.1.2),
    - Poisson distribution (see Subsection A.1.3),
    - gamma distribution (see Subsection A.1.4),

- well developed statistical theory, i.e. a lot of literature related to the theory.

- computer software supporting the framework of the model.

The generalized linear models (GLMs) satisfy all of the requirements. Therefore, we dedicate the complete next chapter to these important regression models.

3   The generalized linear models

On the following pages we will present important results about generalized linear models. The content presented below is also explained in Czado et al. (2013)(Sections 3.1, 3.2 and 3.3). As mentioned before, we now also allow for normal, binomial, Poisson and gamma responses all being members of the exponential family. For this we will introduce the exponential family and clarify the parameters for the single distributions (the calculation can be found in the Section A.1). Then, we will formulate the generalized linear model (see Section 3.1). In Section 3.2 we will focus on the theory of estimating the vector of regression parameters $\boldsymbol{\beta}$ in this setting. The central role is played by the maximum likelihood estimation. The big difference to linear models is that we now have to solve non-linear equations, for which we will derive the iteratively weighted least squares algorithm. We will also focus on the concept of families and link functions and we will show, how they are implemented in the statistical programming environment of R (see Section 3.3). Afterwards we will define the so-called deviance in Section 3.4. The deviance serves as a criterion for assessing the goodness of fit of a generalized linear model. The chapter about generalized linear models will conclude with a short overview and comments on the link functions (see Section 3.5).

3.1   Model description

According to Fahrmeir and Tutz (2001)(p. 19ff.), the density (or probability mass function, respectively) of the response Y_i in a GLM (for $i \in \{1, \ldots, n\}$) is a member of the *exponential family*. This is a very useful class of distributions, which we will now define.

Definition 3.1 (Exponential family)
A random variable Y follows a distribution function of the exponential family, if its density (or probability mass function, respectively) can be written in the following way:

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

where

- $b(\cdot)$ and $c(\cdot)$ are specified functions determined by the distribution.

- $\phi \in \mathbb{R}^+$ is the so-called *scale* or *dispersion parameter*.

- $\theta \in \mathbb{R}$ is called *canonical* or *natural parameter*.

- ω is the weight.

Remark 3.2
For ease of notation we will often write

$$f(y \mid \theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \tag{3.1}$$

for random variables with distributions belonging to the exponential family. This form is given in McCullagh and Nelder (1983)(p. 20f) and it is valid, since we commonly have

$$a(\phi) = \frac{\phi}{\omega}. \tag{3.2}$$

Example 3.3 (Members of the exponential family)
According to Hardin and Hilbe (2007)(p. 9), the exponential family includes the following distributions:

- normal or Gaussian distribution
- binomial distribution
- Poisson distribution
- gamma distribution

- inverse Gaussian distribution
- geometric distribution
- negative binomial distribution

   In the following table we summarize important components of the exponential family distribution for the most important distributions in the setting of generalized linear models. For each of the following distributions we show in the appendix (see Section A.1), that they belong to the exponential family by deriving the single components. A similar table can be found in Fahrmeir and Tutz (2001)(p. 21).

Distribution	$\theta(\mu)$	$b(\theta)$	ϕ	ϕ known	ω	$a(\phi) = \frac{\phi}{\omega}$
$\mathcal{N}(\mu, \sigma^2)$	μ	$\frac{\theta^2}{2}$	σ^2	✗	1	σ^2
ScaledBin(n,p)	$\ln(\frac{p}{1-p})$	$\ln(1 + \exp\{\theta\})$	1	✓	n	$\frac{1}{n}$
$Poi(\lambda)$	$\ln(\lambda)$	$\exp\{\theta\}$	1	✓	1	1
$\Gamma(\mu, \nu)$	$-\frac{1}{\mu}$	$-\ln(-\theta)$	$\frac{1}{\nu}$	✗	1	$\frac{1}{\nu}$

Table 3.1: Components of the exponential family distributions for important families.

   Similar to the definitions given in Fahrmeir and Tutz (2001)(p. 434) and McCullagh and Nelder (1983)(p. 18), we will now define the components of a generalized linear model.

Definition 3.4 (Generalized linear model)
A *generalized linear model* will be described by means of the following three components

(i) **the random component:**
   for each observation $i \in \{1, \cdots, n\}$ the corresponding random response Y_i is independent of the other responses and follows a distribution belonging to the exponential family, i.e. its density (or probability mass function, respectively) is of the form:
   $$f(y_i \mid \theta_i, \phi) = \exp\left\{ \frac{y_i \theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right\}$$

(ii) **the systematic component:**
   for each observation $i \in \{1, \cdots, n\}$ we define the *linear predictor* η_i by:
   $$\eta_i = \eta_i(\boldsymbol{\beta}) := \boldsymbol{x}_i^T \boldsymbol{\beta} \overset{Eq.\ (2.2)}{\underset{Eq.\ (2.3)}{=}} \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} \tag{3.3}$$

   where $\beta_0 \in \mathbb{R}$ is called *intercept* and $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of *regression parameters*.

(iii) **the parametric link component:**
it relates the random component with the systematic component. Therefore, we consider (comparable to Equation (2.4)) the mean $\mu_i = \mathbb{E}[Y_i]$ for each observation $i \in \{1, \cdots, n\}$. The difference is, that we do not assume, that the mean is exactly equal to the linear predictor. Instead we assume a relationship according to the so-called *link function* $g : \mathcal{G} \to \mathcal{H}$ (with $\mathcal{G}, \mathcal{H} \subset \mathbb{R}$):

$$g(\mu_i) = \eta_i(\boldsymbol{\beta}) = \boldsymbol{x}_i^T \boldsymbol{\beta} \tag{3.4}$$

Remark 3.5
- similar to Equation (2.4), Definition 2.4 and Remark 2.5 we can rewrite the *linear predictor* in matrix-vector notation:

$$\boldsymbol{\eta}(\boldsymbol{\beta}) = \boldsymbol{\eta} = X\boldsymbol{\beta} \in \mathbb{R}^n \tag{3.5}$$

- the sets $\mathcal{G}, \mathcal{H}$ represent restrictions coming from the assumption $\mu_i = \mathbb{E}[Y_i]$. Surely this restriction of μ_i yields to restrictions on η_i (e.g. through the domain of the linkinverse $F(\eta)$ we get $\eta \overset{!}{\in} \mathcal{H}$).

- throughout the work we will denote by $F(\eta) = F(\eta_i)$ the inverse of the link function, i.e. $F(\cdot) = g^{-1}(\cdot)$ is a function of η. We have $F : \mathcal{H} \to \mathcal{G}$ and $\mu_i = F(\eta_i)$.

- Hence, we have: g is a function of μ (i.e. $g(\mu)$) with $g : \mathcal{G} \to \mathcal{H}$. Thus, we have $\mu \in \mathcal{G}$. For the linkinverse we have F: $\mathcal{H} \to \mathcal{G}$ is a function of η (i.e. F(η)). Thus, we have $\eta \in \mathcal{H}$. We can see that the link component of a GLM relates the linear predictor η_i to the expectation μ_i.

Theorem 3.6 (Expectation and variance of the exponential family)
Assume Y has a distribution from the exponential family, than we have

$$\mathbb{E}[Y] = b'(\theta)$$
$$\mathbb{V}ar[Y] = b''(\theta)a(\phi)$$

Proof:
A proof can be found in McCullagh and Nelder (1983)(p. 20f.).

$\square$

Remark 3.7 (Variance function $v(\mu_i)$)
As described in Fahrmeir and Tutz (2001)(p. 20), the canonical parameter θ_i is a function of the mean μ_i (i.e. $\theta_i = \theta(\mu_i)$). Further the variance is of the following form:

$$\mathbb{V}ar[y_i \mid \boldsymbol{x}_i] = \sigma^2(\mu_i) = \phi \frac{v(\mu_i)}{\omega_i}$$

where the variance function $v(\cdot)$ is determined by $v(\mu_i) = b''(\theta_i) = \frac{\partial^2 b(\theta_i)}{\partial^2 \theta_i}$. This separation for the variance is made, because $b''(\theta_i)$ depends on θ_i (and thus on μ_i) while the other part $a(\phi)$ is independent of θ_i (see McCullagh and Nelder (1983)(p. 21)).

Example 3.8 (Expectation and variance for the exponential family)
As in Fahrmeir and Tutz (2001)(p. 21), we can calculate the expectation and the variance for the members of the exponential family using Theorem 3.6 and the components we have presented in Table 3.1:

Distribution	$\mathbb{E}\left[Y\right] = b'(\theta)$	variance funct. $= b''(\theta)$	$\mathbb{V}ar\left[Y\right] = b''(\theta)\frac{\phi}{\omega}$
$\mathcal{N}(\mu,\sigma^2)$	$\mu = \theta$	1	$\frac{\sigma^2}{\omega} = \sigma^2$
ScaledBin(n,p)	$p = \frac{\exp\{\theta\}}{1+\exp\{\theta\}}$	$p(1-p) = \frac{\exp\{\theta\}}{(1+\exp\{\theta\})^2}$	$\frac{p(1-p)}{\omega} = \frac{p(1-p)}{n}$
$Poi(\lambda)$	$\lambda = \exp\{\theta\}$	λ	$\frac{\lambda}{\omega} = \lambda$
$\Gamma(\mu,\nu)$	$\mu = -\frac{1}{\theta}$	μ^2	$\frac{\mu^2}{\nu\omega} = \frac{\mu^2}{\nu}$

Table 3.2: Moments of generalized linear model families.

3.2   Parameter estimation in generalized linear models

We want to estimate the unknown parameter vector $\boldsymbol{\beta} \in \mathbb{R}^p$ in the following setting. We assume that $\boldsymbol{Y}$ (as defined in Equation (2.5)) fits in the setting of a GLM with covariate vector $\boldsymbol{x}_i$ for the i-th response Y_i given by:

$$\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})^\top \in \mathbb{R}^p \tag{3.6}$$

Remark 3.9 (Change in notation)
Now we want to introduce a more advantageous way to write the vectors $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$ introduced in the Equations (2.2) and (2.3).

- Equation (3.6) does not contain any ones for the intercept ($x_{i1} = 1$). Instead it begins with x_{i1} as its first component while now x_{ip} is the last component. It is still a vector of dimension p (p = k+1), i.e. $\boldsymbol{x}_i \in \mathbb{R}^p$ and hence, the design matrix X has still the same form (i.e. $X \in \mathbb{R}^{n \times p}$).

- therefore our vectors are shifted in the following sense. We can w.l.o.g. assume the intercept to be one of the parameters. We therefore denote the intercept β_0 by β_1 with the following notation:

$$(1, x_{i1}, \ldots, x_{ik}) \to (x_{i1}, \ldots, x_{ip})$$
$$(\beta_0, \beta_1, \ldots, \beta_k) \to (\beta_1, \beta_2, \ldots, \beta_p)$$

- this is just a change in notation proving a more comfortable notation. Nothing changes in the mathematical theory we have developed so far.

The estimation of $\boldsymbol{\beta}$ can be done by using the *maximum likelihood estimation* (MLE) in GLMs (see Definition 3.11). To develop the theory about MLE we need some definitions. By Theorem 3.6 and Remark 3.7 we can rewrite the canonical parameter θ_i in terms of the mean μ_i. This motivates the following definition:

Definition 3.10 (Inverse mean function in GLMs)
The inverse mean function $h(\cdot)$ in a GLM is defined through:

$$h(\cdot) = (b')^{-1}(\cdot)$$

and satisfies

$$h(\mu_i) = \theta_i \qquad\qquad \forall i \in \{1, \dots, n\} \qquad\qquad (3.7)$$

Further μ_i is a function of η_i (see Remark 3.5) and η_i depends on the regression parameters $\beta_1, \dots, \beta_p$ (see Equation (3.4)). Now we want to concentrate on the estimation of $\boldsymbol{\beta}$ and thus other parameters are assumed to be known.

In the following we want to introduce the method of maximum likelihood estimation comparable to Czado and Schmidt (2011)(Section 3.3). The maximum likelihood estimation is the most important method to derive an estimator for the unknown parameter or parameter vector, respectively. In our case this is the vector of regression parameters $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p) \in \mathbb{R}^p$. Therefore, this method determines the *maximum likelihood estimate* (MLE) denoted by $\widehat{\boldsymbol{\beta}}$ by maximizing the so-called *likelihood function*. The estimator (random variable) is called *maximum likelihood estimator* (also abbreviated by MLE) and unfortunately it is also denoted by $\widehat{\boldsymbol{\beta}}$. This notional inconvenience has also been addressed in Wood (2006)(p. 60). Hence, we want to find p maximum likelihood estimates $\widehat{\beta}_j$ for $j \in \{1, \dots, p\}$ from the data given by the observations y_i we can observe from Y_i (for $i \in \{1, \dots, n\}$).

Definition 3.11 (Maximum likelihood estimation in GLMs)
Given one single observation y_i, the likelihood function of the parameter $\boldsymbol{\beta}$ is given by its density (or probability mass function, respectively). In GLMs the response Y_i follows a distribution of the exponential family (see Definition 3.4). Therefore, the likelihood function is given by:

$$L_i(\boldsymbol{\beta}, \phi \mid y_i) := f(y_i \mid \theta_i, \phi) \overset{\substack{Y_i \sim \text{ Exp. Fam.} \\ \text{see Def. 3.1}}}{=} \exp\left\{\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right\} \qquad (3.8)$$

Maximizing the likelihood function to obtain the MLE $\widehat{\boldsymbol{\beta}}$ is equivalent to optimizing the so-called *log likelihood*. The log likelihood for observation y_i is given by:

$$l_i = l_i(\boldsymbol{\beta}, \phi \mid y_i) := \ln\left[L_i(\boldsymbol{\beta}, \phi \mid y_i)\right] \overset{Eq. (3.8)}{=} \frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \qquad (3.9)$$

By Definition 3.4 the random responses Y_i of a GLM are independent. Therefore, the joint density is simply the product of all marginal densities (the same holds for probability mass functions). Consequently, the likelihood function for the vector of observations $\boldsymbol{y} := (y_1, \dots, y_n)^\top$ (we observe from the vector of responses $\boldsymbol{Y}$) is the product of the likelihood functions $L_i(\boldsymbol{\beta}, \phi \mid y_i)$ for the single observations. Hence, we get:

$$L(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}) = \prod_{i=1}^n L_i(\boldsymbol{\beta}, \phi \mid y_i) \overset{Eq. (3.8)}{=} \exp\left\{\sum_{i=1}^n \left(\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right)\right\}$$

$$\overset{Eq. (3.9)}{=} \exp\left\{\sum_{i=1}^n l_i(\boldsymbol{\beta}, \phi \mid y_i)\right\}$$

In this setting the log likelihood is given by:

$$l(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}) := \ln\left[L(\boldsymbol{\beta}, \phi \mid \boldsymbol{y})\right] = \sum_{i=1}^{n} l_i(\boldsymbol{\beta}, \phi \mid y_i) = \sum_{i=1}^{n} \left(\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right) \quad (3.10)$$

Therefore, the log likelihood in a GLM is given by the sum of the log likelihoods for the single observations y_i for $i \in \{1, \ldots, n\}$. This derivation can be verified by comparing the steps to the calculation made in Wood (2006)(p. 61ff.).

Now our goal is to maximize the log likelihood given in Equation (3.10). For this we must at least satisfy the first order optimization criterion: $\frac{\partial l(\boldsymbol{\beta}, \phi \mid \boldsymbol{y})}{\partial \beta_j} = 0 \; \forall j \in \{1, \ldots, p\}$ and hence we need partial derivatives for iterative gradient descendant methods (see Ulbrich and Ulbrich (2012)). As calculated in Dobson (1999)(p. 146) we receive:

$$\frac{\partial l_i}{\partial \beta_j} = \frac{y_i - \mu_i}{\mathbb{V}ar\left[Y_i\right]} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) x_{ij} \overset{Thm. \ 3.6}{=} \frac{y_i - \mu_i}{b''(\theta_i)a(\phi)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) x_{ij} \quad (3.11)$$

$$\frac{\partial l(\boldsymbol{\beta}, \phi \mid \boldsymbol{y})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{\partial l_i}{\partial \beta_j} \overset{Eq. \ (3.11)}{=} \sum_{i=1}^{n} \left(\frac{y_i - \mu_i}{b''(\theta_i)a(\phi)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right) x_{ij}\right) \quad (3.12)$$

Now we define the *weights in GLMs* (see McCullagh and Nelder (1983)(p. 33)) to rewrite the optimization condition in Equation (3.12):

Definition 3.12 (Weights in GLMs)
Let $b''(\cdot)$ be the variance (see Remark 3.7), then the weights in generalized linear models are defined through:

$$W_i = W_i(\boldsymbol{\beta}) := \frac{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}{b''(\theta_i)}$$

Remark 3.13
With this definition the formula in Equation (3.12) is equivalent to:

$$\frac{\partial l(\boldsymbol{\beta}, \phi \mid \boldsymbol{y})}{\partial \beta_j} = \sum_{i=1}^{n} \frac{1}{a(\phi)}(y_i - \mu_i)W_i(\boldsymbol{\beta})\left(\frac{\partial \eta_i}{\partial \mu_i}\right) x_{ij}$$

In the following calculations we will derive the quantities presented in Fahrmeir and Tutz (2001)(p. 38ff.) and McCullagh and Nelder (1983)(p. 31ff.) using our notation.

Remark 3.14
Since $a(\phi)$ is independent of $\boldsymbol{\beta}$, we don't need to consider the scale parameter ϕ while optimizing. In Fahrmeir and Tutz (2001)(p. 38) this is reflected by assuming w.l.o.g. $\phi = 1$. For this reason the following definitions are given in their unscaled forms, not including the parameter ϕ. This is sufficient to derive point estimates of $\boldsymbol{\beta}$.

Following the calculation in McCullagh and Nelder (1983)(p. 32ff.), it is sufficient to solve the so-called *unscaled score equations of a GLM* in order to get the MLE $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$:

Definition 3.15 (Unscaled score equations in GLMs)
We define the unscaled score through:

$$s_j(\boldsymbol{\beta}, \boldsymbol{y}) := \sum_{i=1}^n (y_i - \mu_i) W_i(\boldsymbol{\beta}) \left(\frac{\partial \eta_i}{\partial \mu_i} \right) x_{ij} \overset{!}{=} 0$$

The set of equations defined through $s_1(\boldsymbol{\beta}, \boldsymbol{y}), \ldots, s_p(\boldsymbol{\beta}, \boldsymbol{y})$ is called the *unscaled score equations*. Again we define the corresponding vector through:

$$s(\boldsymbol{\beta}, \boldsymbol{y}) := (s_1(\boldsymbol{\beta}, \boldsymbol{y}), \ldots, s_p(\boldsymbol{\beta}, \boldsymbol{y}))^\top \in \mathbb{R}^p \tag{3.13}$$

This vector is called the p-dimensional score function.

Since the likelihood equations are non-linear, they often can be solved only numerically through iterative algorithms (see Fahrmeir and Tutz (2001)(p. 42)). One of them is the *Fisher scoring algorithm*, which will be derived now.

Definition 3.16 (Unscaled Hessian matrix in GLMs)
The unscaled Hessian matrix in a GLM is given by:

$$H = H(\boldsymbol{\beta}, \boldsymbol{y}) := \frac{\partial s(\boldsymbol{\beta}, \boldsymbol{y})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial s_1(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_1} & \frac{\partial s_1(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_2} & \cdots & \frac{\partial s_1(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_p} \\ \frac{\partial s_2(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_1} & \frac{\partial s_2(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_2} & \cdots & \frac{\partial s_2(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial s_p(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_1} & \frac{\partial s_p(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_2} & \cdots & \frac{\partial s_p(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_p} \end{pmatrix}$$

$$= \left(\frac{\partial^2 l(\boldsymbol{\beta}, \boldsymbol{y})}{\partial \beta_i \partial \beta_j} \right)_{i,j \in \{1,\ldots,p\}} = \begin{pmatrix} \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial^2 \beta_1} & \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_1 \partial \beta_2} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_1 \partial \beta_p} \\ \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_2 \partial \beta_1} & \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial^2 \beta_2} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_2 \partial \beta_p} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_p \partial \beta_1} & \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial \beta_p \partial \beta_2} & \cdots & \frac{\partial^2 l(\boldsymbol{\beta},\boldsymbol{y})}{\partial^2 \beta_p} \end{pmatrix} \in \mathbb{R}^{p \times p}$$

Remark 3.17 (Observed Fisher information matrix)
The negative of H is called the observed Fisher information matrix:

$$\mathcal{I}_{obs}(\boldsymbol{\beta}) = -H(\boldsymbol{\beta}, \boldsymbol{y}) = - \left(\frac{\partial^2 l(\boldsymbol{\beta}, \boldsymbol{y})}{\partial \beta_i \partial \beta_j} \right)_{i,j \in \{1,\ldots,p\}} \in \mathbb{R}^{p \times p} \tag{3.14}$$

Definition 3.18 (Unscaled Fisher information matrix in GLMs)
The unscaled Fisher information matrix (also called the expected Fisher information) is given by the expectation of the negative of the unscaled Hessian matrix:

$$\mathcal{I} := \mathcal{I}(\boldsymbol{\beta}) := \mathbb{E}\left[-H(\boldsymbol{\beta}, \boldsymbol{y})\right] = \mathbb{E}\left[\mathcal{I}_{obs}(\boldsymbol{\beta})\right] \in \mathbb{R}^{p \times p}$$

We now compute the component $\mathcal{I}_{ij}$ (i.e. the entry in the i-th row and the j-th column of the unscaled Fisher information matrix $\mathcal{I}$) as in McCullagh and Nelder (1983)(p. 32). Therefore, we first compute one entry of the unscaled Hessian matrix.

$$
\begin{aligned}
H_{ij} &= \frac{\partial^2 l(\boldsymbol{\beta}, \boldsymbol{y})}{\partial \beta_i \partial \beta_j} = \frac{\partial s_i(\boldsymbol{\beta}, \boldsymbol{y})}{\partial \beta_j} \overset{Def.\ 3.15}{=} \frac{\partial}{\partial \beta_j} \left[\sum_{z=1}^{n} (y_z - \mu_z) W_z(\boldsymbol{\beta}) \left(\frac{\partial \eta_z}{\partial \mu_z} \right) x_{zi} \right] \\
&= \sum_{z=1}^{n} (y_z - \mu_z) \underbrace{\frac{\partial}{\partial \beta_j} \left[W_z(\boldsymbol{\beta}) \left(\frac{\partial \eta_z}{\partial \mu_z} \right) x_{zi} \right]}_{= \frac{\partial}{\partial \beta_j} \left[\frac{1}{b''(\theta_z)} \left(\frac{\partial \mu_z}{\partial \eta_z} \right) x_{zi} \right]} + \sum_{z=1}^{n} \underbrace{\frac{\partial}{\partial \beta_j} [(y_z - \mu_z)]}_{= -\left(\frac{\partial \mu_z}{\partial \eta_z} \right) x_{zj}} \underbrace{W_z(\boldsymbol{\beta}) \left(\frac{\partial \eta_z}{\partial \mu_z} \right) x_{zi}}_{= \frac{1}{b''(\theta_z)} \left(\frac{\partial \mu_z}{\partial \eta_z} \right) x_{zi}} \\
&= \sum_{z=1}^{n} (y_z - \mu_z) \frac{\partial}{\partial \beta_j} \left[\frac{1}{b''(\theta_z)} \left(\frac{\partial \mu_z}{\partial \eta_z} \right) x_{zi} \right] + \sum_{z=1}^{n} -\left(\frac{\partial \mu_z}{\partial \eta_z} \right) x_{zj} \frac{1}{b''(\theta_z)} \left(\frac{\partial \mu_z}{\partial \eta_z} \right) x_{zi} \\
&= \sum_{z=1}^{n} (y_z - \mu_z) \frac{\partial}{\partial \beta_j} \left[\frac{1}{b''(\theta_z)} \left(\frac{\partial \mu_z}{\partial \eta_z} \right) x_{zj} \right] - \sum_{z=1}^{n} \frac{1}{b''(\theta_z)} \left(\frac{\partial \mu_z}{\partial \eta_z} \right)^2 x_{zi} x_{zi} \qquad (3.15)
\end{aligned}
$$

Now we have to take the expectation of the negative of Equation (3.15) to get the corresponding entry of the unscaled Fisher information matrix:

$$
\mathcal{I}_{ij} = \mathbb{E}\left[-H(\boldsymbol{\beta}, \boldsymbol{y})\right]_{ij} = \mathbb{E}\left[-\frac{\partial s_i(\boldsymbol{\beta}, \boldsymbol{y})}{\partial \beta_j} \right]
$$

$$
\overset{\mathbb{E}[Y_z]=\mu_z}{=} \sum_{z=1}^{n} \frac{1}{b''(\theta_z)} \left(\frac{\partial \mu_z}{\partial \eta_z} \right)^2 x_{zi} x_{zj} = \sum_{z=1}^{n} W_z x_{zi} x_{zj} \qquad (3.16)
$$

We define:

$$
W := W(\boldsymbol{\beta}) \overset{Def.\ 3.12}{:=} diag(W_1(\boldsymbol{\beta}), \ldots, W_n(\boldsymbol{\beta})) \qquad (3.17)
$$

$$
= diag(W_1, \ldots, W_n) = \begin{pmatrix} W_1 & & & & \\ & W_2 & & \text{\Large 0} & \\ & & \ddots & & \\ \text{\Large 0} & & & W_{n-1} & \\ & & & & W_n \end{pmatrix}
$$

Using this very advantageous notation we can rewrite the unscaled Fisher information matrix as (see also Fahrmeir and Tutz (2001)(p. 41)):

$$
\mathcal{I}(\boldsymbol{\beta}) = X^\top W(\boldsymbol{\beta}) X
$$

Remark 3.19

Here $X \in \mathbb{R}^{n \times p}$ is defined as in Definition 2.3. We had:

$$
X = \begin{pmatrix} \underline{\quad} & \boldsymbol{x}_1^\top & \underline{\quad} \\ \underline{\quad} & \boldsymbol{x}_2^\top & \underline{\quad} \\ & \vdots & \\ \underline{\quad} & \boldsymbol{x}_n^\top & \underline{\quad} \end{pmatrix} \Rightarrow X^\top = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n) \in \mathbb{R}^{p \times n}
$$

Remark 3.20 (Finding stationary points using the Newton algorithm)
A stationary point of a function $f : \mathbb{R}^n \to \mathbb{R}$ ($f \in C^2$) can be found using the so-called
Newton algorithm. The algorithm seeks for a stationary point x^* with $\nabla f(x^*) = 0$ by
using the following iterative scheme (see Ulbrich and Ulbrich (2012)(algorithm 10.6)):

$$x^{n+1} = x^n - \underbrace{\nabla^2 f(x^n)^{-1}}_{\textit{inverse of the Hessian matrix}} \nabla f(x^n)$$

Our function to be optimized (i.e. maximized) is the log likelihood function, i.e.

$$f = l(\boldsymbol{\beta}, \phi \mid \boldsymbol{y})$$
$$\Rightarrow \nabla f = s(\boldsymbol{\beta}, \boldsymbol{y})$$
$$\Rightarrow \nabla^2 f = H(f) \stackrel{Def.\ 3.16}{=} H(\boldsymbol{\beta}, \boldsymbol{y})$$

Therefore, we would imagine that an iterative scheme such as:

$$\boldsymbol{\beta}^{n+1} = \boldsymbol{\beta}^n - H^{-1}(\boldsymbol{\beta}^n, \boldsymbol{y})s(\boldsymbol{\beta}^n)$$

would suffice to find the maximum likelihood estimate $\widehat{\boldsymbol{\beta}}$. However, we have to face the
following problems:

- only for canonical links the expected Fisher matrix is equal to the observed Fisher
  information matrix. We would like to avoid dependence on the data.
  $\Rightarrow$ We take the expected Fisher information matrix in our further calculations.
  Another advantage is that it is easier to evaluate and always positive semi-definite
  (see Fahrmeir and Tutz (2001)(p. 42)).

- the Newton algorithm opens up a whole theory about convergence and starting val-
  ues (i.e. one could also find a minimum using this algorithm, no global convergence).
  $\Rightarrow$ Not regarded, but the ambitious reader can read more about it in Ulbrich and
  Ulbrich (2012).

- MLEs might not exist and one has to convince oneself that it is (in case of existence)
  unique.
  $\Rightarrow$ Not regarded, but the ambitious reader can read more about it in Fahrmeir and
  Tutz (2001)(p. 43ff.) and the literature given there.

The algorithm, which is derived out of these considerations is the so-called *Fisher
scoring algorithm* (compare to Fahrmeir and Tutz (2001)(p. 39ff.)). The result $\widehat{\boldsymbol{\beta}}$ may
not be a global maximum of $l(\boldsymbol{\beta}, \phi \mid \boldsymbol{y})$ but it is a solution to $s(\boldsymbol{\beta}, \boldsymbol{y}) = \boldsymbol{0}$.

Definition 3.21 (Fisher scoring algorithm)
  (i) Choose initial values $\boldsymbol{\beta^0}$ (i.e. an initial estimate $\widehat{\boldsymbol{\beta^0}}$) and accuracy $\varepsilon \in \mathbb{R}^+$. Let us
     denote by $\boldsymbol{\beta^r}$ the current estimates of $\boldsymbol{\beta}$.

 (ii) while $\|\boldsymbol{\beta^r} - \boldsymbol{\beta^{r+1}}\| \geq \varepsilon$ do: for each step $r \in \mathbb{N}_0$

$$\boldsymbol{\beta^{r+1}} := \boldsymbol{\beta^r} + \mathcal{I}^{-1}(\boldsymbol{\beta^r})s(\boldsymbol{\beta^r}, \boldsymbol{y}) \tag{3.18}$$

(iii) Set $\widehat{\boldsymbol{\beta}} := \boldsymbol{\beta^{r+1}}$

Remark 3.22 (Rewrite the Fisher scoring algorithm)
The Fisher scoring algorithm determines the MLE $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. There is however another way to rewrite this algorithm using iteratively weighted least squares. An advantage is, that we can use statistical software to estimate the regression parameter vector $\boldsymbol{\beta}$ (see Hardin and Hilbe (2007)(p. 29)). Hence, we will now derive the iterative weighted least squares algorithm based on the Fisher scoring algorithm.

The iteration in Equation (3.18) can be written as

$$\boldsymbol{\beta^{r+1}} = \boldsymbol{\beta^r} + \mathcal{I}^{-1}(\boldsymbol{\beta^r})s(\boldsymbol{\beta^r}, \boldsymbol{y})$$
$$\Leftrightarrow \mathcal{I}(\boldsymbol{\beta^r})\boldsymbol{\beta^{r+1}} = \mathcal{I}(\boldsymbol{\beta^r})\boldsymbol{\beta^r} + s(\boldsymbol{\beta^r}, \boldsymbol{y}) \tag{3.19}$$

Therefore, we calculate the j -th element of the right hand side in Equation (3.19):

$$(\mathcal{I}(\boldsymbol{\beta^r})\boldsymbol{\beta^r} + s(\boldsymbol{\beta^r}, \boldsymbol{y}))_j$$

$$= \left(\begin{pmatrix} \mathcal{I}_{11}(\boldsymbol{\beta^r}) & \mathcal{I}_{12}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{1p}(\boldsymbol{\beta^r}) \\ \mathcal{I}_{21}(\boldsymbol{\beta^r}) & \mathcal{I}_{22}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{2p}(\boldsymbol{\beta^r}) \\ \vdots & & \ddots & \vdots \\ \mathcal{I}_{p1}(\boldsymbol{\beta^r}) & \mathcal{I}_{p2}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{pp}(\boldsymbol{\beta^r}) \end{pmatrix} (\beta_1^r, \dots, \beta_p^r)^\top + (s_1(\boldsymbol{\beta^r}, \boldsymbol{y}), \dots, s_p(\boldsymbol{\beta^r}, \boldsymbol{y}))^\top \right)_j$$

$$= \left(\begin{pmatrix} \mathcal{I}_{11}(\boldsymbol{\beta^r}) & \mathcal{I}_{12}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{1p}(\boldsymbol{\beta^r}) \\ \mathcal{I}_{21}(\boldsymbol{\beta^r}) & \mathcal{I}_{22}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{2p}(\boldsymbol{\beta^r}) \\ \vdots & & \ddots & \vdots \\ \mathcal{I}_{p1}(\boldsymbol{\beta^r}) & \mathcal{I}_{p2}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{pp}(\boldsymbol{\beta^r}) \end{pmatrix} \begin{pmatrix} \beta_1^r \\ \beta_2^r \\ \vdots \\ \beta_p^r \end{pmatrix} + \begin{pmatrix} s_1(\boldsymbol{\beta^r}, \boldsymbol{y}) \\ s_2(\boldsymbol{\beta^r}, \boldsymbol{y}) \\ \vdots \\ s_p(\boldsymbol{\beta^r}, \boldsymbol{y}) \end{pmatrix} \right)_j$$

$$= \begin{pmatrix} \sum_{z=1}^p \mathcal{I}_{1z}(\boldsymbol{\beta^r})\beta_z^r + s_1(\boldsymbol{\beta^r}, \boldsymbol{y}) \\ \sum_{z=1}^p \mathcal{I}_{2z}(\boldsymbol{\beta^r})\beta_z^r + s_2(\boldsymbol{\beta^r}, \boldsymbol{y}) \\ \vdots \\ \sum_{z=1}^p \mathcal{I}_{pz}(\boldsymbol{\beta^r})\beta_z^r + s_p(\boldsymbol{\beta^r}, \boldsymbol{y}) \end{pmatrix}_j$$

$$= \sum_{z=1}^p \underbrace{\mathcal{I}_{jz}(\boldsymbol{\beta^r})}_{\substack{Eq.\ (3.16) \\ = \sum_{m=1}^n W_m(\boldsymbol{\beta^r})x_{mj}x_{mz}}} \beta_z^r + \underbrace{s_j(\boldsymbol{\beta^r}, \boldsymbol{y})}_{\substack{Def.\ 3.15 \\ = \sum_{i=1}^n (y_i - \mu_i^r)W_i(\boldsymbol{\beta^r})\left(\frac{\partial \eta_i^r}{\partial \mu_i^r}\right)x_{ij}}}$$

$$= \sum_{z=1}^p \sum_{m=1}^n W_m(\boldsymbol{\beta^r})x_{mj}x_{mz}\beta_z^r + \sum_{i=1}^n (y_i - \mu_i^r)W_i(\boldsymbol{\beta^r})\left(\frac{\partial \eta_i^r}{\partial \mu_i^r}\right)x_{ij}$$

$$= \sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij} \left[\underbrace{\sum_{z=1}^p x_{iz}\beta_z^r}_{\substack{= \boldsymbol{x}_i^\top \boldsymbol{\beta^r} \stackrel{Def.\ 3.4}{\underset{Rem.\ 3.9}{=}} \eta_i^r}} + (y_i - \underbrace{\mu_i^r}_{= g^{-1}(\eta_i^r)})\left(\frac{\partial \eta_i^r}{\partial \mu_i^r}\right) \right]$$

$$= \sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij} \underbrace{\left[\eta_i^r + (y_i - \mu_i^r)\left(\frac{\partial \eta_i^r}{\partial \mu_i^r}\right) \right]}_{:= Z_i^r} = \sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij}Z_i^r \tag{3.20}$$

Remark 3.23
The variable $Z_i^r := \eta_i^r + (y_i - \mu_i^r)\left(\frac{\partial \eta_i^r}{\partial \mu_i^r}\right)$ is called *adjusted dependent variable* (or "working observation vector" in Fahrmeir and Tutz (2001)(p. 42)).

Calculating the j -th element of the left hand side in Equation (3.19), we get:

$$
\left(\mathcal{I}(\boldsymbol{\beta^r})\boldsymbol{\beta^{r+1}}\right)_j
$$
$$
= \left(\begin{pmatrix} \mathcal{I}_{11}(\boldsymbol{\beta^r}) & \mathcal{I}_{12}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{1p}(\boldsymbol{\beta^r}) \\ \mathcal{I}_{21}(\boldsymbol{\beta^r}) & \mathcal{I}_{22}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{2p}(\boldsymbol{\beta^r}) \\ \vdots & & \ddots & \vdots \\ \mathcal{I}_{p1}(\boldsymbol{\beta^r}) & \mathcal{I}_{p2}(\boldsymbol{\beta^r}) & \dots & \mathcal{I}_{pp}(\boldsymbol{\beta^r}) \end{pmatrix} \begin{pmatrix} \beta_1^{r+1} \\ \beta_2^{r+1} \\ \vdots \\ \beta_p^{r+1} \end{pmatrix}\right)_j
$$
$$
= \begin{pmatrix} \sum_{z=1}^p \mathcal{I}_{1z}(\boldsymbol{\beta^r})\beta_z^{r+1} \\ \sum_{z=1}^p \mathcal{I}_{2z}(\boldsymbol{\beta^r})\beta_z^{r+1} \\ \vdots \\ \sum_{z=1}^p \mathcal{I}_{pz}(\boldsymbol{\beta^r})\beta_z^{r+1} \end{pmatrix}_j
$$
$$
= \sum_{z=1}^p \underbrace{\mathcal{I}_{jz}(\boldsymbol{\beta^r})}_{Eq.\ \stackrel{(3.16)}{=}\sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij}x_{iz}} \beta_z^{r+1} = \sum_{z=1}^p \sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij}x_{iz}\beta_z^{r+1}
$$
$$
= \sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij} \underbrace{\sum_{z=1}^p x_{iz}\beta_z^{r+1}}_{=\boldsymbol{x}_i^\top \boldsymbol{\beta^{r+1}} \stackrel{Def.\ 3.4}{\underset{Rem.\ 3.9}{=}} \eta_i^{r+1}} = \sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij}\eta_i^{r+1} \tag{3.21}
$$

If we combine Equation (3.20) and Equation (3.21), we get:

$$
\sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij}\eta_i^{r+1} = \sum_{i=1}^n W_i(\boldsymbol{\beta^r})x_{ij}Z_i^r \qquad \forall j \in \{1,\dots,p\}
$$

With $W(\boldsymbol{\beta^r})$ defined similarly to Equation (3.17), we can rewrite this equivalently in matrix-vector notation (compare to Fahrmeir and Tutz (2001)(p. 42)):

$$
X^\top \underbrace{W(\boldsymbol{\beta^r})}_{:=W^r} X\boldsymbol{\beta^{r+1}} = X^\top \underbrace{W(\boldsymbol{\beta^r})}_{:=W^r} \underbrace{\begin{pmatrix} Z_1^r \\ Z_2^r \\ \vdots \\ Z_n^r \end{pmatrix}}_{:=\boldsymbol{Z^r}}
$$
$$
\Leftrightarrow X^\top W^r X\boldsymbol{\beta^{r+1}} = X^\top W^r \boldsymbol{Z^r} \tag{3.22}
$$
$$
\Leftrightarrow \boldsymbol{\beta^{r+1}} = (X^\top W^r X)^{-1} X^\top W^r \boldsymbol{Z^r} \tag{3.23}
$$

Remark 3.24 (Advantages of the IWLS)
According to Fahrmeir and Tutz (2001)(p. 42), the advantage of the IWLS is that one can use results for the least squares estimation for the iteratively weighted least squares after adequate modifications.

Remark 3.25 (Origin of the name "iterative weighted least squares")
In Subsection 2.2.1 we derived the ordinary least squares estimation in linear models. Now we want to derive the concept of weighted least squares. The calculation is similar to the calculation we made for the ordinary case and can be found for instance in Myers et al. (2002)(p. 49ff.).

Definition 3.26 (Generalized least squares estimator of β)
We consider the model:

$$\boldsymbol{Z} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \text{with} \quad \boldsymbol{\varepsilon} \sim \mathcal{N}_n(0, W)$$

with $W \in \mathbb{R}^{n \times n}$ known, positive definite (i.e. only positive eigenvalues). Then, we get $\widehat{\boldsymbol{\beta}}$ by solving the so-called *generalized normal equation* (as seen in Remark 2.7 for the least squares estimator):

$$X^\top W^{-1} X \boldsymbol{\beta} = X^\top W^{-1} \boldsymbol{Z} \tag{3.24}$$
$$\Leftrightarrow \widehat{\boldsymbol{\beta}} = (X^\top W^{-1} X)^{-1} X^\top W^{-1} \boldsymbol{Z}$$

Here $\widehat{\boldsymbol{\beta}}$ depends on the *weights* (more precisely on the variance-covariance matrix W). Thus, $\widehat{\boldsymbol{\beta}}$ is called the weighted least squares estimator.

With this intuition we can understand that Equation (3.22) corresponds to the generalized normal equation (as given in Equation (3.24)). We examine the weighted least squares of the response $\boldsymbol{Z^r}$ with design matrix X and weights $(W^r)^{-1}$. Hence, it seems reasonable to speak of *iteratively weighted least squares (IWLS)*.

Finally, we can present the *iterative weighted least squares (IWLS)* algorithm, which is derived from the iterative Fisher scoring algorithm (see Definition 3.21). This routine is also used in the `glm` function in R in the default method (see `help(glm.fit)`).

Definition 3.27 (IWLS for estimation of β in GLMs)
(i) Choose initial values $\boldsymbol{\beta^0}$ and accuracy $\varepsilon \in \mathbb{R}^+$. Let us denote by $\boldsymbol{\beta^r}$ the current estimates of $\boldsymbol{\beta}$ (for $r \in \mathbb{N}_0$).

(ii) while $\|\boldsymbol{\beta^r} - \boldsymbol{\beta^{r+1}}\| \geq \varepsilon$ do: determine for each observation $i \in \{1, \dots, n\}$

- the current linear predictors: $\widehat{\eta_i^r} := \boldsymbol{x}_i^\top \boldsymbol{\beta^r}$

- the current fitted means: $\widehat{\mu_i^r} := g^{-1}(\widehat{\eta_i^r}) \overset{Rem.\ 3.5}{=} F(\widehat{\eta_i^r})$

- current canonical parameters: $\widehat{\theta_i^r} := h(\widehat{\mu_i^r})$

- adjusted dependent variables: $Z_i^r := \widehat{\eta_i^r} + (y_i - \widehat{\mu_i^r}) \left(\frac{d\eta_i}{d\mu_i} \Big|_{\mu_i = \widehat{\mu_i^r}} \right)$

- $W_i^r := \left[b''(\theta_i)|_{\theta_i = \widehat{\theta_i^r}} \left(\frac{d\eta_i}{d\mu_i} \Big|_{\mu_i = \widehat{\mu_i^r}} \right)^2 \right]^{-1}$

Regress Z_i^r on $\boldsymbol{x}_i$ (i.e. $x_{i1}, \dots, x_{ip}$) with weights $(W_i^r)^{-1}$ to obtain new estimates $\boldsymbol{\beta^{r+1}}$, i.e. derive $\boldsymbol{\beta^{r+1}}$ using (see Equation (3.23) and Remark 3.25):

$$\boldsymbol{\beta^{r+1}} = (X^\top W^r X)^{-1} X^\top W^r \boldsymbol{Z^r}$$

Remark 3.28 (Asymptotic normality of the MLE)
According to Fahrmeir and Tutz (2001)(p. 44), the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$ is asymptotically normally distributed. We have the following asymptotic behavior for a large number of observations (n large):

$$\widehat{\boldsymbol{\beta}} \sim \mathcal{N}_p(\boldsymbol{\beta}, \mathcal{I}^{-1}(\boldsymbol{\beta}))$$

Therefore, we also get an asymptotic result for the variance-covariance matrix of the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$:

$$\mathbb{C}ov\left(\widehat{\boldsymbol{\beta}}\right) \approx \mathcal{I}^{-1}(\boldsymbol{\beta})$$

Remark 3.29 (Derivation of the IWLS)
For the derivation of the IWLS in our notation we cited McCullagh and Nelder (1983) and Fahrmeir and Tutz (2001) for the most important steps. Similar derivations can also be found in other literature about GLMs, for example in Dobson (1999)(see Section 4.4). But as every book (or author, respectively) has its own notation, it is advisable to restrict oneself to only few resources.

3.3   Families and link functions

By the term family we refer to the distribution of the error term and the link function in the model. In R this is one of the arguments, which have to be specified in the `glm` function (see `help(glm)`). If we call a family without specifying the link function, then the default link will be used. The link functions that are already implemented in R can be seen in Figure 3.1. The default links are printed with blue letters. However, it is also possible to call the `glm` function with user-defined links (see Chapter 6). We need to choose the link according to the data we want to examine. The following diagram may help (see Fahrmeir and Tutz (2001) and `help(family)` in R):
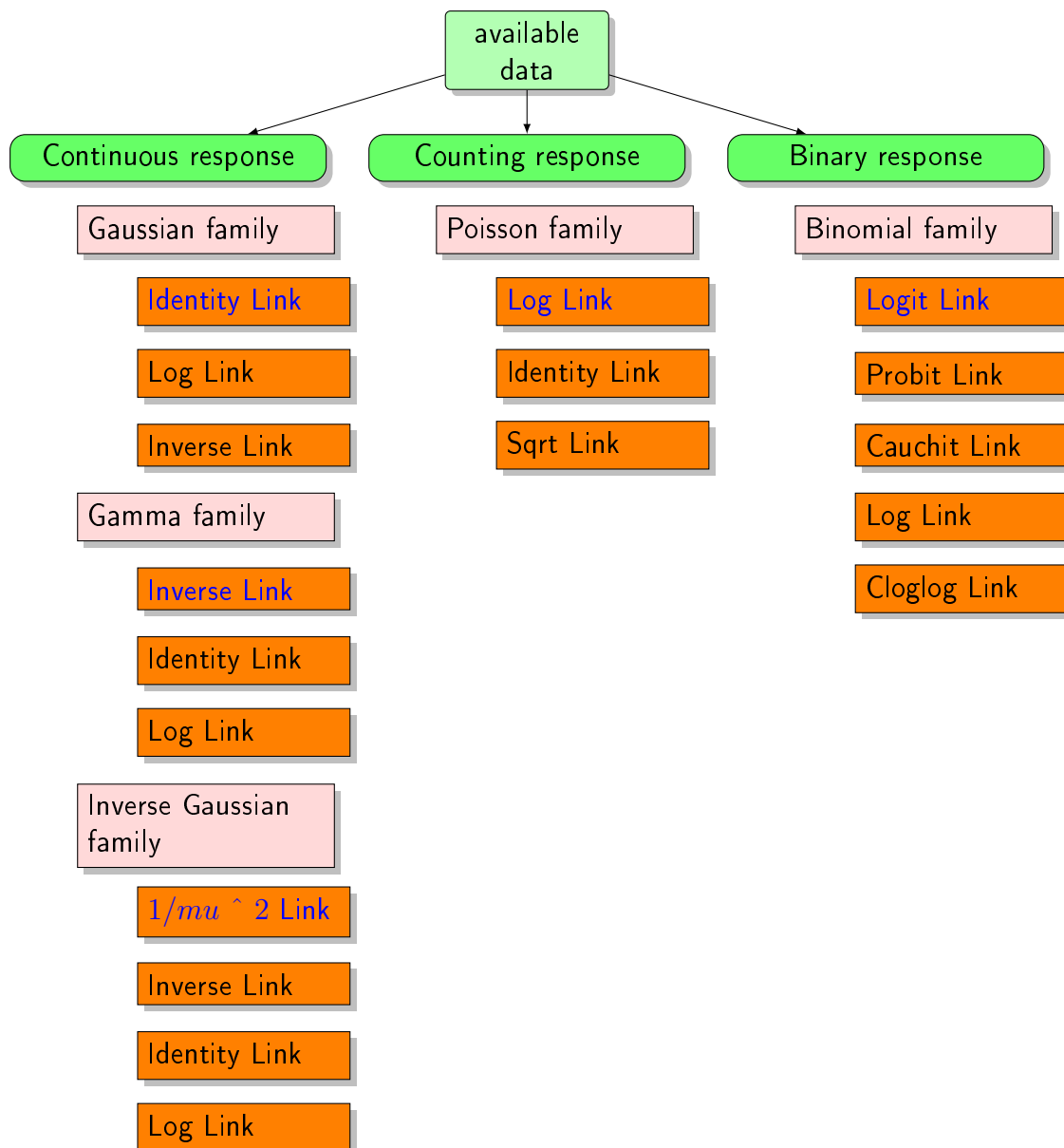

Figure 3.1: Families implemented in R (quasi families are missing) with the names of accepted link functions. The default links are printed with blue letters.

Remark 3.30
In the following we will not examine every possible link function and every possible family. In particular, we will not focus on the inverse Gaussian family. All the other families will be discussed with their corresponding default link functions. For the binomial distribution the probit link function will also be discussed.

For a link function in R we need the following specifications (see `help(make.link)`):

(i) `linkfun`: the link function, i.e. a function of the parameter μ.

(ii) `linkinv`: the inverse link function, i.e. a function of the parameter η.

(iii) `mu.eta`: the derivative $\left(\frac{d\mu}{d\eta}\right)$, i.e. the first derivative of the linkinv function. It is a function depending on η.

(iv) `valideta`: a function of η which states "TRUE" if η is in the domain of linkinv.

(v) `name`: the name to be used for the link function.

3.3.1   Gaussian family

We consider the case of a normal distribution (i.e. we assume that the errors follow a normal distribution) and choose the identity link. This leads us to the simple linear regression model as introduced in Definition 2.1. Consequently, the linear predictor and the mean are equal (see Equation (2.4)). The relationship between the linear predictor and the mean is reflected by the identity link. It is the most common used link for the Gaussian family and thus it is set as the default link in R (i.e. `link` = "identity"). However, if we notice that a non-linear relationship seems more appropriate, we can also choose another link (see Fahrmeir and Tutz (2001)(p. 23)). The log link (i.e. $g(\mu) = \ln(\mu)$) or the inverse link (i.e. $g(\mu) = \frac{1}{\mu}$) are allowed (see also Figure 3.1). We would like to refer to this situation using the term *Gaussian regression model*.

Remark 3.31
- the mean of the Gaussian distribution satisfies $\mu \in \mathbb{R}$. Hence, we have $\mathcal{G} = \mathbb{R}$.

- since the link is the identity we have $\eta \in \mathbb{R}$ and consequently we have $\mathcal{H} = \mathbb{R}$. This is also the restriction encoded through `valideta` in the link function.

Example 3.32 (Gaussian family (`link` = "identity"))
In R the identity link is defined through:

```
make.link("identity")

## $linkfun
## function (mu)
## mu
## <environment: namespace:stats>
##
## $linkinv
```

```
## function (eta)
## eta
## <environment: namespace:stats>
##
## $mu.eta
## function (eta)
## rep.int(1, length(eta))
## <environment: namespace:stats>
##
## $valideta
## function (eta)
## TRUE
## <environment: namespace:stats>
##
## $name
## [1] "identity"
##
## attr(,"class")
## [1] "link-glm"
```

Therefore, the identity link function in a Gaussian regression model is defined through:

(i) `linkfun`: $\eta = g(\mu) = \mu$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = \eta$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = 1$

(iv) `valideta`: $\mathbb{1}_{\{\eta \in \mathbb{R}\}} = \text{TRUE}$

3.3.2   Binomial family

Let us consider binomial responses. According to Fahrmeir and Tutz (2001)(p. 24), the setting of a *binomial regression* is the following.

Definition 3.33 (Binomial regression model)
Consider we are given the data for n observations i.e. for $i \in \{1, \ldots n\}$ we are given the realizations y_i of the responses and the values of the known covariates $\boldsymbol{x}_i$. Recall that the y_i are realizations from the random variable Y_i, where $Y_1, \ldots, Y_n$ are independent (see Definition 3.4). Since the responses are binary, they can only take values in $\{0,1\}$, i.e. $\forall i \in \{1, \ldots n\}$ we have $Y_i = 0$ or $Y_i = 1$. Therefore, we can determine the binary variable completely by its success probability. Given a covariate vector $\boldsymbol{x}_i$, the success probability is defined through:
$$p_i := p(\boldsymbol{x}_i) := \mathbb{P}\left(Y_i = 1 \mid \boldsymbol{x}_i\right) = \mathbb{E}\left[Y_i \mid \boldsymbol{x}_i\right]$$

Remark 3.34
- the success probability must fulfill the constraint : $p(\boldsymbol{x_i}) \in [0, 1] \quad \forall i \in \{1, \ldots n\}$.

- please notice that for a binary random variable the expectation is the success probability, i.e. we have $p_i = \mathbb{E}[Y_i] =: \mu_i$.

In the following we want to restrict ourselves to two important models for the success probability: the *logistic regression model* and the *probit regression model* as given in Fahrmeir and Tutz (2001)(p. 24ff.).

Definition 3.35 (Logistic regression model)
In the logistic regression model we take:

$$p(\boldsymbol{x}_i) = \mathbb{P}\left(Y_i = 1 \mid \boldsymbol{x}_i\right) = \frac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}$$

By replacing through known quantities we get:

$$\mu_i = p_i := p(\boldsymbol{x}_i) = F(\eta_i) = \frac{\exp\{\eta_i\}}{1 + \exp\{\eta_i\}}$$

Definition 3.36 (Probit regression model)
In the probit regression model we take:

$$p(\boldsymbol{x}_i) = \mathbb{P}\left(Y_i = 1 \mid \boldsymbol{x}_i\right) \stackrel{Rem. A.3}{=} \Phi(\boldsymbol{x}_i^\top \boldsymbol{\beta})$$

By replacing through known quantities we get:

$$\mu_i = p_i := p(\boldsymbol{x}_i) = F(\eta_i) = \Phi(\eta_i)$$

Remark 3.37
As described in Fahrmeir and Tutz (2001)(p. 25), we do usually consider scaled binomial responses when examining binomial responses. I.e. we consider $Y_i^* := \frac{Y_i}{n_i}$ as responses (for $i \in \{1, \ldots, n\}$). For the distribution of Y_i^* we introduce the term "scaled binomial distribution". In the following we will only refer to GLMs with scaled binomial responses and thus we introduce the following notation.

Definition 3.38 (Scaled binomial distribution)
For $Y \sim Bin(n, p)$ we say $Y^* := \frac{Y}{n} \sim ScaledBin(n, p)$ follows a *scaled binomial distribution*. For the ordinary binomial distribution we can take values $k \in \{0, 1, \ldots, n\}$, while for $Y^* \sim ScaledBin(n, p)$ we have $k^* := \frac{k}{n} \in \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$). In Subsection A.1.2 we show that the scaled binomial distribution is a member of the exponential family.

For $Y^* \sim ScaledBin(n, p)$ the expectation is $\mathbb{E}[Y^*] = \mu = p \in (0, 1)$. Therefore, we seek for a link function $g : (0, 1) \to \mathbb{R}$. Using the Definitions 3.35 and 3.36 and $g(\cdot) = F^{-1}(\cdot)$ we can see that such functions are given by:

- $g(\mu_i) = \ln\left(\frac{\mu_i}{1-\mu_i}\right)$ (inverse of the distribution function of the logistic distribution)

- $g(\mu_i) = \Phi^{-1}(\mu_i)$ (inverse of the distribution function of the standard normal distribution)

If we take the inverse of the distribution function of the logistic distribution as link function, we speak of the *logistic regression*. Likewise, if we take the inverse of the distribution function of the standard normal distribution, we speak of the *probit regression*.

Remark 3.39

- we have for the expectation $\mu \in (0, 1)$. Therefore, we choose $\mathcal{G} = (0, 1)$.

- since the link is either logit or probit we receive $\eta \in \mathbb{R}$ as restriction. Hence, we have $\mathcal{H} = \mathbb{R}$. This is also the restriction encoded through `valideta` in the link function.

Example 3.40 (Binomial family (`link` = "logit"))
In R the logit link is defined through:

```
make.link("logit")

## $linkfun
## function (mu)
## .Call(C_logit_link, mu)
## <environment: namespace:stats>
##
## $linkinv
## function (eta)
## .Call(C_logit_linkinv, eta)
## <environment: namespace:stats>
##
## $mu.eta
## function (eta)
## .Call(C_logit_mu_eta, eta)
## <environment: namespace:stats>
##
## $valideta
## function (eta)
## TRUE
## <environment: namespace:stats>
##
## $name
## [1] "logit"
##
## attr(,"class")
## [1] "link-glm"
```

We made the efforts to see how these C-code functions are defined:

- `linkfun`: $.Call(C\_logit\_link, mu)$

- `linkinv`: $.Call(C\_logit\_linkinv, eta)$

- `mu.eta`: $.Call(C\_logit\_mu\_eta, eta)$

Remark 3.41 (Assessing C-Code in R)
Since the code is written in the programming language C, we don't have access to it directly. We are also not able to view it with R without further ado. We are following Ligges (2006) to assess the underlying C-code. Especially the section "Compiled Code Sources" is of interest. Therefore we proceed taking the following steps:

(i) we download the *R source bundle* from the CRAN mirror (e.g. GWDG Goettingen under `http://ftp5.gwdg.de/pub/misc/cran/src/base/R-3/R-3.1.0.tar.gz`). It is important to download the R source bundle, since the source files are not included in the binary version of R, nor in the included packages. This way we can examine the original sources R has been installed from.

(ii) we receive a file ending with "....tar.gz". This file is compressed twice. If you have unpacked it entirely, you can find the source code under

$$\text{".../src/library/stats/src/family.c"}$$

(e.g. if we download "R-3.1.0.tar.gz" we can find the C source file "family" under "R-3.1.0/src/library/stats/src" in the decompressed folder). For other source code in different packages or package bundles we can find the code under "Package-Name/src/" or "BundleName/PackageName/src/".

Therefore, we can define the logit link function in a binomial regression model through:

(i) `linkfun`: $\eta = g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = \frac{\exp\{\eta\}}{1+\exp\{\eta\}}$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = \frac{\exp\{\eta\}}{(1+\exp\{\eta\})^2}$

(iv) `valideta`: $\mathbb{1}_{\{\eta\in\mathbb{R}\}} = \text{TRUE}$

Example 3.42 (Binomial family (`link` = "probit"))
In R the probit link is defined through:

```
make.link("probit")

## $linkfun
## function (mu)
## qnorm(mu)
## <environment: namespace:stats>
##
## $linkinv
## function (eta)
## {
##      thresh <- -qnorm(.Machine$double.eps)
##      eta <- pmin(pmax(eta, -thresh), thresh)
##      pnorm(eta)
## }
## <environment: namespace:stats>
##
## $mu.eta
## function (eta)
## pmax(dnorm(eta), .Machine$double.eps)
## <environment: namespace:stats>
##
## $valideta
## function (eta)
## TRUE
## <environment: namespace:stats>
##
## $name
## [1] "probit"
##
## attr(,"class")
## [1] "link-glm"
```

Therefore, the probit link function in a binomial regression model is defined through:

(i) `linkfun`: $\eta = g(\mu) = qnorm(\mu) \overset{quantile\ function}{=\joinrel=} \Phi^{-1}(\mu)$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = pnorm(\eta) \overset{distr.\ func.\ see}{\underset{Rem.\ A.3}{=\joinrel=}} \Phi(\eta)$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = dnorm(\eta) \overset{density\ see}{\underset{Rem.\ A.3}{=\joinrel=}} f(\eta \mid 0, 1) := \varphi(\eta)$

(iv) `valideta`: $\mathbb{1}_{\{\eta \in \mathbb{R}\}} = \text{TRUE}$

3.3.3 Poisson family

As explained in Fahrmeir and Tutz (2001)(p. 36) we can use the Poisson distribution to model count data (i.e. the number of events occurring in a fixed time period). Hence, we have a discrete and non-negative response with values in $\mathbb{N}_0$. We expect $\mathbb{E}[Y] = \lambda = \mu > 0$ (see Remark A.7). In R the default link is the log link. Two other possible links are the identity link (i.e. $g(\mu) = \mu$) and the sqrt link (i.e. $g(\mu) = \sqrt{\mu}$) (see Figure 3.1).

Definition 3.43 (Poisson regression model)
We want to refer to the following setting using the term *Poisson regression model*. Assume we want to model count data and take the Poisson family with the log link. Then, we have:

$$\eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} = g(\mu_i) = \ln(\mu)$$
$$\mu_i = F(\eta_i) = \exp\{\eta_i\}$$

Remark 3.44
- the mean of the Poisson distribution fulfills $\mu = \lambda \in \mathbb{R}^+$. Hence, we have $\mathcal{G} = \mathbb{R}^+$.

- since we take the log link we have $\eta \in \mathbb{R}$ (the domain of $\exp(\cdot)$ is $\mathbb{R}$) and consequently we have $\mathcal{H} = \mathbb{R}$. This is also the restriction encoded through `valideta` in the link function.

Example 3.45 (Poisson family (`link` = "log"))
In R the log link is defined through:

```
make.link("log")

## $linkfun
## function (mu)
## log(mu)
## <environment: namespace:stats>
##
## $linkinv
## function (eta)
## pmax(exp(eta), .Machine$double.eps)
## <environment: namespace:stats>
##
## $mu.eta
## function (eta)
## pmax(exp(eta), .Machine$double.eps)
## <environment: namespace:stats>
##
## $valideta
## function (eta)
## TRUE
## <environment: namespace:stats>
##
```

```
## $name
## [1] "log"
##
## attr(,"class")
## [1] "link-glm"
```

Therefore, the log link function in a Poisson regression model is defined through:

(i) `linkfun`: $\eta = g(\mu) = \ln(\mu)$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = \exp\{\eta\}$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = \exp\{\eta\}$

(iv) `valideta`: $\mathbb{1}_{\{\eta \in \mathbb{R}\}} = \text{TRUE}$

3.3.4   Gamma family

As described in Fahrmeir and Tutz (2001)(p. 23) we can use the gamma distribution for continuous and non-negative responses. Hence, we expect $\mathbb{E}[Y] = \mu > 0$ and thus the shape parameter ν is positive (i.e. $\nu > 0$). This can also be derived from Remark A.8. For instance, data sets about insurance claims or the amount of rainfall would fit in the setting of a gamma regression. In R the default link is the inverse link. Also the log link (i.e. $g(\mu) = \ln(\mu)$) and the identity link (i.e. $g(\mu) = \mu$) are allowed (see Figure 3.1).

Definition 3.46 (Gamma regression model)
We want to refer to the following setting using the term *gamma regression model*. Assume we model a continuous and non-negative response taking the gamma family with the inverse link. Then, we have:

$$\mu_i = F(\eta_i) = \frac{1}{\eta_i}$$
$$\Rightarrow \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} = g(\mu_i) = \frac{1}{\mu_i}$$

Remark 3.47
- the expectation of the gamma distribution is positive (i.e. $\mu \in \mathbb{R}^+$). Therefore, we choose $\mathcal{G} = \mathbb{R}^+$.

- since the link is the reciprocal we have $\eta \neq 0$ as restriction. Hence, we have $\mathcal{H} = \mathbb{R} \setminus \{0\}$. This is also the restriction encoded through `valideta` in the link function.

Example 3.48 (Gamma family (`link` = "inverse"))
In R the inverse link is defined through:

```r
make.link("inverse")

## $linkfun
## function (mu)
## 1/mu
## <environment: namespace:stats>
##
## $linkinv
## function (eta)
## 1/eta
## <environment: namespace:stats>
##
## $mu.eta
## function (eta)
## -1/(eta^2)
## <environment: namespace:stats>
##
## $valideta
## function (eta)
## all(is.finite(eta)) && all(eta != 0)
## <environment: namespace:stats>
##
## $name
## [1] "inverse"
##
## attr(,"class")
## [1] "link-glm"
```

Therefore, the inverse link function in a gamma regression model is defined through:

(i) `linkfun`: $\eta = g(\mu) = \frac{1}{\mu}$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = \frac{1}{\eta}$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = -\frac{1}{\eta^2}$

(iv) `valideta`: $\mathbb{1}_{\{\{\eta \in \mathbb{R}\} \cap \{\eta \neq 0\}\}}$

3.4   Goodness of fit of a generalized linear model

Assume we have chosen a family with a suitable link function in a generalized linear model for our response. Now we would like to assess how good the GLM of choice fits to the given data. We will follow McCullagh and Nelder (1983)(p. 24ff.), to introduce the *deviance* as a measure for the goodness of fit.

Definition 3.49 (Fitted mean)
For one observation ($i \in \{1, \ldots, n\}$) we are able to estimate the mean μ_i of Y_i by (using Equation (3.3) and the link function g as defined in Equation (3.4)):

$$\widehat{\mu}_i = g^{-1}(\boldsymbol{x}_i^\top \widehat{\boldsymbol{\beta}})$$

If our model is good, we would expect, that $\|\widehat{\boldsymbol{\mu}} - \boldsymbol{y}\|_2$ is small (i.e. there is not much discrepancy and the vector of the fitted means $\widehat{\boldsymbol{\mu}}$ is close to the vector of observations $\boldsymbol{y}$). In the following we want to derive a method to measure this discrepancy. Therefore we introduce a notation to describe how many parameters or covariates, respectively (since p = k+1) our model should contain:

Definition 3.50 (Null model and saturated model)
We have to decide how many parameters our model should contain. Given n observations $y_1, \ldots, y_n$, we could fit models containing between 1 and n parameters (i.e. $p \in \{1, \ldots, n\}$). Therefore, we will have two extreme models:

- **the null model**
  the null model is the simplest model. It does not contain any covariates at all and it consists of only one parameter: β_0. Using Equation (2.4) we obtain:

  $$\mu_i = \mathbb{E}\left[Y_i\right] = \beta_0$$

  Therefore, this model implies that the responses $Y_1, \ldots, Y_n$ have a common mean.

- **the saturated model**
  the saturated model (also called full or maximal model) is the largest well defined model for n responses. In this model n parameters are included (one for each observation). We have $k = n - 1$ covariates and with Equation (2.4) we get:

  $$\widehat{\mu}_i = \mathbb{E}\left[Y_i\right] = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} \stackrel{Eq.\ (2.1)}{=} y_i \tag{3.25}$$

  Therefore, the mean fits perfectly on the data (i.e. no discrepancy).

Remark 3.51 (Reasonable GLMs)
Any informative and acceptable GLM will range between the null model and the saturated model. The null model is considered being too simple while the saturated model only repeats information about the given data.

Since we derived the log likelihood for GLMs in Equation (3.10), we can use it to assess the goodness of fit for a model with p parameters. We rewrite the log likelihood in terms of the mean vector $\boldsymbol{\mu}$ instead of the vector of canonical parameters $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_n)^\top$.

Definition 3.52 (Mean parameterization of the log likelihood)
We can rewrite the log likelihood function in terms of the mean vector $\boldsymbol{\mu}$.

$$l(\boldsymbol{\beta}, \phi \mid \boldsymbol{y}) = \sum_{i=1}^{n} \left(\frac{\theta_i y_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi) \right)$$

$$\stackrel{see\ Eq.\ (3.7)}{=} \sum_{i=1}^{n} \left(\frac{h(\mu_i) y_i - b(h(\mu_i))}{a(\phi)} + c(y_i, \phi) \right) := l(\boldsymbol{\mu}, \phi \mid \boldsymbol{y})$$

This is called the *mean parameterization of the log likelihood*.

Definition 3.53 (Scaled deviance $D_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi)$)
Let us denote by $\widehat{\boldsymbol{\mu}} := (\widehat{\mu_1}, \ldots, \widehat{\mu_n})^\top$ the vector of fitted means. Further we will denote by $l(\widehat{\boldsymbol{\beta}}_{max}, \phi \mid \boldsymbol{y})$ the maximized log likelihood of the saturated model while $l(\widehat{\boldsymbol{\beta}}, \phi \mid \boldsymbol{y})$ denotes the maximized log likelihood for the model of interest. According to Wood (2006)(p. 70) the scaled deviance $D_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi)$ is then given by:

$$
\begin{aligned}
D_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi) :=& 2\left[l(\widehat{\boldsymbol{\beta}}_{max}, \phi \mid \boldsymbol{y}) - l(\widehat{\boldsymbol{\beta}}, \phi \mid \boldsymbol{y}) \right] \\
\overset{see\ Eq.\ (3.25)}{=}& 2\big[\quad \underbrace{l(\boldsymbol{y}, \phi \mid \boldsymbol{y}) - l(\widehat{\boldsymbol{\mu}}, \phi \mid \boldsymbol{y})}_{\overset{Def.\ 3.52}{=}\sum_{i=1}^{n}\left(\frac{h(y_i)y_i - b(h(y_i))}{a(\phi)} + c(y_i,\phi) - \frac{h(\widehat{\mu_i})y_i - b(h(\widehat{\mu_i}))}{a(\phi)} - c(y_i,\phi) \right)} \quad \big] \\
=& 2\sum_{i=1}^{n}\left(\frac{h(y_i)y_i - b(h(y_i)) - h(\widehat{\mu_i})y_i + b(h(\widehat{\mu_i}))}{a(\phi)} \right) \\
=& 2\sum_{i=1}^{n}\left(\frac{(\overbrace{h(y_i)}^{:=\tilde{\theta}_i} - \overbrace{h(\widehat{\mu_i})}^{:=\widehat{\theta}_i})y_i - b(h(y_i)) + b(h(\widehat{\mu_i}))}{a(\phi)} \right) \\
=& 2\sum_{i=1}^{n} \frac{(\tilde{\theta}_i - \widehat{\theta}_i)y_i - b(\tilde{\theta}_i) + b(\widehat{\theta}_i)}{a(\phi)}
\end{aligned}
\tag{3.26}
$$

Assuming that $a(\phi) = \frac{\phi}{\omega}$ (see Equation (3.2)), we can rewrite Equation (3.26) in an unscaled version. Often one refers to the unscaled deviance using the term *deviance*.

Definition 3.54 ((Unscaled) deviance $D(\widehat{\boldsymbol{\mu}}, \boldsymbol{y})$)
Let the scaled deviance be defined as in Equation (3.26):

$$
D_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi) := 2\sum_{i=1}^{n} \frac{(\tilde{\theta}_i - \widehat{\theta}_i)y_i - b(\tilde{\theta}_i) + b(\widehat{\theta}_i)}{a(\phi)} = 2\sum_{i=1}^{n} w_i \frac{(\tilde{\theta}_i - \widehat{\theta}_i)y_i - b(\tilde{\theta}_i) + b(\widehat{\theta}_i)}{\phi}
$$

Then, according to McCullagh and Nelder (1983)(p. 24) and Wood (2006)(p. 70), the *(unscaled) deviance* $D(\widehat{\boldsymbol{\mu}}, \boldsymbol{y})$ is given by:

$$
D(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}) := \phi D_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi) = 2\sum_{i=1}^{n} w_i \left[(\tilde{\theta}_i - \widehat{\theta}_i)y_i - b(\tilde{\theta}_i) + b(\widehat{\theta}_i) \right]
$$

The unscaled deviance is independent of ϕ.

Remark 3.55 (Distribution of the deviance)
- **distribution of the scaled deviance:**
  According to Wood (2006)(p. 70) we will have

$$
D_s(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi) \sim \chi^2_{n-p}
$$

  if the model is good (i.e. if it describes the data in a good way).

- **distribution of the unscaled deviance:**
  Following Fahrmeir and Tutz (2001)(p. 50f.), we can assume for a sufficiently large number of observations that:

$$D(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}, \phi) \sim \phi \chi^2_{n-p}$$

Remark 3.56
All in all we want to take the GLM delivering the minimal deviance, i.e. the minimal discrepancy between the fitted means $\widehat{\mu}_i$ and the observations of the response y_i.

Remark 3.57 (Calculation of the deviance)
The deviance of the distributions used throughout this thesis can be found in McCullagh and Nelder (1983)(p. 25). In R the unscaled deviance is calculated as the value for the deviance according to the common formulas (see also Czado et al. (2013)(p. 41f. and p. 49) and Wood (2006)(p. 61 and p. 70)). In Section A.3 we verify the calculation of the deviance for two examples of the Gaussian regression.

3.5 Overview and comments

The sections before lead to the following table:

Component	Notation	Gaussian Identity	Binomial Logit	Binomial Probit	Poisson Log	Gamma Inverse
linkfun	$g(\mu)$	μ	$\ln\left(\frac{\mu}{1-\mu}\right)$	$\Phi^{-1}(\mu)$	$\ln(\mu)$	$\frac{1}{\mu}$
linkinverse	$F(\eta)$	η	$\frac{\exp\{\eta\}}{1+\exp\{\eta\}}$	$\Phi(\eta)$	$\exp\{\eta\}$	$\frac{1}{\eta}$
mu.eta	$\frac{d\mu}{d\eta}(F(\eta))$	1	$\frac{\exp\{\eta\}}{(1+\exp\{\eta\})^2}$	$\varphi(\eta)$	$\exp\{\eta\}$	$-\frac{1}{\eta^2}$
valideta	$\mathbb{1}_{\{res.\ to\ \eta\}}$	$\mathbb{1}_{\{\eta\in\mathbb{R}\}}$	$\mathbb{1}_{\{\eta\in\mathbb{R}\}}$	$\mathbb{1}_{\{\eta\in\mathbb{R}\}}$	$\mathbb{1}_{\{\eta\in\mathbb{R}\}}$	$\mathbb{1}_{\{\{\eta\in\mathbb{R}\}\cap\{\eta\neq0\}\}}$

Table 3.3: Overview: common link functions with their components in R.

Remark 3.58
Further interesting tables are given in Hardin and Hilbe (2007)(Appendix A p. 356ff.).

Definition 3.59 (Canonical link function)
According to Fahrmeir and Tutz (2001)(p. 20), a link function $g(\cdot) = g(\mu_i) = g(\mathbb{E}\left[Y_i\right])$ is called *canonical* or *natural* if the following holds:

$$\theta = \theta(\mu_i) = \eta_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}$$
$$\Rightarrow g(\mathbb{E}\left[Y_i\right]) = g(\mu_i) = \theta_i = \theta(\mu_i) \qquad \forall\ i \in \{1, \ldots, n\}$$

Remark 3.60
Definition 3.59 is a restriction to the function $g(\cdot)$ since $\eta_i = g(\mu_i) = g(b'(\theta_i))$ (see Equation (3.4) and Theorem 3.6).

Example 3.61 (Canonical links)
According to McCullagh and Nelder (1983)(p. 24) we have:

- for the Gaussian regression model the identity link (*i.e.* $g(\mu) = \mu$) is canonical.

- for the binomial regression model the logit link $\left(i.e.\ g(\mu) = \ln \left(\frac{\mu}{1-\mu} \right) \right)$ is canonical.

- for the Poisson regression model the log link (*i.e.* $g(\mu) = \ln(\mu)$) is canonical.

- for the gamma regression model the inverse link $\left(i.e.\ g(\mu) = \frac{1}{\mu} \right)$ is canonical.

As mentioned in Remark 3.5 we can derive restrictions for the expectation μ (i.e. $\mu \in \mathcal{G}$) and restrictions for the linear predictor η (i.e. $\eta \in \mathcal{H}$). These restrictions are summarized in the following table:

Distr. of error	Res. for μ_i (by $\mathcal{G}$)	Linkinverse $F(\eta)$	Res. for η_i (by $\mathcal{H}$)
Normal	$\mathcal{G} = \mathbb{R}$	$F(\eta) = \eta$	$\mathcal{H} = \mathbb{R}$
Scaled binomial	$\mathcal{G} = (0,1)$	$F(\eta) = \frac{\exp\{\eta\}}{1+\exp\{\eta\}}$ $F(\eta) = \Phi(\eta)$	$\mathcal{H} = \mathbb{R}$ $\mathcal{H} = \mathbb{R}$
Poisson	$\mathcal{G} = \mathbb{R}^+$	$F(\eta) = \exp\{\eta\}$	$\mathcal{H} = \mathbb{R}$
Gamma	$\mathcal{G} = \mathbb{R}^+$	$F(\eta) = \frac{1}{\eta}$	$\mathcal{H} = \mathbb{R} \setminus \{0\}$

Table 3.4: Summary: restrictions for the common link functions.

4 Data sets

In this chapter we introduce the data sets, which we will examine throughout this thesis. Five of them were also studied in Czado (2007) (namely: `pcb.ex`, `beetle.ex`, `bys.ex`, `rotifier.ex` and `mining.ex`). We also provide an example of the gamma regression model by presenting data about the Canadian automobile experience as given in the data set `carinsur.ex`. This example can be found in Bailey and Simon (1960). Sometimes it is necessary to transform the covariates or the response to obtain a linear relationship.

4.1 Gaussian family

Example 4.1 (PCB concentration in lake trouts)
As in Bates and Watts (2007)(see p. 3 and p. 267), we examine the data gathered about 28 lake trouts of the Cayuga Lake in New York. The first column (i.e. *pcb*) in Table 4.1 contains the observed concentrations of PCB (= polychlorinated biphenyl) in the trouts in parts per million. We also have information about the age of the trouts as given in the third column *age*. The authors recommend using a simple linear regression model (i.e. `family` = "gaussian", `link` = "identity") with response $y_i = \ln(pcb[i])$ as given in the second column *log.pcb*. The covariate will be the centered age as given in the fourth column *age.cen* (i.e. $x_{1i} = x_i = age[i] - mean(age)$). Figure 4.1 illustrates the linear interaction between the centered age (covariate) and the logarithm of the PCB concentration (response).

pcb	log.pcb	age	age.cen	pcb	log.pcb	age	age.cen
0.6	-0.511	1	-4.536	3.4	1.224	6	0.464
1.6	0.470	1	-4.536	9.7	2.272	6	0.464
0.5	-0.693	1	-4.536	8.6	2.152	6	0.464
1.2	0.182	1	-4.536	4.0	1.386	7	1.464
2.0	0.693	2	-3.536	5.5	1.705	7	1.464
1.3	0.262	2	-3.536	10.5	2.351	7	1.464
2.5	0.916	2	-3.536	17.5	2.862	8	2.464
2.2	0.788	3	-2.536	13.4	2.595	8	2.464
2.4	0.875	3	-2.536	4.5	1.504	8	2.464
1.2	0.182	3	-2.536	30.4	3.414	9	3.464
3.5	1.253	4	-1.536	12.4	2.518	11	5.464
4.1	1.411	4	-1.536	13.4	2.595	12	6.464
5.1	1.629	4	-1.536	26.2	3.266	12	6.464
5.7	1.740	5	-0.536	7.4	2.001	12	6.464

Table 4.1: Concentration of PCB in lake trouts as given in the data frame `pcb.ex`.

In Example 5.1 the output of the ordinary `glm` function on this data set will be presented (using `family` = "gaussian", `link` = "identity"). We will also examine this data set using a user-defined link function. In Example 6.11 we will demonstrate the effect of a right tail modification using the `glm` function in R. The parameter for this modification is derived in Example 7.1 by using the `glmProfile` function.

4.2   Binomial family

Example 4.2 (Beetle mortality)

Bliss (1935)(see p. 154) reports two data series about beetles being exposed to the harmful gas CS_2 (= carbon disulphide, measured in $\frac{mg}{l}$) for five hours. Table 4.2 summarizes the two given data series in eight observations.

y_i	n_i	$log_{10}(CS_2\frac{mg}{l})$	dose.cen	y_i	n_i	$log_{10}(CS_2\frac{mg}{l})$	dose.cen
6	59	1.6907	-0.1027	52	63	1.8113	0.0179
13	60	1.7242	-0.0692	53	59	1.8369	0.0435
18	62	1.7552	-0.0382	61	62	1.8610	0.0676
28	56	1.7842	-0.0092	60	60	1.8839	0.0905

Table 4.2: Mortality of beetles being exposed to CS_2 as given in the data frame `beetle.ex`.

The number of beetles being found dead after the exposure are given by y_i in the first column. The second column gives the total number of insects n_i being exposed to the gas in this observation. The third column (i.e. $log_{10}(CS_2\frac{mg}{l})$) contains the logarithm (to the base ten) of the different concentrations of CS_2. Centering the covariate (i.e. $log_{10}(CS_2\frac{mg}{l})$) yields to *dose.cen* as given in the fourth column.
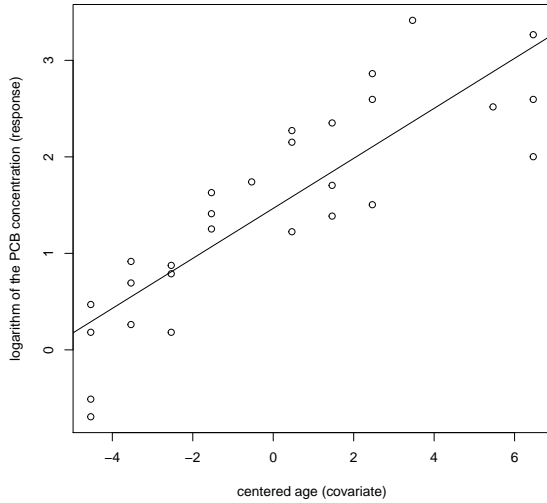
Figure 4.1: Plot of the covariate age.cen against the response log.pcb for the data given in `pcb.ex`.
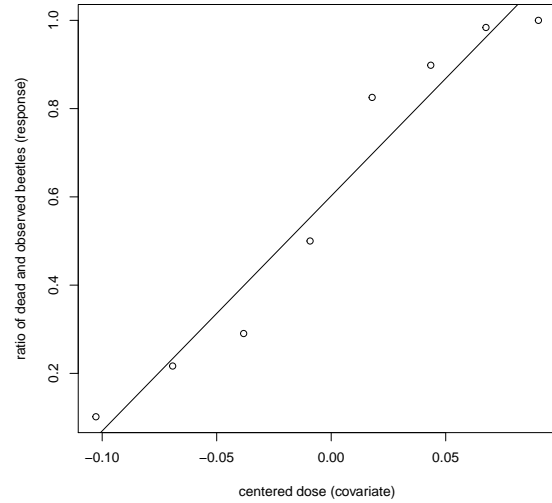
Figure 4.2: Plot of the covariate dose.cen against the response ratio $\frac{y_i}{n_i}$ for the data given in `beetle.ex`.

The number of dead beetles y_i correspond to the binomial family. In each observation ($i \in \{1, \ldots, 8\}$) each single beetle can be either dead or still alive after the exposure. Hence, the state of a single beetle corresponds to a Bernoulli distribution. Examining the state of n_i independent beetles yields to a binomial distribution. Considering the

distribution of $\frac{y_i}{n_i}$ for each observation corresponds to a scaled binomial distribution. Therefore, the response is $\frac{y_i}{n_i}$, i.e. the ratio of dead beetles and the number of observed beetles in the specific observation ($= \frac{\# \ of \ dead \ beetles \ in \ observation \ i}{\# \ beetles \ in \ observation \ i}$). Figure 4.2 shows the dependence between the response ratio and the centered dose of CS_2. We can see a linear interaction as indicated by the straight line.

In Example 5.2 the output of the ordinary `glm` function on this data set will be presented (using `family` = "binomial", `link` = "logit"). In addition we examine this data set using a user-defined link function. In Example 6.18 we will demonstrate the effect of a left tail modification using the `glm` function in R. The parameter for this modification is derived in Example 7.2 by using the `glmProfile` function.

Example 4.3 (Byssinosis among cotton workers)
The following data is about the medical conditions of cotton textile workers and was reported by Higgins and Koch (1977). The data was gathered in a large cross-sectional occupational health survey. Higgins and Koch (1977) focused on the relationship between the complaints of the workers of byssinosis and the variables sex, race, length of employment (employment years), smoking habit (smoking) and the dustiness of the work area (type of workplace). Following Czado (2007), we can use the scheme given in Table 4.3 to group our sample (5419 workers) by these factors. The incidence of byssinosis is reflected by a binary response on whether or not the workers suffer from symptoms related to this respiratory ailment.

factor	factor level	-1 $\widehat{=}$	0 $\widehat{=}$	1 $\widehat{=}$
type of workplace	factor	most dusty	less dusty	least dusty
employment years	description	< 10 yrs.	10 - 19 yrs.	> 19 yrs.
smoking			nonsmoker	smoker in last 5 yrs.
sex			female	male
race			others	white
byssinosis			no complaint	complaint

Table 4.3: Classification of cotton textile workers in health survey.

Czado (2007) and Higgins and Koch (1977) identified three covariates of particular importance. They are *employment years*, *smoking* and *workplace*. Thus, the data used in our further analysis will not include the other covariates and can be found in Table 4.4.

Employment years	Smoking	most dusty	Workplace less dusty	least dusty
< 10	yes	30/233	3/403	11/951
	no	7/126	5/283	7/733
10 − 19	yes	16/67	2/94	3/320
	no	3/20	1/51	1/160
> 19	yes	41/155	4/237	15/733
	no	8/72	3/232	5/553

Table 4.4: Data gathered about cotton textile workers.

Using the factor levels as described in Table 4.3 and resorting the data set w.r.t. the covariate *workplace* (abbreviated by work) yields to Table 4.5. All in all, the data frame `bys.ex` consist of 18 observations.

y_i	n_i	work	smoking	employ	y_i	n_i	work	smoking	employ
30	233	-1	1	-1	1	51	0	0	0
7	126	-1	0	-1	4	237	0	1	1
16	67	-1	1	0	3	232	0	0	1
3	20	-1	0	0	11	951	1	1	-1
41	151	-1	1	1	7	733	1	0	-1
8	72	-1	0	1	3	320	1	1	0
3	403	0	1	-1	1	160	1	0	0
5	283	0	0	-1	15	733	1	1	1
2	94	0	1	0	5	553	1	0	1

Table 4.5: Incidence of byssinosis among cotton workers as given in `bys.ex`.

The first column (i.e. y_i) contains the number of workers complaining about symptoms of byssinosis. The second column gives the total number n_i of workers falling in this group (i.e. $\sum_{i=1}^{18} n_i = 5419$). The factor levels of the covariate *workplace* are contained in the third column (i.e. *work*). The fourth column shows the factor levels of the covariate *smoking*. In the last column (i.e. *employ*) the factor levels of the covariate *employment years* are presented. Again we can see that the number of workers, suffering from byssinosis (i.e. y_i) corresponds to the binomial distribution. Therefore, considering the distribution of $\frac{y_i}{n_i}$ yields to a scaled binomial distribution. Thus, we consider the ratio $\frac{y_i}{n_i}$ as the response in our model, i.e. the ratio of workers suffering from byssinosis and the number of workers falling in this specific group ($= \frac{workers\ suffering\ from\ byssinosis\ in\ group\ i}{workers\ falling\ in\ group\ i}$).

In Example 5.3 the output of the ordinary `glm` function on this data set will be presented (using `family` = "binomial", `link` = "logit"). We will also examine this data set using a user-defined link function. In Example 6.20 we will demonstrate the effect of a left tail modification using the `glm` function in R. The parameter for this modification can be found in the analysis of the `bys.ex` data frame in Czado (2007).

Example 4.4 (Rotifer suspension)
The following example can be found in Collett (1999)(see p. 217). Two species of rotifer were investigated: the Polyarthra major (denoted by *species* = 1) and the Keratella cochlearis (denoted by *species* = 0). It was examined how many rotifers remained in suspension (i.e. y_i) after having introduced a certain number n_i into 40 tubes containing different relative *densities* of Ficoll. Thus, the data frame `rotifer.ex` contains 40 observations (i.e. $i \in \{1, \ldots, 40\}$) as we can see in Table 4.6.

y_i	n_i	density	species	den.cen		y_i	n_i	density	species	den.cen
11	58	1.019	1	-2.565		13	161	1.019	0	-2.565
7	86	1.020	1	-2.465		14	248	1.020	0	-2.465
10	76	1.021	1	-2.365		30	234	1.021	0	-2.365
19	83	1.030	1	-1.465		10	283	1.030	0	-1.465
9	56	1.030	1	-1.465		14	129	1.030	0	-1.465
21	73	1.030	1	-1.465		35	161	1.030	0	-1.465
13	29	1.031	1	-1.365		26	167	1.031	0	-1.365
34	44	1.040	1	-0.465		32	286	1.040	0	-0.465
10	31	1.040	1	-0.465		22	117	1.040	0	-0.465
36	56	1.041	1	-0.365		23	162	1.041	0	-0.365
20	27	1.048	1	0.335		7	42	1.048	0	0.335
54	59	1.049	1	0.435		22	48	1.049	0	0.435
20	22	1.050	1	0.535		9	49	1.050	0	0.535
9	14	1.050	1	0.535		34	160	1.050	0	0.535
14	17	1.060	1	1.535		71	74	1.060	0	1.535
10	22	1.061	1	1.635		25	45	1.061	0	1.635
64	66	1.063	1	1.835		94	101	1.063	0	1.835
68	86	1.070	1	2.535		63	68	1.070	0	2.535
488	492	1.070	1	2.535		178	190	1.070	0	2.535
88	89	1.070	1	2.535		154	154	1.070	0	2.535

Table 4.6: Data about rotifers in suspension as given in `rotifer.ex`.

The first column (i.e. y_i) contains the number of rotifers remaining in suspension in tube i (for $i \in \{1, \ldots, 40\}$). The number n_i of rotifers, which were introduced in the very same suspension are contained in the second column. The suspensions in the tubes contained different relative densities of Ficoll as specified by *density* in the third column. The column *species* gives information about what species of rotifer (1 $\hat{=}$ Polyarthra major, 0 $\hat{=}$ Keratella cochlearis) was examined. The centered density is given by *den.cen* in the last column. As before we consider the response $\frac{y_i}{n_i}$, i.e. the ratio of the rotifers remaining in suspension and the number of rotifers being introduced in this suspension $\left(= \frac{\#\ number\ of\ rotifers\ remaining\ in\ suspension}{\#\ number\ of\ rotifers\ introduced\ in\ this\ suspension} \right)$. Again the distribution of $\frac{y_i}{n_i}$ corresponds to a scaled binomial distribution.

In Example 5.4 the output of the ordinary `glm` function on this data set will be presented (using `family` = "binomial", `link` = "probit"). In addition we examine this data set using a user-defined link function. In Example 6.27 we will demonstrate the effect of a both tail modification using the `glm` function in R. The parameter for this modification can be found in the analysis of the `rotifer.ex` data frame in Czado (2007).

4.3 Poisson family

Example 4.5 (Coal mining fractures)
The following example is reported in Myers (1990) and consists of the observations made from 44 coal mines in the coal fields of the Appalachian region in western Virginia. It contains the number of miners being either injured or having a fracture denoted by y_i.

Moreover, we have information about the inner burden thickness (= *INB*, in ft., i.e. the shortest distance between seam floor and lower seam), the amount of previously mined seam in percent, i.e. the percentage of extraction (= *EX*) and the time in years that the mine has been opened (= *T*). The data set `mining.ex` is given in Table 4.7:

y_i	INB	EX	T	INB.cen	EX.cen	y_i	INB	EX	T	INB.cen	EX.cen
2	50	70	1.0	-119.23	-5.93	3	65	75	5.0	-104.23	-0.93
1	230	65	6.0	60.77	-10.93	3	470	90	9.0	300.77	14.07
0	125	70	1.0	-44.23	-5.93	2	300	80	9.0	130.77	4.07
4	75	65	0.5	-94.23	-10.93	2	275	90	4.0	105.77	14.07
1	70	65	0.5	-99.23	-10.93	0	420	50	17.0	250.77	-25.93
2	65	70	3.0	-104.23	-5.93	1	65	80	15.0	-104.23	4.07
0	65	60	1.0	-104.23	-15.93	5	40	75	15.0	-129.23	-0.93
0	350	60	0.5	180.77	-15.93	2	900	90	35.0	730.77	14.07
4	350	90	0.5	180.77	14.07	3	95	88	20.0	-74.23	12.07
4	160	80	0.0	-9.23	4.07	3	40	85	10.0	-129.23	9.07
1	145	65	10.0	-24.23	-10.93	3	140	90	7.0	-29.23	14.07
4	145	85	0.0	-24.23	9.07	0	150	50	5.0	-19.23	-25.93
1	180	70	2.0	10.77	-5.93	0	80	60	5.0	-89.23	-15.93
5	43	80	0.0	-126.23	4.07	2	80	85	5.0	-89.23	9.07
2	42	85	12.0	-127.23	9.07	0	145	65	9.0	-24.23	-10.93
5	42	85	0.0	-127.23	9.07	0	100	65	9.0	-69.23	-10.93
5	45	85	0.0	-124.23	9.07	3	150	80	3.0	-19.23	4.07
5	83	85	10.0	-86.23	9.07	2	150	80	0.0	-19.23	4.07
0	300	65	10.0	130.77	-10.93	3	210	75	2.0	40.77	-0.93
5	190	90	6.0	20.77	14.07	5	11	75	0.0	-158.23	-0.93
1	145	90	12.0	-24.23	14.07	0	100	65	25.0	-69.23	-10.93
1	510	80	10.0	340.77	4.07	3	50	88	20.0	-119.23	12.07

Table 4.7: Data about injuries and fractures of miners as given in `mining.ex`.

The first column (i.e. y_i) contains the number of miners being either injured or having a fracture. The second column (i.e. *INB*) contains the data about the inner burden thickness in this mine. The third column (i.e. *EX*) gives the percentage of extraction. The third covariate (i.e. *T*) is the time that the mine has been opened and it is given in the fourth column. In our further analysis we do only consider the covariates *INB* and *EX*, since they have a linear influence, while the covariate *T* has no effect on our response y_i. Centered versions of these two covariates are given in *INB.cen* and *EX.cen* in the last two columns. As explained in Subsection 3.3.3 we can use the Poisson distribution to model the number of accidents in a fixed time period. Hence, we choose a Poisson regression model for our response y_i (i.e. the number of injured or fractured miners).

In Example 5.5 the output of the ordinary `glm` function on this data set will be presented (using `family` = "poisson", `link` = "log"). This data set will also be examined using a user-defined link function. In Example 6.34 we will demonstrate the effect of a right tail modification using the `glm` function in R. The parameter for this modification is derived in Example 7.3 by using the `glmProfile` function.

4.4   Gamma family

Example 4.6 (Canadian automobile insurance)
The data presented in the following is about the Canadian automobile experience and was analyzed in Bailey and Simon (1960). It reflects the policy years 1957 and 1958 (as of June 30, 1959) for private passenger automobile liability for non farmers (for all of Canada excluding Saskatchewan). It includes the experience of almost all insurance companies operating in Canada and was collected by the Statistical Agency (Statistical Department of the Canadian Underwriters' Association) acting under instructions from the Superintendent of Insurance. Table 4.8 shows the factors and definitions of the variables *Merit* and *Class*.

Merit	Description	Original definition
3	licensed and accident free three or more years	A
2	licensed and accident free two years	X
1	licensed and accident free one year	Y
0	all others	B

Class	Description	
1	pleasure, no male operator under 25	
2	pleasure, non-principal male operator under 25	
3	business use	
4	unmarried owner or principal operator under 25	
5	married owner or principal operator under 25	

Table 4.8: Merit rating definition and class definitions.

The data frame given in Table 4.9 contains 20 observations on groups of Canadian private passenger automobile insurance holders. The insurance holders were classified using a multiple classification system. The column *Merit* gives information about the accidental behavior of several classes of car insurance holders. The merit rating plan is a classification according to previous accidents and conviction records. The second column *Class* represents the so-called class plan. It is a collective of the variables age, sex, use and occupation. Also we are given information about the earned car years, which is contained in the column *Insured*. Of further importance is the variable *Premium* containing the earned premiums at present given in thousand dollars. The premiums were adjusted to what they would have been if all the cars had been written at 1B rates. The fifth column *Claims* presents the number of claims incurred as given in Bailey and Simon (1959)(Table 1, p. 162). The last column *Cost* contains the losses incurred given in thousand dollars. The data will be primarily grouped by the merit rating and then by the class plan as in Bailey and Simon (1960). We can use the gamma distribution to model data sets about insurance claims (see Subsection 3.3.4). In our further work we will focus on the ratio $\frac{Cost}{Claims}$, which will be our response. We choose a gamma regression model with the inverse link. Finally we include weights given by the covariate *Claims* (i.e. `weights` = "Claims").

Merit	Class	Insured	Premium	Claims	Cost
3	1	2757520	159108	217151	63191
3	2	130535	7175	14506	4598
3	3	247424	15663	31964	9589
3	4	156871	7694	22884	7964
3	5	64130	3241	6560	1752
2	1	130706	7910	13792	4055
2	2	7233	431	1001	380
2	3	15868	1080	2695	701
2	4	17707	888	3054	983
2	5	4039	209	487	114
1	1	163544	9862	19346	5552
1	2	9726	572	1430	439
1	3	20369	1382	3546	1011
1	4	21089	1052	3618	1281
1	5	4869	250	613	178
0	1	273944	17226	37730	11809
0	2	21504	1207	3421	1088
0	3	37666	2502	7565	2383
0	4	56730	2756	11345	3971
0	5	8601	461	1291	382

Table 4.9: Data about Canadian automobile insurance claims contained in `carinsur.ex`.

In Example 5.6 the output of the ordinary `glm` function on this data set will be presented (using `family` = "Gamma", `link` = "inverse" and `weights` = "Claims"). We will also examine this data set using a user-defined link function. In Example 6.41 we will demonstrate the effect of a left tail modification using the `glm` function in R. The parameter for this modification is derived in Example 7.4 by using the `glmProfile` function.

5 Examples: the ordinary `glm` function in R

In this section we want to apply the R function `glm` (see `help(glm)`). It is already implemented in R and requires the following input:

(i) `formula`: is an expression of the form: response $\sim\ covariate_1 + \cdots + covariate_k$ and for every data set we have to decide which covariates we will include in our model.

(ii) `family`: one of the families introduced in Section 3.3 (e.g. `family` = "binomial").

(iii) `data`: data set containing n rows of observations of both the response and the covariates.

(iv) ... further things could be specified (e.g. `weights`, `method` (IWLS by default), `intercept` (is of type logical), `start` (is a vector of starting values for the parameters in the linear predictor), ...)

5.1 Gaussian family

Example 5.1 (PCB concentration in lake trouts)
Now we want to give an example of the Gaussian regression (see Subsection 3.3.1). For this we call the `glm` function with `family` = "gaussian" and the identity link. The data set we will examine is `pcb.ex` as described in Example 4.1.

```
pcb_glm<- glm(formula = log.pcb~age.cen,
              family = gaussian(link="identity"),
              data = pcb.ex)
summary(pcb_glm)

##
## Call:
## glm(formula = log.pcb ~ age.cen, family = gaussian(link = "identity"),
##     data = pcb.ex)
##
## Deviance Residuals:
##     Min        1Q    Median        3Q       Max
## -1.1395   -0.3879    0.0957    0.4327    1.0508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.4659     0.1072   13.68  2.2e-13 ***
## age.cen       0.2591     0.0308    8.41  6.8e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.3215)
##
```

```
##      Null deviance: 31.120  on 27  degrees of freedom
## Residual deviance:  8.359  on 26  degrees of freedom
## AIC: 51.61
##
## Number of Fisher Scoring iterations: 2
```

We are especially interested in:

```
summary(pcb_glm)$deviance
```

```
## [1] 8.359
```

5.2   Binomial family

In Subsection 3.3.2 we introduced two common types of the binomial regression. For both we want to give examples. In the following we will present the output of the `glm` function on the three data sets presented in Section 4.2. First we will exemplify the logistic regression in two examples and afterwards we will present an example of the probit regression.

5.2.1   Logistic regression

Example 5.2 (Beetle mortality)
Now we want to give the first example of the logistic regression (see Definition 3.35). For this we call the `glm` function with `family` = "binomial" and the logit link. The data set we will examine is `beetle.ex` as described in Example 4.2.

```
beetle_glm<- glm(formula = cbind(y, n-y)~ dose.cen,
                 family = binomial(link="logit"),
                 data = beetle.ex)
summary(beetle_glm)
```

```
##
## Call:
## glm(formula = cbind(y, n - y) ~ dose.cen,
##     family = binomial(link = "logit"),
##     data = beetle.ex)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.594  -0.394   0.833   1.259   1.594
##
## Coefficients:
##            Estimate Std. Error z value Pr(>|z|)
## (Intercept)   0.744      0.138     5.4  6.8e-08 ***
```

```
## dose.cen        34.270       2.912      11.8   < 2e-16 ***
## ---
## Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.202  on 7  degrees of freedom
## Residual deviance:  11.232  on 6  degrees of freedom
## AIC: 41.43
##
## Number of Fisher Scoring iterations: 4
```

We are especially interested in:

```
summary(beetle_glm)$deviance
```

```
## [1] 11.23
```

Example 5.3 (Byssinosis among cotton workers)
We illustrate the logistic regression (see Definition 3.35) with a second example. Again
we call the `glm` function with `family` = "binomial" and the logit link. The data set we
will examine is `bys.ex` as described in Example 4.3.

```
bys_glm<- glm(formula = cbind(y, n-y)~ workplace+smoking+employment,
              family = binomial(link="logit"),
              data = bys.ex)
summary(bys_glm)
```

```
##
## Call:
## glm(formula = cbind(y, n - y) ~ workplace + smoking + employment,
##     family = binomial(link = "logit"), data = bys.ex)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -3.336  -0.482   0.162   1.160   2.105
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -3.7626     0.1652  -22.78   < 2e-16 ***
## workplace    -1.4657     0.1058  -13.86   < 2e-16 ***
## smoking       0.6778     0.1887    3.59   0.00033 ***
## employment    0.3331     0.0886    3.76   0.00017 ***
## ---
## Signif. codes:   0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 290.739  on 17  degrees of freedom
## Residual deviance:  40.774  on 14  degrees of freedom
## AIC: 112.1
##
## Number of Fisher Scoring iterations: 5
```

We are especially interested in:

```
summary(bys_glm)$deviance
```

```
## [1] 40.77
```

5.2.2   Probit regression

Example 5.4 (Rotifer suspension)
In Subsection 3.3.2 we defined another important regression model. This was the probit regression (see Definition 3.36). In the following example we call the `glm` function with `family` = "binomial" and the probit link examining the data set `rotifer.ex` as described in Example 4.4.

```
rotifer_glm<- glm(formula = cbind(y, n-y)~species
                                        +density.cen
                                        +species*density.cen,
                family = binomial(link="probit"),
                data = rotifer.ex)
summary(rotifer_glm)
```

```
##
## Call:
## glm(formula = cbind(y, n - y) ~ species + density.cen + species *
##     density.cen, family = binomial(link = "probit"), data = rotifer.ex)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -6.836  -2.319   0.369   2.412   6.423
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)         -0.3939     0.0297  -13.26   <2e-16 ***
## species              0.7379     0.0531   13.90   <2e-16 ***
## density.cen          0.6085     0.0195   31.13   <2e-16 ***
## species:density.cen -0.0135     0.0315   -0.43     0.67
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3180.99  on 39  degrees of freedom
## Residual deviance:  471.25  on 36  degrees of freedom
## AIC: 633.8
##
## Number of Fisher Scoring iterations: 6
```

We are especially interested in:

```
summary(rotifer_glm)$deviance
```

```
## [1] 471.3
```

5.3   Poisson family

Example 5.5 (Coal mining fractures)
Now we want to give an example of the Poisson regression (see Subsection 3.3.3). For this
we call the `glm` function with `family` = "poisson" and the log link. The data set we will
analyze is `mining.ex` as described in Example 4.5.

```
mining_glm<- glm(formula = y~inb.cen+extrp.cen,
                 family = poisson(link="log"),
                 data = mining.ex)
summary(mining_glm)
```

```
##
## Call:
## glm(formula = y ~ inb.cen + extrp.cen, family = poisson(link = "log"),
##     data = mining.ex)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.926  -0.948  -0.188   0.534   2.092
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.599236   0.123749    4.84  1.3e-06 ***
## inb.cen     -0.001708   0.000747   -2.29    0.022 *
## extrp.cen    0.058420   0.011811    4.95  7.6e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 74.984  on 43  degrees of freedom
## Residual deviance: 42.094  on 41  degrees of freedom
## AIC: 144.4
##
## Number of Fisher Scoring iterations: 5
```

We are especially interested in:

```
summary(mining_glm)$deviance
```

```
## [1] 42.09
```

5.4   Gamma family

Example 5.6 (Canadian automobile insurance)
The last example we will present in this chapter is for the gamma regression (see Subsection 3.3.4). For this we call the `glm` function with `family` = "Gamma" and the inverse link. We will examine the data set `carinsur.ex` as described in Example 4.6.

```
carinsur_glm<-glm(formula = Cost/Claims~Merit+Class,
                  weights = Claims,
                  family = Gamma(link="inverse"),
                  data = carinsur.ex)
summary(carinsur_glm)

##
## Call:
## glm(formula = Cost/Claims ~ Merit + Class,
##     family = Gamma(link = "inverse"),
##     data = carinsur.ex, weights = Claims)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -6.012  -1.889  -0.335   2.192   6.388
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.2466     0.0510   63.72  < 2e-16 ***
## Merit1        0.2153     0.0888    2.42   0.0321 *
## Merit2        0.2237     0.0994    2.25   0.0440 *
## Merit3        0.1773     0.0534    3.32   0.0061 **
## Class2       -0.2678     0.0859   -3.12   0.0089 **
```

```
## Class3         -0.0539      0.0635    -0.85    0.4122
## Class4         -0.4984      0.0594    -8.39    2.3e-06 ***
## Class5          0.2874      0.1486     1.93    0.0770 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 14.15)
##
##     Null deviance: 1556.01  on 19  degrees of freedom
## Residual deviance:  167.43  on 12  degrees of freedom
## AIC: -2972929
##
## Number of Fisher Scoring iterations: 4
```

We are especially interested in:

```
summary(carinsur_glm)$deviance
```

```
## [1] 167.4
```

6   Generalized linear models with parametric link

There are many cases in which we obtain a misspecification of the link function. The reason is simple: we have to choose the link function before having sufficient information about the choice of the link. Thus, we would like to describe a way to improve the goodness of fit of the GLMs by reducing the deviance (see Section 3.4). It will turn out, that one elegant way to improve the models is to allow link functions coming from the parametric link families specified in Czado (2007). This advantageous parametric class of link transformations was developed by Czado (1992). The general $h(\cdot)$ - power transformation functions are the key elements for modifying the tails of a graph. We will introduce them in Section 6.1. We then will be able to define the parametric link families in Section 6.2. Further it will be demonstrated, how the use of a parametric link reduces the deviance of the GLMs presented in Chapter 5. In Section 6.3 we will describe the `glmProfile` function. It helps to find the optimal parameter for the parametric link functions and thus it will be applied on the data sets in Chapter 7.

6.1   General $h(\cdot)$ - power transformations

In the following we want to investigate the behavior of the h($\cdot$) - power transformation functions for specific values of $\boldsymbol{\psi}$. Therefore we have to pass the function a start value $\eta_0 \in \mathbb{R}$ and either ψ_1 or $\psi_2 \in \mathbb{R}$ for a single tail modification or the values of the vector $\boldsymbol{\psi} \in \mathbb{R}^2$ for a both tail modification. For a left tail modification every point $< \eta_0$ will be modified whereas for a right tail modification every point $\geq \eta_0$ is. A both tail modification modifies both tails (i.e. all points $< \eta_0$ and $\geq \eta_0$ are modified).

Remark 6.1 (Remarks on the choice of the parameter $\boldsymbol{\psi}$)
- by setting the parameters to 1 we obtain no modification (i.e. a straight line).

- if we set one parameter to 1 in the both tail modification, we get a single tail modification (e.g. $\boldsymbol{\psi} = (1, \psi_2)$ modifies the left tail).

- in a right tail modification a parameter $\psi_1 < 1$ *decreases* the slope, while setting $\psi_1 > 1$ *increases* it. In the left tail modification it is the other way round for ψ_2.

- in the cases $\psi_1 \vee \psi_2 = 0$ we had to look at e.g.:

$$\lim_{\psi_1 \to 0} \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} \overset{L'Hospital}{=} \lim_{\psi_1 \to 0} \frac{\ln(\eta - \eta_0 + 1)(\eta - \eta_0 + 1)^{\psi_1}}{1}$$
$$= \ln(\eta - \eta_0 + 1) \underbrace{\lim_{\psi_1 \to 0} (\eta - \eta_0 + 1)^{\psi_1}}_{=1} = \ln(\eta - \eta_0 + 1)$$

  The other cases can be shown analogously.

- for numerical reasons, we had to weaken the statement $\psi_1 \vee \psi_2 = 0$ to $|\psi_1| \vee |\psi_2| < \varepsilon$, where in our case the tiny number $\varepsilon = $ `1e-14` $\widehat{=} 1 \cdot 10^{-14}$.

Definition 6.2 (Right tail modification)

$$h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1) = \begin{cases} \eta_0 + \ln(\eta - \eta_0 + 1) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \eta & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}$$
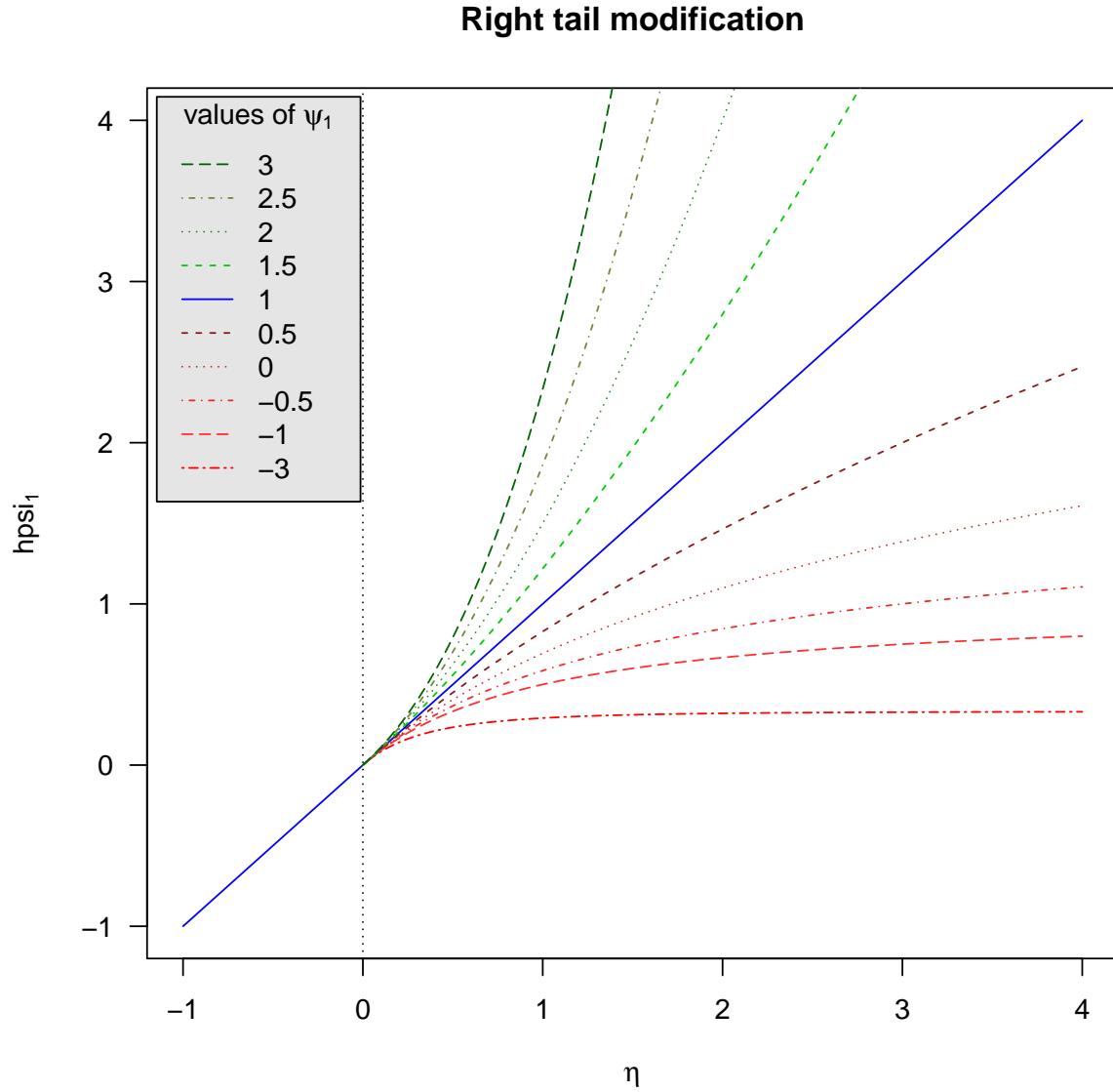
Right tail modification

Figure 6.1: Right tail modification for different values of ψ_1 as implemented in the function `hpsi1` ($\eta_0 = 0$ as indicated by the dotted vertical line).

Definition 6.3 (Left tail modification)

$$h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2) = \begin{cases} \eta & \text{if } \eta \geq \eta_0 \\ \eta_0 - \ln(-\eta + \eta_0 + 1) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \eta_0 - \frac{(-\eta+\eta_0+1)^{\psi_2}-1}{\psi_2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

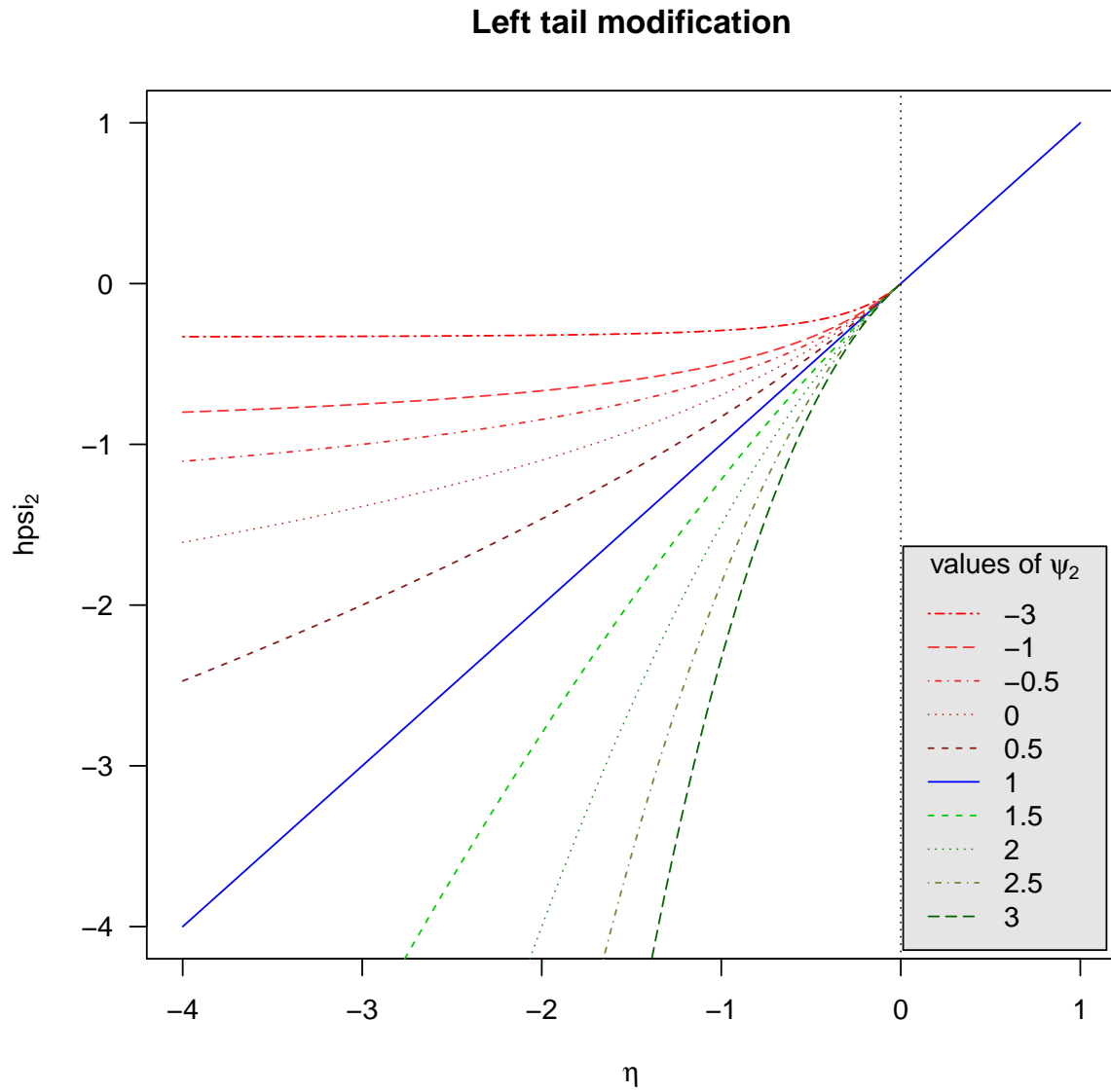Figure 6.2: Left tail modification for different values of ψ_2 as implemented in the function `hpsi2` ($\eta_0 = 0$ as indicated by the dotted vertical line).

Definition 6.4 (Both tail modification)

$$h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2)) = \begin{cases} \eta_0 + \ln\left(\eta - \eta_0 + 1\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \eta_0 - \ln\left(-\eta + \eta_0 + 1\right) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$
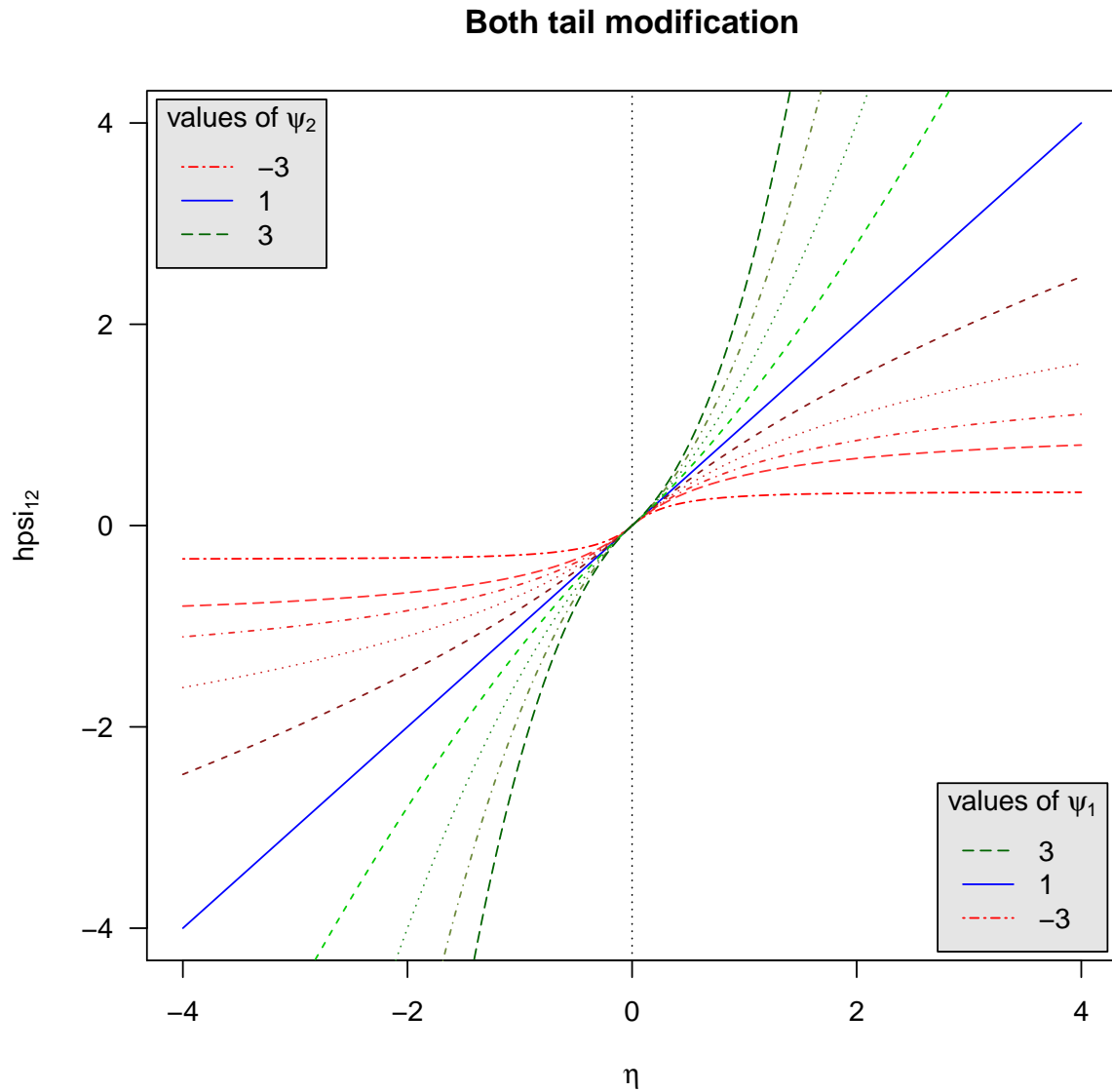
Both tail modification

Figure 6.3: Both tail modification for different values of $\boldsymbol{\psi}$ as implemented in the function `hpsi12`. Arbitrary combinations of the parameters ψ_1 and ψ_2 can be made to transform the tail. Again $\eta_0 = 0$ as indicated by the dotted vertical line.

Remark 6.5 (More about the general $h(\cdot)$ - power transformations)
The implementation of the single `hpsi` functions can be found in Section A.2. Further we present related mathematical quantities including the domains of the `hpsi` functions, the inverses, the partial derivatives w.r.t. η and the partial derivatives w.r.t. $\boldsymbol{\psi}$. I.e. we calculated the derivative w.r.t. ψ_1 in the case of a right tail modification (w.r.t. ψ_2 in the case of a left tail modification) and for the both tail modification we calculated both partial derivatives w.r.t. ψ_1 and ψ_2. In addition we present the limits as $|\eta| \to \infty$ (i.e. $\lim_{|\eta| \to \infty}$) for the single tail modifications.

6.2   Parametric link families

Now we want to improve the GLMs presented in Chapter 5 using user-defined link functions. Therefore, we modify the link functions introduced in Section 3.3. The topic of this chapter will be how such link functions can be implemented in R and what specific link functions we will use for the different families. We will proceed with the implementation of the methods and functions described in Czado (2007), modifying them slightly where it is needed. The following table gives an overview of parametric link families using the general $h(\cdot)$ - power transformations we have introduced in Section 6.1. If we take a link function from these parametric link families, we speak of a tail modified GLM since we use a tail modified link function. We restricted ourselves to the most common families in GLMs and specified the allowable modifications as it can be seen in Table 6.1.

Error distr.	Linkinv. $F(\eta)$	Link family $\mathcal{F} = \{F(\cdot, \boldsymbol{\psi}) \mid \psi \in \Psi\}$	Allow. modif.
Normal	$F(\eta) = \eta$	$F(\eta, \boldsymbol{\psi}) = h_{\eta_0}(\eta, \boldsymbol{\psi})$	all
Scaled	$F(\eta) = \frac{\exp\{\eta\}}{1+\exp\{\eta\}}$	$F(\eta, \boldsymbol{\psi}) = \frac{\exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}{1+\exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}$	all
binomial	$F(\eta) = \Phi(\eta)$	$F(\eta, \boldsymbol{\psi}) = \Phi(h_{\eta_0}(\eta, \boldsymbol{\psi}))$	all
Poisson	$F(\eta) = \exp\{\eta\}$	$F(\eta, \boldsymbol{\psi}) = \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}$	right
Gamma	$F(\eta) = \frac{1}{\eta}$	$F(\eta, \boldsymbol{\psi}) = \frac{1}{h_{\eta_0}(\eta, \boldsymbol{\psi})}$	left

Table 6.1: Overview of the link families using η_0-standardized tail modifications.

To implement the parametric link families we have to create user-defined link functions. One way to define a user-defined link function in R is to specify all of the following components: `linkfun`, `linkinv`, `mu.eta`, `valideta` and `name` in an object of `class` "link-glm" (see also `help(make.link)` and Section 3.3).

6.2.1   Gaussian family with tail modified identity link

In the following we want to derive the tail modified identity link function for all possible modifications (i.e. "right", "left" and "both").

Remark 6.6 (Gaussian family with identity link)
In Subsection 3.3.1 we defined the ordinary identity link function through the following specifications:

(i) `linkfun`: $\eta = g(\mu) = \mu$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = \eta$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = 1$

(iv) `valideta`: $\mathbb{1}_{\{\eta \in \mathbb{R}\}} = \text{TRUE}$

We can define the tail modified Gaussian regression by modifying the linkinverse in the ordinary Gaussian regression as follows.

Definition 6.7 (Tail modified Gaussian regression)
Let $h_{\eta_0}(\eta, \boldsymbol{\psi})$ be a tail modification as defined in Section 6.1 (see Definitions 6.2, 6.3 and 6.4). Then, the tail modified Gaussian regression has an inverse link function of the following form:

$$F(\eta, \boldsymbol{\psi}) = h_{\eta_0}(\eta, \boldsymbol{\psi}) \tag{6.1}$$

Definition 6.8 (Tail modified identity link)
Using Equation (6.1), we can define the tail modified identity link through the following specifications:

(i) `linkfun`: $g(\mu, \boldsymbol{\psi})$:
   We had $\eta = g(\mu) = \mu$ for the identity link function. Hence, now we have:

$$h_{\eta_0}(\eta, \boldsymbol{\psi}) = \mu$$
$$\Leftrightarrow \eta = (h_{\eta_0}(\eta, \boldsymbol{\psi}))^{-1}(\mu)$$
$$\Rightarrow g(\mu, \boldsymbol{\psi}) = (h_{\eta_0}(\eta, \boldsymbol{\psi}))^{-1}(\mu)$$

(ii) `linkinv`: $F(\eta, \boldsymbol{\psi}) = h_{\eta_0}(\eta, \boldsymbol{\psi})$

(iii) `mu.eta`: $\frac{\partial \mu}{\partial \eta} = \frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi})$:

$$\frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi}) = \frac{\partial}{\partial \eta}(h_{\eta_0}(\eta, \boldsymbol{\psi}))$$

(iv) `valideta`: TRUE, if η is in the domain of $F(\eta, \boldsymbol{\psi})$.

(v) `name`: "psi1GAUSS" for the right tail modification, "psi2GAUSS" for the left tail modification and "psi12GAUSS" for the both tail modification.

Remark 6.9
If we use a right tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ as defined in Definition 6.2 (a left tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ as defined in Definition 6.3) for $h_{\eta_0}(\eta, \boldsymbol{\psi})$, we speak of a *right tail modified Gaussian regression* (*left tail modified Gaussian regression*). If we use $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ as defined in Definition 6.4 for $h_{\eta_0}(\eta, \boldsymbol{\psi})$, we speak of a *both tail modified Gaussian regression*.

Example 6.10 (Right tail modified Gaussian regression)
Let the right tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ be given as in Definition 6.2, then the inverse of it is given as calculated in Equation (A.2) and the partial derivative w.r.t. η is given as calculated in Equation (A.3). We determine the values, which are in the domain of $F(\eta, \boldsymbol{\psi})$. As we derived above we have:

$$F(\eta, \boldsymbol{\psi}) = h_{\eta_0}(\eta, \boldsymbol{\psi})$$

Therefore, the domain of $F(\eta, \boldsymbol{\psi})$ is the domain of the corresponding `hpsi` function $h_{\eta_0}(\eta, \boldsymbol{\psi})$ as given in Equation (A.1) (and in Equations A.4 and A.7). We derived the following implementation of the right tail modified identity link called "psi1GAUSS".

```
psi1GAUSS<-function(psi1 = 1, eta0 = 0)
{linkfun <- function(mu) {hpsi1INV(psi1, mu, eta0)}
 linkinv <- function(eta){hpsi1(psi1, eta, eta0)}
 mu.eta <- function(eta) {hpsi1DERIV(psi1, eta, eta0)}
 valideta<-function(eta) {h <- 1:length(eta)
 for (i in 1:length(eta) ) {
    if (is.finite(linkinv(eta[i]))) {h[i] <- TRUE}
    else {h[i] <- FALSE}
    }
  h
 }
 link <- paste("psi1GAUSS(", psi1, " , " , eta0, ")")
 structure(list(linkfun = linkfun,
                linkinv = linkinv,
                mu.eta = mu.eta,
                valideta = valideta,
                name = link),
          class = "link-glm")
}
```

Example 6.11 (PCB concentration in lake trouts)
Now we present the output of a right tail modified Gaussian regression with two different parameters. As in Example 5.1 our analysis is based on the data frame `pcb.ex` (see Example 4.1). The first parameter is chosen in a way that decreases the deviance drastically (for further information see Example 7.1). For the other parameter we choose the value $\psi_1 = 1$. This yields to an ordinary GLM with `family` = "gaussian" and `link` = "identity".

```
pcb_glm_right<- glm(formula = log.pcb~age.cen,
                    family = gaussian(link=psi1GAUSS(psi1=0.18,eta0=0)),
                    data = pcb.ex)
summary(pcb_glm_right)

##
## Call:
```

```
## glm(formula = log.pcb ~ age.cen,
##     family = gaussian(link = psi1GAUSS(psi1 = 0.18, eta0 = 0)),
##     data = pcb.ex)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -0.798  -0.334   0.020   0.348   1.034
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.477      0.343    10.1  1.6e-10 ***
## age.cen        0.801      0.088     9.1  1.4e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2433)
##
##     Null deviance: 31.1196  on 27  degrees of freedom
## Residual deviance:  6.3253  on 26  degrees of freedom
## AIC: 43.81
##
## Number of Fisher Scoring iterations: 5

summary(pcb_glm_right)$deviance

## [1] 6.325
```

Remark 6.12 (No transformation for $\psi_1 = 1$)

```
pcb_glm_right2<- glm(formula = log.pcb~age.cen,
                     family = gaussian(link=psi1GAUSS(psi1=1,eta0=0)),
                     data = pcb.ex)
summary(pcb_glm_right2)$deviance

## [1] 8.359
```

Setting the link parameter ψ_1 to 1 yields to no transformation (see Remark 6.1). Hence, the summary of `pcb_glm_right2` yields to the same output as in Example 5.1. In particular, the deviance is the same, as can be seen above.

6.2.2   Binomial family with tail modified logit link

In the following we want to derive the tail modified logit link function for all possible modifications (i.e. "right", "left" and "both").

Remark 6.13 (Binomial family with logit link)
In Subsection 3.3.2 we defined the ordinary logit link function through the following specifications:

(i) `linkfun`: $\eta = g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = \frac{\exp\{\eta\}}{1+\exp\{\eta\}}$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = \frac{\exp\{\eta\}}{(1+\exp\{\eta\})^2}$

(iv) `valideta`: $\mathbb{1}_{\{\eta \in \mathbb{R}\}} = \text{TRUE}$

We can define the tail modified logistic regression by modifying the linkinverse in the ordinary logistic regression as follows.

Definition 6.14 (Tail modified logistic regression)
Let $h_{\eta_0}(\eta, \boldsymbol{\psi})$ be a tail modification as defined in Section 6.1 (see Definitions 6.2, 6.3 and 6.4). Then, the tail modified logistic regression has an inverse link function of the following form:

$$F(\eta, \boldsymbol{\psi}) = \frac{\exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}{1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}} \tag{6.2}$$

Definition 6.15 (Tail modified logit link)
Using Equation (6.2) we can define the tail modified logit link through the following specifications:

(i) `linkfun`: $g(\mu, \boldsymbol{\psi})$:

We had $\eta = g(\mu) = \ln\left(\frac{\mu}{1-\mu}\right)$ for the logit link function. Hence, now we have:

$$h_{\eta_0}(\eta, \boldsymbol{\psi}) = \ln\left(\frac{\mu}{1-\mu}\right)$$

$$\Leftrightarrow \eta = (h_{\eta_0}(\eta, \boldsymbol{\psi}))^{-1}\left(\ln\left(\frac{\mu}{1-\mu}\right)\right)$$

$$\Rightarrow g(\mu, \boldsymbol{\psi}) = (h_{\eta_0}(\eta, \boldsymbol{\psi}))^{-1}\left(\ln\left(\frac{\mu}{1-\mu}\right)\right)$$

(ii) `linkinv`: $F(\eta, \boldsymbol{\psi})$:

$$F(\eta, \boldsymbol{\psi}) = \frac{\exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}{1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}$$

$$= \underbrace{\frac{1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}{1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}}_{=1} - \frac{1}{1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}$$

$$= 1 - \frac{1}{1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}$$

(iii) `mu.eta`: $\frac{\partial \mu}{\partial \eta} = \frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi})$:

$$
\begin{aligned}
\frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi}) &= \frac{\partial}{\partial \eta} \left(\frac{\exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}{1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}} \right) \\
&= \frac{\left(\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi})\right) \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\} \left(1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}\right)}{(1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\})^2} \\
&\quad - \frac{\exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\} \left(\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi})\right) \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}{(1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\})^2} \\
&= \frac{\left(\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi})\right) \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}{(1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\})^2} = \left(\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi})\right) F(\eta, \boldsymbol{\psi}) \left(1 - F(\eta, \boldsymbol{\psi})\right)
\end{aligned}
$$

(iv) `valideta`: TRUE, if η is in the domain of $F(\eta, \boldsymbol{\psi})$.

(v) `name`: "psi1LOGIT" for the right tail modification, "psi2LOGIT" for the left tail modification and "psi12LOGIT" for the both tail modification.

Remark 6.16
If we use a right tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ as defined in Definition 6.2 (a left tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ as defined in Definition 6.3) for $h_{\eta_0}(\eta, \boldsymbol{\psi})$, we speak of a *right tail modified logistic regression (left tail modified logistic regression)*. If we use $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ as defined in Definition 6.4 for $h_{\eta_0}(\eta, \boldsymbol{\psi})$, we speak of a *both tail modified logistic regression*.

Example 6.17 (Left tail modified logistic regression)
Let the left tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ be given as in Definition 6.3, then the inverse of it is given as calculated in Equation (A.5) and the partial derivative w.r.t. η is given as calculated in Equation (A.6). We determine the values, which are in the domain of $F(\eta, \boldsymbol{\psi})$. As we derived above we have:

$$
F(\eta, \boldsymbol{\psi}) = 1 - \frac{1}{1 + \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi})\}}
$$

The domain of $\exp\{\cdot\}$ is the real line and $\exp\{\cdot\}$ is non-negative. Therefore, η has to be in the domain of the `hpsi` functions $h_{\eta_0}(\eta, \boldsymbol{\psi})$ as given in Equation (A.4) (and in Equations A.1 and A.7). We derived the following implementation of the left tail modified logit link called "psi2LOGIT".

```
psi2LOGIT<-function(psi2 = 1, eta0 = 0)
{linkfun <- function(mu) hpsi2INV(psi2, log(mu/(1 - mu)), eta0)
 linkinv <- function(eta) 1 - (1/(1 + exp(hpsi2(psi2, eta, eta0))))
 mu.eta <- function(eta){
  (hpsi2DERIV(psi2, eta, eta0)
  *(1 - (1/(1 + exp(hpsi2(psi2, eta, eta0))))))/
   (1 + exp(hpsi2(psi2, eta, eta0)))
```

```
 }
 valideta<-function(eta){h <- 1:length(eta)
 for (i in 1:length(eta) ) {
    if (is.finite(linkinv(eta[i]))) {h[i] <- TRUE}
    else {h[i] <- FALSE}
    }
  h
 }
 link <- paste("psi2LOGIT(", psi2, " , " , eta0, ")")
 structure(list(linkfun = linkfun,
                linkinv = linkinv,
                mu.eta = mu.eta,
                valideta = valideta,
                name = link),
           class = "link-glm")
}
```

Example 6.18 (Beetle mortality)
Now we present the output of a left tail modified logistic regression with two different parameters. As in Example 5.2 our analysis is based on the data frame `beetle.ex` (see Example 4.2). The first parameter is chosen in a way that decreases the deviance drastically (for further information see Example 7.2). For the other parameter we choose $\psi_2 = 1$. This yields to an ordinary GLM with `family` = "binomial" and `link` = "logit".

```
beetle_glm_left<- glm(formula = cbind(y, n-y)~dose.cen,
                      family = binomial(link=psi2LOGIT(psi2=0.16,eta0=0)),
                      data = beetle.ex)
summary(beetle_glm_left)

##
## Call:
## glm(formula = cbind(y, n - y) ~ dose.cen,
##     family = binomial(link = psi2LOGIT(psi2 = 0.16, eta0 = 0)),
##     data = beetle.ex)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -0.967  -0.323   0.198   0.569   0.945
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)    0.514      0.173    2.97   0.0029 **
## dose.cen      48.454      5.457    8.88   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 284.2024  on 7  degrees of freedom
## Residual deviance:   3.0445  on 6  degrees of freedom
## AIC: 33.24
##
## Number of Fisher Scoring iterations: 4

summary(beetle_glm_left)$deviance

## [1] 3.045
```

Remark 6.19 (No transformation for $\psi_2 = 1$)
```
beetle_glm_left2<- glm(formula = cbind(y, n-y)~dose.cen,
                        family = binomial(link=psi2LOGIT(psi2=1,eta0=0)),
                        data = beetle.ex)
summary(beetle_glm_left2)$deviance

## [1] 11.23
```

Setting the link parameter ψ_2 to 1 yields to no transformation (see Remark 6.1). Hence, the summary of `beetle_glm_left2` yields to the same output as in Example 5.2. In particular, the deviance is the same, as can be seen above.

Example 6.20 (Byssinosis among cotton workers)
Again we present the output of a left tail modified logistic regression with two different parameters. As in Example 5.3 our analysis is based on the data frame `bys.ex` (see Example 4.3). The first parameter is chosen in a way that decreases the deviance drastically (for further information see Czado (2007)(pp. 15-18)). For the other parameter we choose $\psi_2 = 1$, which yields to an ordinary GLM with `family` = "binomial" and `link` = "logit".

```
bys_glm_left<- glm(formula = cbind(y, n-y)~workplace+smoking+employment,
                   family = binomial(link=psi2LOGIT(psi2=-1.9626,
                                                     eta0=-3.912)),
                   start = c(-3.8, -1.5, 0.6, 0.3),
                   data = bys.ex)
summary(bys_glm_left)

##
## Call:
## glm(formula = cbind(y, n - y) ~ workplace + smoking + employment,
##     family = binomial(link = psi2LOGIT(psi2 = -1.9626, eta0 = -3.912)),
##     data = bys.ex, start = c(-3.8, -1.5, 0.6, 0.3))
```

```
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.133  -0.462  -0.103   0.492   1.854
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -5.506      0.826   -6.66  2.7e-11 ***
## workplace      -3.139      0.789   -3.98  7.0e-05 ***
## smoking         0.966      0.275    3.51  0.00045 ***
## employment      0.446      0.117    3.80  0.00015 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 290.7386  on 17  degrees of freedom
## Residual deviance:   9.2599  on 14  degrees of freedom
## AIC: 80.63
##
## Number of Fisher Scoring iterations: 7

summary(bys_glm_left)$deviance

## [1] 9.26
```

Remark 6.21 (No transformation for $\psi_2 = 1$)

```
bys_glm_left2<- glm(formula = cbind(y, n-y)~workplace+smoking+employment,
                    family = binomial(link=psi2LOGIT(psi2=1,
                                                     eta0=-3.912)),
                    start = c(-3.8, -1.5, 0.6, 0.3),
                    data = bys.ex)
summary(bys_glm_left2)$deviance

## [1] 40.77
```

Setting the link parameter ψ_2 to 1 yields to no transformation (see Remark 6.1). Hence, the summary of `bys_glm_left2` yields to the same output as in Example 5.3. In particular, the deviance is the same, as can be seen above.

6.2.3 Binomial family with tail modified probit link

In the following we want to derive the tail modified probit link function for all possible modifications (i.e. "right", "left" and "both").

Remark 6.22 (Binomial family with probit link)
In Subsection 3.3.2 we defined the ordinary probit link function through the following specifications:

(i) `linkfun`: $\eta = g(\mu) = qnorm(\mu) \stackrel{quantile\ function}{=} \Phi^{-1}(\mu)$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = pnorm(\eta) = \Phi(\eta)$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = dnorm(\eta) = \varphi(\eta)$

(iv) `valideta`: $\mathbb{1}_{\{\eta \in \mathbb{R}\}} = \text{TRUE}$

We can define the tail modified probit regression by modifying the linkinverse in the ordinary probit regression as follows.

Definition 6.23 (Tail modified probit regression)
Let $h_{\eta_0}(\eta, \boldsymbol{\psi})$ be a tail modification as defined in Section 6.1 (see Definitions 6.2, 6.3 and 6.4). Then, the tail modified probit regression has an inverse link function of the following form:
$$F(\eta, \boldsymbol{\psi}) = \Phi(h_{\eta_0}(\eta, \boldsymbol{\psi})) \tag{6.3}$$

Definition 6.24 (Tail modified probit link)
Using Equation (6.3) we can define the tail modified probit link through the following specifications:

(i) `linkfun`: $g(\mu, \boldsymbol{\psi})$:
We had $\eta = g(\mu) = qnorm(\mu) = \Phi^{-1}(\mu)$ for the probit link function. Hence, now we have:

$$\begin{aligned}
h_{\eta_0}(\eta, \boldsymbol{\psi}) &= \Phi^{-1}(\mu) \\
\Leftrightarrow \eta &= (h_{\eta_0}(\eta, \boldsymbol{\psi}))^{-1} \left(\Phi^{-1}(\mu) \right) \\
\Rightarrow g(\mu, \boldsymbol{\psi}) &= (h_{\eta_0}(\eta, \boldsymbol{\psi}))^{-1} \left(\Phi^{-1}(\mu) \right)
\end{aligned}$$

(ii) `linkinv`: $F(\eta, \boldsymbol{\psi}) = \Phi(h_{\eta_0}(\eta, \boldsymbol{\psi}))$

(iii) `mu.eta`: $\frac{\partial \mu}{\partial \eta} = \frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi})$:

$$\begin{aligned}
\frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi}) &= \frac{\partial}{\partial \eta} \left(\Phi(h_{\eta_0}(\eta, \boldsymbol{\psi})) \right) \\
&= \left(\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi}) \right) \varphi(h_{\eta_0}(\eta, \boldsymbol{\psi}))
\end{aligned}$$

(iv) `valideta`: TRUE, if η is in the domain of $F(\eta, \boldsymbol{\psi})$.

(v) `name`: "psi1PROBIT" for the right tail modification, "psi2PROBIT" for the left tail modification and "psi12PROBIT" for the both tail modification.

Remark 6.25

If we use a right tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ as defined in Definition 6.2 (a left tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ as defined in Definition 6.3) for $h_{\eta_0}(\eta, \boldsymbol{\psi})$, we speak of a *right tail modified probit regression* (*left tail modified probit regression*). If we use $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ as defined in Definition 6.4 for $h_{\eta_0}(\eta, \boldsymbol{\psi})$, we speak of a *both tail modified probit regression*.

Example 6.26 (Both tail modified probit regression)

Let the both tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ be given as in Definition 6.4, then the inverse of it is given as calculated in Equation (A.8) and the partial derivative w.r.t. η is given as calculated in Equation (A.9). We determine the values, which are in the domain of $F(\eta, \boldsymbol{\psi})$. As we derived above we have:

$$F(\eta, \boldsymbol{\psi}) = \Phi(h_{\eta_0}(\eta, \boldsymbol{\psi}))$$

The domain of $\Phi(\cdot)$ is the real line. Therefore, η has to be in the domain of the `hpsi` functions $h_{\eta_0}(\eta, \boldsymbol{\psi})$ as given in Equation (A.7) (and in Equations A.1 and A.4). We derived the following implementation of the both tail modified probit link called "psi12PROBIT".

```r
psi12PROBIT<-function(psi1 = 1, psi2 = 1, eta0 = 0)
{linkfun <- function(mu) {hpsi12INV(psi1, psi2, qnorm(mu), eta0)}
 linkinv <- function(eta){pnorm(hpsi12(psi1, psi2, eta, eta0))}
 mu.eta <- function(eta) {hpsi12DERIV(psi1, psi2, eta, eta0)*
             dnorm(hpsi12(psi1, psi2, eta, eta0))}
 valideta<-function(eta) {h <- 1:length(eta)
 for (i in 1:length(eta) ) {
    if (is.finite(linkinv(eta[i]))) {h[i] <- TRUE}
    else {h[i] <- FALSE}
    }
 h
 }
 link <- paste("psi12PROBIT(", psi1, "," , psi2, " , " , eta0, ")")
 structure(list(linkfun = linkfun,
                linkinv = linkinv,
                mu.eta = mu.eta,
                valideta = valideta,
                name = link),
           class = "link-glm")
}
```

Example 6.27 (Rotifer suspension)
Now we present the output of a both tail modified probit regression with two different parameters. As in Example 5.4 our analysis is based on the data frame `rotifer.ex` (see Example 4.4). The first parameter is chosen in a way that decreases the deviance drastically (for further information see Czado (2007)(pp. 18-22)). The other parameter vector is chosen to be (1,1) (i.e. $\psi = (1,1)$), which yields to an ordinary GLM as in Example 5.4 with `family` = "binomial" and `link` = "probit" (see also Remark 6.1).

```r
rotifer_glm_both<- glm(formula = cbind(y, n-y)~species
                                            +density.cen
                                            +species*density.cen,
                    family = binomial(link=psi12PROBIT(psi1=0,
                                                       psi2=-0.5,
                                                       eta0=0)),
                    data = rotifer.ex)
summary(rotifer_glm_both)

##
## Call:
## glm(formula = cbind(y, n - y) ~ species + density.cen + species *
##     density.cen, family = binomial(link = psi12PROBIT(psi1 = 0,
##     psi2 = -0.5, eta0 = 0)), data = rotifer.ex)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -5.896  -1.392  -0.065   1.740   4.450
##
## Coefficients:
##                    Estimate Std. Error z value Pr(>|z|)
## (Intercept)          -2.606      0.264   -9.87  < 2e-16 ***
## species               3.510      0.290   12.09  < 2e-16 ***
## density.cen           2.729      0.197   13.82  < 2e-16 ***
## species:density.cen  -1.230      0.232   -5.29  1.2e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 3180.99  on 39  degrees of freedom
## Residual deviance:  253.58  on 36  degrees of freedom
## AIC: 416.1
##
## Number of Fisher Scoring iterations: 7

summary(rotifer_glm_both)$deviance

## [1] 253.6
```

Remark 6.28 (No transformation for $\psi = (1,1)$)

```
rotifer_glm_both2<- glm(formula = cbind(y, n-y)~species
                                        +density.cen
                                        +species*density.cen,
                        family = binomial(link=psi12PROBIT(psi1=1,
                                                            psi2=1,
                                                            eta0=0)),
                        data = rotifer.ex)
summary(rotifer_glm_both2)$deviance

## [1] 471.3
```

Setting the link parameter ψ to (1,1) yields to no transformation (see Remark 6.1). Hence, the summary of `rotifer_glm_both2` yields to the same output as in Example 5.4. In particular, the deviance is the same, as can be seen above.

6.2.4   Poisson family with right tail modified log link

In the following we want to derive the right tail modified log link function. In a Poisson regression model with log link no modification other than the right tail modification is allowed.

Remark 6.29 (Poisson family with log link)
In Subsection 3.3.3 we defined the ordinary log link function through the following specifications:

(i) `linkfun`: $\eta = g(\mu) = \ln(\mu)$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = \exp\{\eta\}$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = \exp\{\eta\}$

(iv) `valideta`: $\mathbb{1}_{\{\eta \in \mathbb{R}\}} = \text{TRUE}$

We can define the right tail modified Poisson regression by modifying the linkinverse in the ordinary Poisson regression as follows.

Definition 6.30 (Right tail modified Poisson regression)
Let $h_{\eta_0}(\eta, \psi = \psi_1)$ be the right tail modification as defined in Definition 6.2. Then, the right tail modified Poisson regression has an inverse link function of the following form:

$$F(\eta, \psi = \psi_1) = \exp\{h_{\eta_0}(\eta, \psi = \psi_1)\} \tag{6.4}$$

Definition 6.31 (Right tail modified log link)
Using Equation (6.4) we can define the right tail modified log link through the following specifications:

(i) `linkfun`: $g(\mu, \boldsymbol{\psi} = \psi_1)$:
We had $\eta = g(\mu) = \ln(\mu)$ for the log link function. Hence, now we have:

$$h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1) = \ln(\mu)$$
$$\Leftrightarrow \eta = (h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1))^{-1}(\ln(\mu))$$
$$\Rightarrow g(\mu, \boldsymbol{\psi} = \psi_1) = (h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1))^{-1}(\ln(\mu))$$

(ii) `linkinv`: $F(\eta, \boldsymbol{\psi} = \psi_1) = \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)\}$

(iii) `mu.eta`: $\frac{\partial \mu}{\partial \eta} = \frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi} = \psi_1)$:

$$\frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi} = \psi_1) = \frac{\partial}{\partial \eta}\left(\exp\{h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)\}\right)$$
$$= \left(\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)\right)\exp\{h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)\}$$

(iv) `valideta`: TRUE, if η is in the domain of $F(\eta, \boldsymbol{\psi} = \psi_1)$.

(v) `name`: "psi1POISS" for the right tail modification.

Remark 6.32
In the setting given above we speak of a *right tail modified Poisson regression*.

Example 6.33 (Right tail modified Poisson regression)
Let the right tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ be given as in Definition 6.2, then the inverse of it is given as calculated in Equation (A.2) and the partial derivative w.r.t. η is given as calculated in Equation (A.3). We determine the values, which are in the domain of $F(\eta, \boldsymbol{\psi} = \psi_1)$. As we derived above we have:

$$F(\eta, \boldsymbol{\psi} = \psi_1) = \exp\{h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)\}$$

The domain of $\exp\{\cdot\}$ is the real line. Therefore, η has to be in the domain of the `hpsi` function $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ as given in Equation (A.1). We derived the following implementation of the right tail modified log link called "psi1POISS".

```
psi1POISS<-function(psi1 = 1, eta0 = 0)
{linkfun<-function(mu){hpsi1INV(psi1, log(mu), eta0)}
 linkinv<-function(eta){exp(hpsi1(psi1, eta, eta0))}
 mu.eta<-function(eta) {exp(hpsi1(psi1, eta, eta0))*
              hpsi1DERIV(psi1, eta, eta0)}
 valideta<-function(eta) {h <- 1:length(eta)
 for (i in 1:length(eta) ) {
     if (is.finite(linkinv(eta[i]))) {h[i] <- TRUE}
```

```r
    else {h[i] <- FALSE}
    }
 h
 }
 link <- paste("psi1POISS(", psi1, " , " , eta0, ")")
 structure(list(linkfun = linkfun,
                linkinv = linkinv,
                mu.eta = mu.eta,
                valideta = valideta,
                name = link),
          class = "link-glm")
}
```

Example 6.34 (Coal mining fractures)
Now we present the output of the right tail modified Poisson regression with two differ-
ent parameters. As in Example 5.5 our analysis is based on the data frame `mining.ex`
(see Example 4.5). The first parameter is chosen in a way that decreases the deviance
drastically (for further information see Example 7.3). For the other parameter we choose
$\psi_1 = 1$. This yields to an ordinary GLM with `family` = "poisson" and `link` = "log".

```r
mining_glm_right<- glm(formula = y~inb.cen+extrp.cen,
                       family = poisson(link=psi1POISS(psi1=-0.57,
                                                       eta0=0)),
                       data = mining.ex)
summary(mining_glm_right)

##
## Call:
## glm(formula = y ~ inb.cen + extrp.cen,
##     family = poisson(link = psi1POISS(psi1 = -0.57, eta0 = 0)),
##     data = mining.ex)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.999  -0.633  -0.171   0.419   2.103
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.09432    1.02158    3.03   0.0025 **
## inb.cen     -0.01018    0.00403   -2.53   0.0115 *
## extrp.cen    0.35944    0.11124    3.23   0.0012 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
```

```
##
##      Null deviance: 74.984  on 43  degrees of freedom
## Residual deviance: 30.757  on 41  degrees of freedom
## AIC: 133
##
## Number of Fisher Scoring iterations: 17

summary(mining_glm_right)$deviance

## [1] 30.76
```

Remark 6.35 (No transformation for $\psi_1 = 1$)
```
mining_glm_right2<- glm(formula = y~inb.cen+extrp.cen,
                        family = poisson(link=psi1POISS(psi1=1,
                                                        eta0=0)),
                        data = mining.ex)
summary(mining_glm_right2)$deviance

## [1] 42.09
```

Setting the link parameter ψ_1 to 1 yields to no transformation (see Remark 6.1). Hence, the summary of `mining_glm_right2` yields to the same output as in Example 5.5. In particular, the deviance is the same, as can be seen above.

6.2.5 Gamma family with left tail modified inverse link

In the following we want to derive the left tail modified inverse link function. In a gamma regression model with inverse link no modification other than the left tail modification is allowed.

Remark 6.36 (Gamma family with inverse link)
In Subsection 3.3.4 we defined the ordinary inverse link function through the following specifications:

(i) `linkfun`: $\eta = g(\mu) = \frac{1}{\mu}$

(ii) `linkinv`: $\mu = g^{-1}(\eta) = F(\eta) = \frac{1}{\eta}$

(iii) `mu.eta`: $\frac{d\mu}{d\eta}(F(\eta)) = -\frac{1}{\eta^2}$

(iv) `valideta`: $\mathbb{1}_{\{\{\eta \in \mathbb{R}\} \cap \{\eta \neq 0\}\}}$

We can define the left tail modified gamma regression by modifying the linkinverse in the ordinary gamma regression as follows.

Definition 6.37 (Left tail modified gamma regression)
Let $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ be the left tail modification as defined in Definition 6.3. Then, the left tail modified gamma regression has an inverse link function of the following form:

$$F(\eta, \boldsymbol{\psi} = \psi_2) = \frac{1}{h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)} \tag{6.5}$$

Definition 6.38 (Left tail modified inverse link)
Using Equation (6.5) we can define the left tail modified inverse link through the following specifications:

(i) `linkfun`: $g(\mu, \boldsymbol{\psi} = \psi_2)$:
   We had $\eta = g(\mu) = \frac{1}{\mu}$ for the inverse link function. Hence, now we have:

$$h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2) = \frac{1}{\mu}$$

$$\Leftrightarrow \eta = (h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2))^{-1} \left(\frac{1}{\mu}\right)$$

$$\Rightarrow g(\mu, \boldsymbol{\psi} = \psi_2) = (h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2))^{-1} \left(\frac{1}{\mu}\right)$$

(ii) `linkinv`: $F(\eta, \boldsymbol{\psi} = \psi_2) = \frac{1}{h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)}$

(iii) `mu.eta`: $\frac{\partial \mu}{\partial \eta} = \frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi} = \psi_2)$:

$$\frac{\partial}{\partial \eta} F(\eta, \boldsymbol{\psi} = \psi_2) = \frac{\partial}{\partial \eta} \frac{1}{h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)}$$

$$= -\left(\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)\right) \frac{1}{(h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2))^2}$$

(iv) `valideta`: TRUE, if η is in the domain of $F(\eta, \boldsymbol{\psi} = \psi_2)$.

(v) `name`: "psi2GAMMA" for the left tail modification.

Remark 6.39
In the setting given above we speak of a *left tail modified gamma regression*.

Example 6.40 (Left tail modified gamma regression)
Let the left tail modification $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ be given as in Definition 6.3, then the inverse of it is given as calculated in Equation (A.5) and the partial derivative w.r.t. η is given as calculated in Equation (A.6). We determine the values, which are in the domain of $F(\eta, \boldsymbol{\psi} = \psi_2)$. As we derived above we have:

$$F(\eta, \boldsymbol{\psi} = \psi_2) = \frac{1}{h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)}$$

Therefore, the domain of $F(\eta, \boldsymbol{\psi} = \psi_2)$ is the domain of the $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ except for the values of η, where $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2) = 0$. The domain of the corresponding `hpsi` function is given in Equation (A.4). We derived the following implementation of the left tail modified inverse link called "psi2GAMMA".

```
psi2GAMMA<-function(psi2 = 1, eta0 = 0)
{linkfun <- function(mu) hpsi2INV(psi2, (1/mu), eta0)
 linkinv <- function(eta) 1/hpsi2(psi2, eta, eta0)
 mu.eta <- function(eta) {
  hpsi2DERIV(psi2, eta, eta0)*(- (1/hpsi2(psi2, eta, eta0))^2)
 }
 valideta<-function(eta) {h <- 1:length(eta)
 for (i in 1:length(eta) ) {
    if (is.finite(linkinv(eta[i]))) {h[i] <- TRUE}
    else {h[i] <- FALSE}
    }
  h
 }
 link <- paste("psi2GAMMA(", psi2, " , " , eta0, ")")
 structure(list(linkfun = linkfun,
                linkinv = linkinv,
                mu.eta = mu.eta,
                valideta = valideta,
                name = link),
           class = "link-glm")
}
```

Example 6.41 (Canadian automobile insurance)
Now we present the output of the left tail modified gamma regression with two different
parameters. As in Example 5.6 our analysis is based on the data frame `carinsur.ex`
(see Example 4.6). The first parameter is chosen in a way that decreases the deviance
drastically (for further information see Example 7.4). For the other parameter we choose
$\psi_2 = 1$, which yields to an ordinary GLM with `family` = "Gamma" and `link` = "inverse".

```
carinsur_glm_left<- glm(formula = (Cost/Claims)~Merit+Class,
                        family = Gamma(link=psi2GAMMA(psi2=-1.39,
                                                      eta0=3.6)),
                        weights = Claims,
                        start = c(3.2,0,0,0,-0.3,-0.1,-0.5,0.25),
                        data = carinsur.ex)
summary(carinsur_glm_left)

## 
## Call:
## glm(formula = (Cost/Claims) ~ Merit + Class,
##     family = Gamma(link = psi2GAMMA(psi2 = -1.39, eta0 = 3.6)),
##     data = carinsur.ex,
##     weights = Claims,
##     start = c(3.2, 0, 0, 0, -0.3, -0.1, -0.5, 0.25))
## 
```

```
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -6.108  -1.337   0.000   0.956   5.753
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  2.81e+00   2.68e+00    1.05     0.32
## Merit1       6.57e-01   3.02e+00    0.22     0.83
## Merit2       6.35e-01   3.18e+00    0.20     0.84
## Merit3       5.79e-01   2.72e+00    0.21     0.83
## Class2      -9.12e-01   6.34e+00   -0.14     0.89
## Class3      -9.16e-02   1.50e+00   -0.06     0.95
## Class4      -2.37e+04   5.07e+08    0.00     1.00
## Class5       3.56e-01   1.96e+00    0.18     0.86
##
## (Dispersion parameter for Gamma family taken to be 2050)
##
##     Null deviance: 1556.01  on 19  degrees of freedom
## Residual deviance:  122.19  on 12  degrees of freedom
## AIC: -3100206
##
## Number of Fisher Scoring iterations: 15

summary(carinsur_glm_left)$deviance

## [1] 122.2
```

Remark 6.42 (No transformation for $\psi_2 = 1$)
```
carinsur_glm_left2<- glm(formula = Cost/Claims~Merit+Class,
                    family = Gamma(link=psi2GAMMA(psi2=1,
                                                  eta0=3.6)),
                    weights = Claims,
                    start = c(3.2,0,0,0,-0.3,-0.1,-0.5,0.25),
                    data = carinsur.ex)
summary(carinsur_glm_left2)$deviance

## [1] 167.4
```

Setting the link parameter ψ_2 to 1 yields to no transformation (see Remark 6.1). Thus, the summary of `carinsur_glm_left2` yields to the same output as in Example 5.6. In particular, the deviance is the same, as can be seen above.

6.3 The `glmProfile` function

In the following we will define a function called "glmProfile". It was also presented in Czado (2007) and can be seen as an extension to the ordinary `glm` function in R.

Remark 6.43 (Motivation of the `glmProfile` function)
For every single, appropriate value of $\boldsymbol{\psi}$ we can (given a value for η_0) define a user-defined link function. For each of these user-defined link functions we can run the `glm` function (see examples in Section 6.2). For each value of $\boldsymbol{\psi}$ specified through the arguments `psi1` and `psi2`, we calculate the deviance using a `glm` function with our user-defined link. Now we want to investigate, which value we should take for the parameter $\boldsymbol{\psi}$ in order to receive the best-fitting model (w.r.t. the deviance). Therefore, we would like to take the parameter of $\boldsymbol{\psi}$ corresponding to the model with the minimal deviance. The parameter we should take can be determined with the `glmProfile` function, which we will define now. We can also see a plot describing the decrease of the deviance graphically.

Definition 6.44 (Definition of the `glmProfile` function)
The `glmProfile` function requires the following input:

(i) `formula`: is an expression of the form: response $\sim \; covariate_1 + \cdots + covariate_k$ (as in the ordinary `glm` function). For every data set we have to decide which covariates we want to take in our model.

(ii) values for the parameter $\boldsymbol{\psi}$:

- for a single tail modification we have to specify either ψ_1 (for a right tail modification) or ψ_2 (for a left tail modification). For the single right tail modification (single left tail modification) a grid of ψ_1 values (ψ_2 values) has to be specified in the vector `psi1` (vector `psi2`) which we pass to the `glmProfile` function. The default values for both `psi1` and `psi2` are 1, which yields to no transformation.

- for a both tail modification we have to specify both ψ_1 and ψ_2 in the corresponding vectors `psi1` and `psi2`. This gives a grid of $\boldsymbol{\psi} = (\psi_1, \psi_2)$ values.

(iii) value for η_0: for every kind of modification we have to declare a value for η_0 in the variable `eta0`. If no value is given in the command line, then η_0 is set to 0 by default.

(iv) `family`: we can specify one of the families introduced in Section 3.3. Since we modified both the logistic regression and the probit regression, we have to make little changes stating the family. One of the following families is possible (in the brackets we give the numbers corresponding to this family in the procedure of `glmProfile`):

- `family = "gaussian"` (internally handled with `familyinteger = 1`). This family is set to be the default family as in the ordinary `glm` function in R.

- family = "poisson" (internally handled with `familyinteger = 2`).

- `family = "logit"` (internally handled with `familyinteger = 3`). This statement is slightly different from the specification in the ordinary `glm` function (there it would have been `family = binomial(link = "logit")`).

- `family` = "probit" (internally handled with `familyinteger` = 4).
- `family` = "Gamma" (internally handled with `familyinteger` = 5).

(v) `tail`: we have to specify what tail transformation we like to perform. Table 6.1 gives an overview of allowed tail modifications. The following specifications are possible (again we give the internal numbers corresponding to this transformation):

- `tail` = "left" (internally handled with `tailinteger` = 1) corresponding to Def. 6.3. This modification is as default (i.e. when no specification of `tail` is assigned).
- `tail` = "right" (internally handled with `tailinteger` = 2) corresponding to Def. 6.2.
- `tail` = "both" (internally handled with `tailinteger` = 3) corresponding to Def. 6.4.

(vi) `weights`: optional statement with default value NULL. It contains the "prior weights" to be used in the fitting process as in the ordinary `glm` function in R (see `help(glm)`).

(vii) `start`: optional statement with default value NULL. It contains starting values for the parameters in the linear predictor as in the ordinary `glm` function in R (see `help(glm)`). Thus, we have to specify a vector of length p = k+1 (since $\boldsymbol{\eta}$ is depending on $\boldsymbol{\beta} \in \mathbb{R}^p$, see Equation (3.5)). We will specify a vector, which is close to the estimates of the regression parameters in the ordinary GLM (see Chapter 5).

(viii) `data`: data set containing n rows of observations of both the response and the covariates as for the ordinary `glm` function.

If we specify all of the required arguments of the `glmProfile` function we get the following output:

(i) <u>plot</u>:

- for a single tail modification: a deviance profile plot for the link parameter. Therefore we plot the link parameter (either ψ_1 or ψ_2) versus the corresponding deviance for these values. We also include an approximate 95% confidence interval for the link parameter (printed with dotted lines).
- for a both tail modification: a deviance profile contour plot for the link parameters. Therefore we plot the link parameter ψ_1 on the x-axis and ψ_2 and the y-axis together with the deviance surface for the grid of (ψ_1, ψ_2) values. In addition we include an approximate 95% joint confidence region for the link parameters (ψ_1, ψ_2) (printed with dotted lines).

(ii) <u>computed values</u>: the `glmProfile` function returns a list containing the values of $\boldsymbol{\psi}$ which we can access by `$psi` for a single tail modification (returning either the values of ψ_1 for a right tail modification or those of ψ_2 for a left tail modification) and by `$psi1` and `$psi2` for a both tail modification. The list also contains the deviance for the values of $\boldsymbol{\psi}$ which we can access by `$dev`.

(iii) <u>summarizing statement</u>: furthermore the **glmProfile** function gives the minimum
deviance and the value of ψ for which it was attained.

```r
glmProfile<-function(formula = formula(data),
                     psi1 = 1, psi2 = 1, eta0 = 0,
                     family = "gaussian",
                     tail = "left",
                     weights = NULL,
                     start = NULL,
                     data = stop("data missing")){
w <- weights
if(any(w < 0)){stop("negative weights not allowed")}
familyinteger <- charmatch(family, c("gaussian",
                                     "poisson",
                                     "logit",
                                     "probit",
                                     "Gamma"))
tailinteger <- charmatch(tail, c("left",
                                 "right",
                                 "both"))

 ##############################################################
 # Gaussian: all modifications allowed  #####################
 ##############################################################
 if(tailinteger == 1 & familyinteger == 1){
  psi <- psi2
  linkset <- psi2GAUSS
  familyname<<-gaussian
 }
 else if(tailinteger == 2 & familyinteger == 1){
  psi <- psi1
  linkset <- psi1GAUSS
  familyname<<-gaussian
 }
 else if(tailinteger == 3 & familyinteger == 1){
  linkset <- psi12GAUSS
  familyname<<-gaussian
 }

 ##############################################################
 # Poisson: only right modification allowed  ################
 ##############################################################
 else if(tailinteger == 2 & familyinteger == 2){
  psi <- psi1
  linkset <- psi1POISS
```

```
 familyname<<-poisson
}


############################################################
# Logit: all modifications allowed  ########################
############################################################
else if(tailinteger == 1 & familyinteger == 3){
 psi <- psi2
 linkset <- psi2LOGIT
 familyname<<-binomial
}
else if(tailinteger == 2 & familyinteger == 3){
 psi <- psi1
 linkset <- psi1LOGIT
 familyname<<-binomial
}
else if(tailinteger == 3 & familyinteger == 3){
 linkset <- psi12LOGIT
 familyname<<-binomial
}


############################################################
# Probit: all modifications allowed  #######################
############################################################
else if(tailinteger == 1 & familyinteger == 4){
 psi <- psi2
 linkset <- psi2PROBIT
 familyname<<-binomial
}
else if(tailinteger == 2 & familyinteger == 4){
 psi <- psi1
 linkset <- psi1PROBIT
 familyname<<-binomial
}
else if(tailinteger == 3 & familyinteger == 4){
 linkset <- psi12PROBIT
 familyname<<-binomial
}


############################################################
# Gamma: only left modification allowed  ###################
############################################################
else if(tailinteger == 1 & familyinteger == 5){
psi <- psi2
linkset <- psi2GAMMA
```

```
familyname<<-Gamma
}

else{stop(paste(tail, "tail modification is not available for the",
                family, "family"))}
Weights <<- w
#########################################
# BOTH TAIL MODIFICATIONS                #
#########################################
if(tailinteger == 3){
 psi1 <- sort(psi1)
 psi2 <- sort(psi2)
 r <- matrix(0, nrow = length(psi1), ncol = length(psi2))
 for(j in 1:length(psi1)){
  for(k in 1:length(psi2)){
   r[j, k] <- glm(formula,
           family = familyname(link = linkset(psi1[j],
                                               psi2[k],
                                               eta0)),
                            weights = Weights, start=start,
                            data = data)$deviance
  }
 }
 dimnames(r) <- list(as.character(psi1), as.character(psi2))
 out <- list(psi1 = psi1, psi2 = psi2, dev = r)
 rpsi1 <- range(psi1)
 rpsi2 <- range(psi2)
 par(oma = c(2, 0, 0, 0))
 contour(psi1, psi2, r, xlim = rpsi1, ylim = rpsi2,
         xlab = "psi1", ylab = "psi2")
 title(main = "Deviance Profile Contour Plot for the Link Parameters",
       sub = paste(family, "GLM with",
                   tail, "tails modified with 95% confidence region"))
 sortr <- sort.list(r)[1]
 sortpsi2 <- floor(sortr/length(psi1))
 temp <- sortr - sortpsi2 * length(psi1)
 if(temp == 0) {sortpsi2 <- sortpsi2 - 1}
 sortpsi1 <- sortr - sortpsi2 * length(psi1)
 sortpsi2 <- ceiling(sortr/length(psi1))
 cat(paste("\nminimum deviance=", min(r), "attained for",
           "psi1=", psi1[sortpsi1], ",psi2=", psi2[sortpsi2],".\n\n"))
 conlvl <- min(r) + qchisq(0.95, 2)

 if(conlvl < max(r)) {
  par(new = T)
```

```r
  contour(psi1, psi2, r, xlim = rpsi1, ylim = rpsi2,
          levels = conlvl, lty = 2)
  par(new = T)
  plot(psi1[sortpsi1],
       psi2[sortpsi2],
       xlim = rpsi1, ylim = rpsi2)
 }
}
##########################################
# SINGLE TAIL MODIFICATIONS             #
##########################################
else {
  psi <- sort(psi)
  r <- rep(0, length(psi))
  for(j in 1:length(psi)){
    r[j] <- glm(formula,
                family = familyname(link = linkset(psi[j], eta0)),
                weights = Weights, start=start,
                data = data)$deviance
}

out <- list(psi = psi, dev = r)
ry <- range(r, min(r) - (max(r) - min(r)) * 0.05)
plot(psi, r, ylim = ry, ylab = "deviance", xlab = "psi", type = "l",
main = "Deviance Profile Plot for the Link Parameter",
sub = paste(family, "GLM with",
            tail, "tail modified with 95% confidence interval"))
abline(h = min(r) + 2, lty = 2)
devcon <- min(r) + 2
rlow <- r[psi < psi[sort.list(r)[1]]] - devcon
rhigh <- r[psi > psi[sort.list(r)[1]]] - devcon
if(max(rlow) > 0) {
 psilow <- sort.list(abs(rlow))[1]
 segments(psi[psilow], ry[1], psi[psilow], devcon, lty = 2)
}
if(max(rhigh) > 0) {
 psihigh <- sort.list(abs(rhigh))[1]
 segments(psi[psihigh + sort.list(r)[1]], ry[1], psi[
    psihigh + sort.list(r)[1]], devcon, lty = 2)
 }
 cat(paste("\nminimum deviance=", min(r),
           "attained for", psi[sort.list(r)[1]], ".\n\n"))
}
 out
}
```

Remark 6.45 (Interpretation of the 95% - confidence interval)
The 95% - confidence interval allows us to assess whether the link modification is necessary. We saw when setting ψ_1, $\psi_2 = 1$ or $\boldsymbol{\psi} = (1, 1)$ we obtain no transformation (see Remark 6.1) and thus we use the same link function as in the ordinary GLM. If consequently these values (i.e. $\boldsymbol{\psi} = (1, 1)$ or ψ_1, $\psi_2 = 1$) lie within the confidence interval (or confidence region, respectively) no tail modification would have been necessary, i.e. the analysis of the ordinary `glm` function suffices. In these cases the decrease of the deviance is only weak. Whenever $\psi_1 = 1$ is not contained in the confidence interval given in the *deviance profile plot* a right tail modification seems appropriate. Analogously if $\psi_2 = 1$ is not contained in the confidence interval, then a left tail modification yields to an improved fit of the model. The same results can be adopted for the both tail modifications.

Remark 6.46 (Starting value for η_0)
According to Czado (2007) the following choices for the starting value η_0 are reasonable:

- it is advisable to choose $\eta_0 = \beta_0$ using centered covariates, varying around their center β_0.

- if in a binary regression the data sets show symmetric proportions around 0.5 one can take $\eta_0 = 0$.

- if the proportions are not symmetric one has to estimate the success probability p_0. This can be done via the observed proportions for discrete random variables. For continuous random variables one should group the data primarily.

In Chapter 7 we will present the output of the `glmProfile` function. We restrict ourselves to apply the `glmProfile` function only on the examples for the single tail modifications with the default link functions. All in all we will derive the parameters for the parametric link by applying the `glmProfile` function on the following data sets and link functions:

Family	Link	Data set	As seen in	Ordinary GLM
Gaussian	identity	pcb.ex	Example 4.1	Example 5.1
Binomial	logit	beetle.ex	Example 4.2	Example 5.2
Poisson	log	mining.ex	Example 4.5	Example 5.5
Gamma	inverse	carinsur.ex	Example 4.6	Example 5.6

Table 6.2: Overview: data sets on which `glmProfile` will be applied.

Remark 6.47 (Further examples)
Two other examples were discussed extensively in Czado (2007). From this technical report we took the optimal values for the corresponding parameters for the tail modifying link functions (see Example 6.20 and Example 6.27).
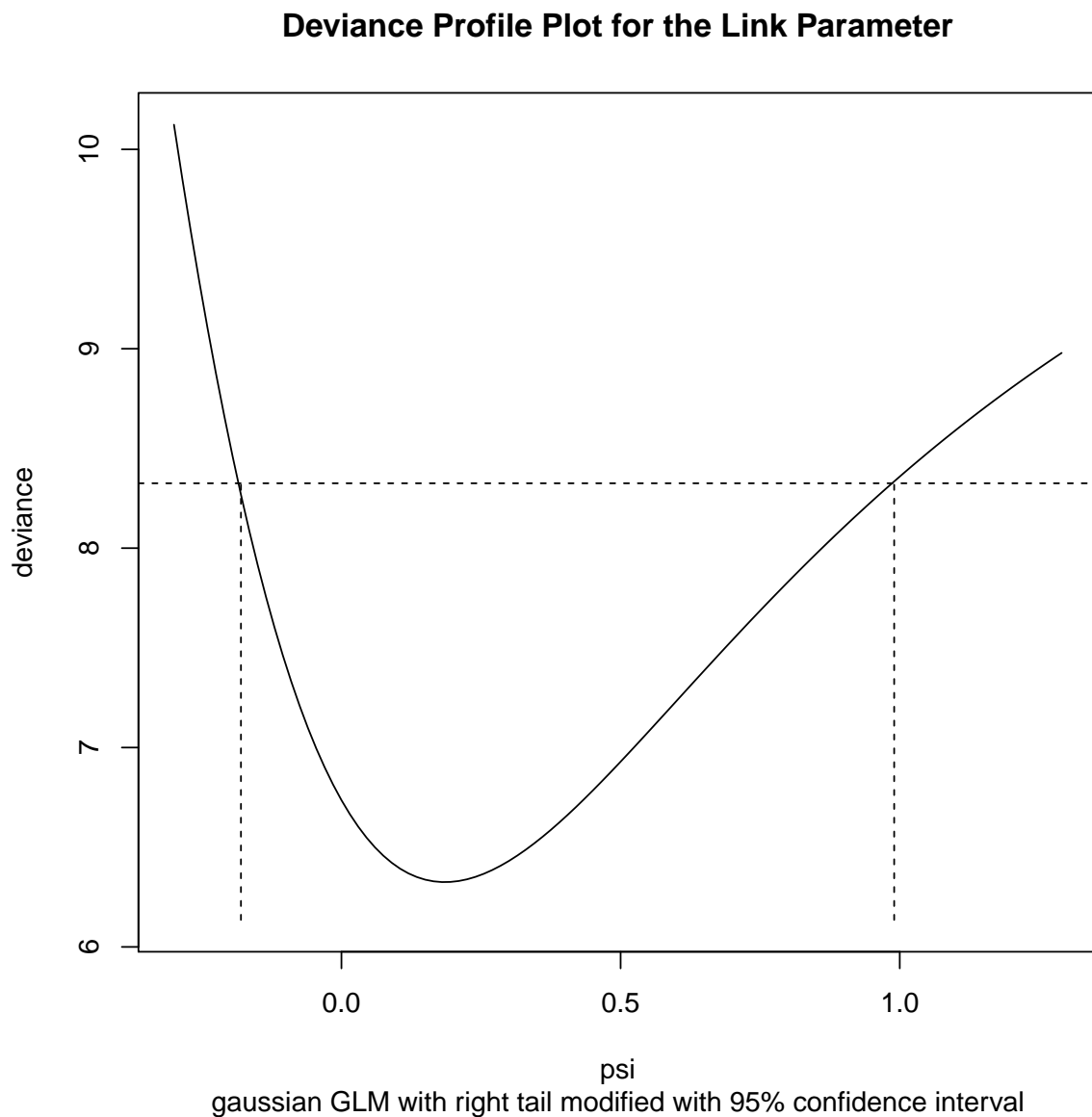
7   Examples: the `glmProfile` function

7.1   Modified Gaussian regression

Example 7.1 (Applying `glmProfile` on PCB data)
Now we derive the optimal value for the parameter ψ_1 we have used in the right tail modified Gaussian regression (see Example 6.11). Therefore we apply the `glmProfile` function on the data frame `pcb.ex` with the following specifications:

```
glmProfile(formula = log.pcb~age.cen,
           psi1 = seq(-0.3, 1.3, by=0.015),
           family = "gaussian",
           tail = "right",
           data = pcb.ex)
```

Deviance Profile Plot for the Link Parameter

psi
gaussian GLM with right tail modified with 95% confidence interval

```
##
## minimum deviance= 6.32526546684224 attained for 0.18 .
## $psi
##   [1] -0.300 -0.285 -0.270 -0.255 -0.240 -0.225 -0.210 -0.195 -0.180
##  [10] -0.165 -0.150 -0.135 -0.120 -0.105 -0.090 -0.075 -0.060 -0.045
##  [19] -0.030 -0.015  0.000  0.015  0.030  0.045  0.060  0.075  0.090
##  [28]  0.105  0.120  0.135  0.150  0.165  0.180  0.195  0.210  0.225
##  [37]  0.240  0.255  0.270  0.285  0.300  0.315  0.330  0.345  0.360
##  [46]  0.375  0.390  0.405  0.420  0.435  0.450  0.465  0.480  0.495
##  [55]  0.510  0.525  0.540  0.555  0.570  0.585  0.600  0.615  0.630
##  [64]  0.645  0.660  0.675  0.690  0.705  0.720  0.735  0.750  0.765
##  [73]  0.780  0.795  0.810  0.825  0.840  0.855  0.870  0.885  0.900
##  [82]  0.915  0.930  0.945  0.960  0.975  0.990  1.005  1.020  1.035
##  [91]  1.050  1.065  1.080  1.095  1.110  1.125  1.140  1.155  1.170
## [100]  1.185  1.200  1.215  1.230  1.245  1.260  1.275  1.290
##
## $dev
##   [1] 10.124  9.852  9.592  9.343  9.106  8.879  8.665  8.461  8.268
##  [10]  8.086  7.915  7.754  7.603  7.462  7.332  7.210  7.098  6.995
##  [19]  6.900  6.814  6.736  6.666  6.603  6.548  6.500  6.458  6.422
##  [28]  6.393  6.369  6.350  6.337  6.329  6.325  6.326  6.331  6.339
##  [37]  6.351  6.367  6.385  6.407  6.431  6.458  6.487  6.519  6.552
##  [46]  6.587  6.624  6.662  6.701  6.742  6.784  6.827  6.870  6.914
##  [55]  6.959  7.004  7.049  7.095  7.141  7.187  7.233  7.279  7.326
##  [64]  7.371  7.417  7.463  7.508  7.553  7.598  7.643  7.687  7.731
##  [73]  7.774  7.817  7.860  7.902  7.943  7.985  8.025  8.066  8.106
##  [82]  8.145  8.184  8.222  8.260  8.298  8.335  8.371  8.407  8.443
##  [91]  8.478  8.512  8.547  8.580  8.614  8.646  8.679  8.711  8.742
## [100]  8.773  8.804  8.834  8.864  8.893  8.922  8.951  8.979
```

7.2   Modified logistic regression

Example 7.2 (Applying `glmProfile` on beetle data)
Now we derive the optimal value for the parameter ψ_2 we have used in the left tail modified logistic regression (see Example 6.18). Therefore we apply the `glmProfile` function on the data frame `beetle.ex` with the following specifications:

```
glmProfile(formula = cbind(y, n-y)~dose.cen,
           psi2 = seq(-0.4, 0.8, by=0.01), eta0=0,
           family = "logit",
           tail = "left",
           data = beetle.ex)
```

Deviance Profile Plot for the Link Parameter

psi
logit GLM with left tail modified with 95% confidence interval
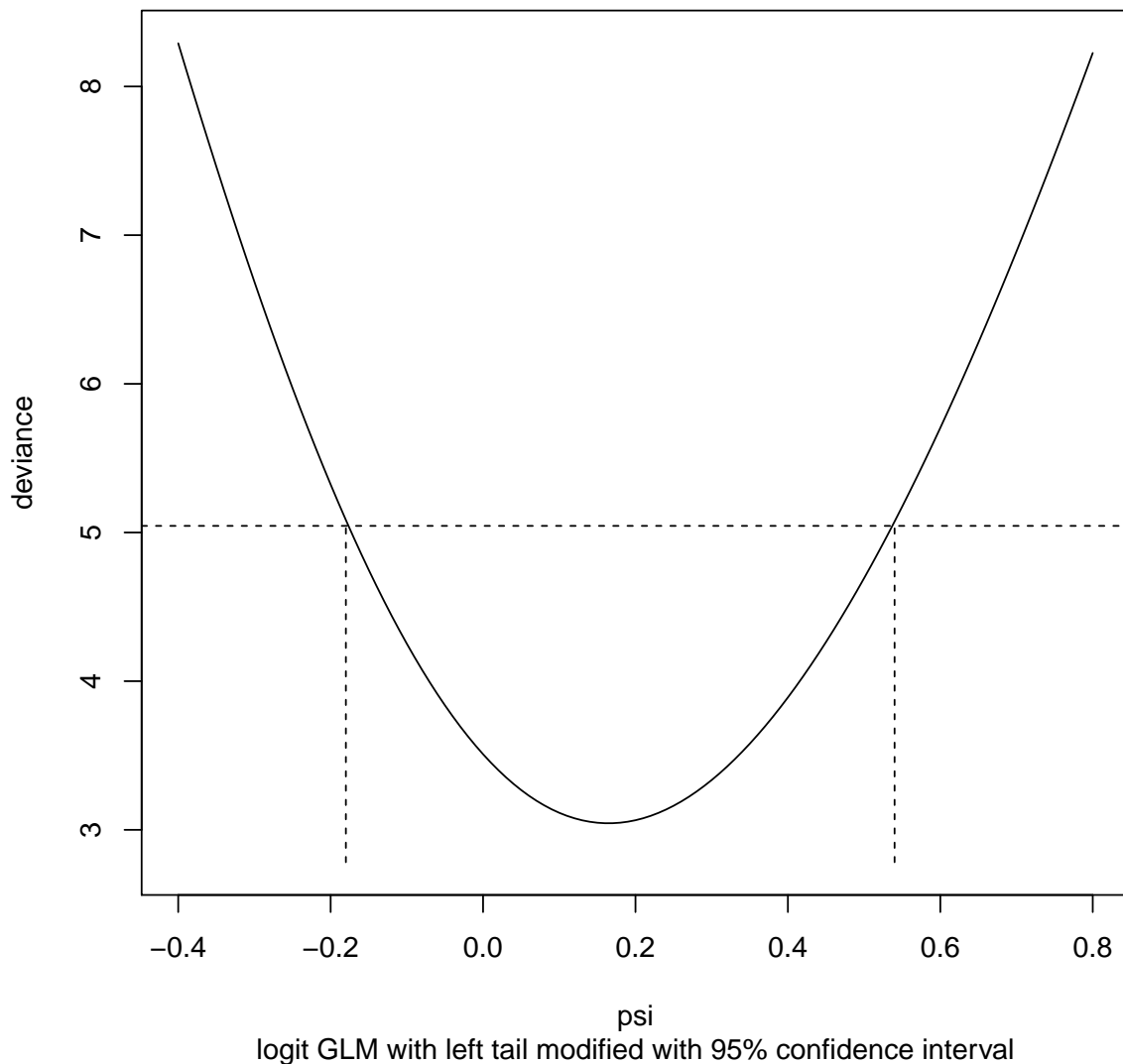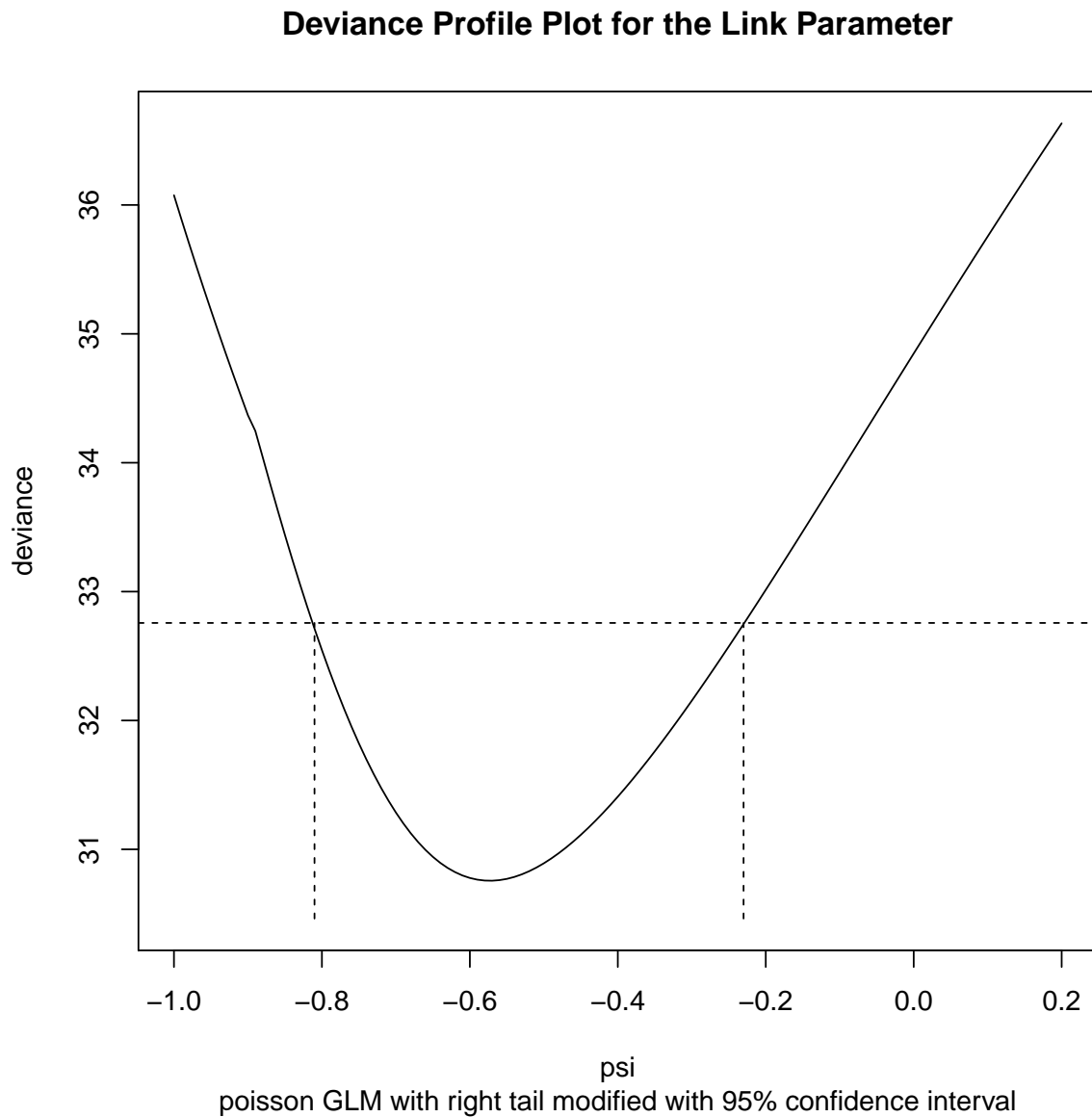
```
##
## minimum deviance= 3.04454405636305 attained for 0.16 .
## $psi
##   [1] -0.40 -0.39 -0.38 -0.37 -0.36 -0.35 -0.34 -0.33 -0.32 -0.31
##  [11] -0.30 -0.29 -0.28 -0.27 -0.26 -0.25 -0.24 -0.23 -0.22 -0.21
##  [21] -0.20 -0.19 -0.18 -0.17 -0.16 -0.15 -0.14 -0.13 -0.12 -0.11
##  [31] -0.10 -0.09 -0.08 -0.07 -0.06 -0.05 -0.04 -0.03 -0.02 -0.01
##  [41]  0.00  0.01  0.02  0.03  0.04  0.05  0.06  0.07  0.08  0.09
##  [51]  0.10  0.11  0.12  0.13  0.14  0.15  0.16  0.17  0.18  0.19
##  [61]  0.20  0.21  0.22  0.23  0.24  0.25  0.26  0.27  0.28  0.29
##  [71]  0.30  0.31  0.32  0.33  0.34  0.35  0.36  0.37  0.38  0.39
##  [81]  0.40  0.41  0.42  0.43  0.44  0.45  0.46  0.47  0.48  0.49
##  [91]  0.50  0.51  0.52  0.53  0.54  0.55  0.56  0.57  0.58  0.59
## [101]  0.60  0.61  0.62  0.63  0.64  0.65  0.66  0.67  0.68  0.69
## [111]  0.70  0.71  0.72  0.73  0.74  0.75  0.76  0.77  0.78  0.79
## [121]  0.80
##
## $dev
##   [1] 8.291 8.121 7.953 7.787 7.623 7.461 7.302 7.144 6.989 6.836
##  [11] 6.686 6.538 6.392 6.249 6.109 5.971 5.836 5.703 5.574 5.447
##  [21] 5.323 5.202 5.084 4.969 4.857 4.748 4.643 4.540 4.441 4.344
##  [31] 4.252 4.162 4.076 3.993 3.913 3.837 3.765 3.695 3.630 3.568
##  [41] 3.509 3.454 3.402 3.354 3.309 3.268 3.230 3.196 3.165 3.138
##  [51] 3.115 3.094 3.078 3.064 3.054 3.048 3.045 3.045 3.048 3.055
##  [61] 3.065 3.078 3.094 3.114 3.136 3.162 3.191 3.222 3.257 3.295
##  [71] 3.335 3.378 3.425 3.474 3.525 3.580 3.637 3.696 3.759 3.823
##  [81] 3.891 3.960 4.033 4.107 4.184 4.263 4.345 4.428 4.514 4.602
##  [91] 4.693 4.785 4.879 4.976 5.074 5.174 5.277 5.381 5.487 5.595
## [101] 5.704 5.815 5.928 6.043 6.160 6.278 6.397 6.519 6.641 6.766
## [111] 6.891 7.019 7.147 7.277 7.409 7.542 7.676 7.811 7.948 8.086
## [121] 8.225
```

7.3   Modified Poisson regression

Example 7.3 (Applying `glmProfile` on mining data)
Now we derive the optimal value for the parameter ψ_1 we have used in the right tail modified Poisson regression (see Example 6.34). Therefore we apply the `glmProfile` function on the data frame `mining.ex` with the following specifications:

```
glmProfile(formula = y~inb.cen+extrp.cen,
           psi1 = seq(-1, 0.2, by=0.01), eta0=0,
           family = "poisson",
           tail = "right",
           data = mining.ex)
```

Deviance Profile Plot for the Link Parameter

poisson GLM with right tail modified with 95% confidence interval

```
##
## minimum deviance= 30.7566766282853 attained for -0.57 .
## $psi
##    [1] -1.00 -0.99 -0.98 -0.97 -0.96 -0.95 -0.94 -0.93 -0.92 -0.91
##   [11] -0.90 -0.89 -0.88 -0.87 -0.86 -0.85 -0.84 -0.83 -0.82 -0.81
##   [21] -0.80 -0.79 -0.78 -0.77 -0.76 -0.75 -0.74 -0.73 -0.72 -0.71
##   [31] -0.70 -0.69 -0.68 -0.67 -0.66 -0.65 -0.64 -0.63 -0.62 -0.61
##   [41] -0.60 -0.59 -0.58 -0.57 -0.56 -0.55 -0.54 -0.53 -0.52 -0.51
##   [51] -0.50 -0.49 -0.48 -0.47 -0.46 -0.45 -0.44 -0.43 -0.42 -0.41
##   [61] -0.40 -0.39 -0.38 -0.37 -0.36 -0.35 -0.34 -0.33 -0.32 -0.31
##   [71] -0.30 -0.29 -0.28 -0.27 -0.26 -0.25 -0.24 -0.23 -0.22 -0.21
##   [81] -0.20 -0.19 -0.18 -0.17 -0.16 -0.15 -0.14 -0.13 -0.12 -0.11
##   [91] -0.10 -0.09 -0.08 -0.07 -0.06 -0.05 -0.04 -0.03 -0.02 -0.01
## [101]  0.00  0.01  0.02  0.03  0.04  0.05  0.06  0.07  0.08  0.09
## [111]  0.10  0.11  0.12  0.13  0.14  0.15  0.16  0.17  0.18  0.19
## [121]  0.20
##
## $dev
##    [1] 36.08 35.89 35.71 35.53 35.35 35.18 35.01 34.85 34.69 34.53
##   [11] 34.37 34.25 34.04 33.84 33.64 33.44 33.25 33.06 32.89 32.71
##   [21] 32.55 32.39 32.23 32.09 31.95 31.82 31.70 31.58 31.48 31.38
##   [31] 31.29 31.20 31.13 31.06 31.00 30.94 30.90 30.86 30.82 30.80
##   [41] 30.78 30.76 30.76 30.76 30.76 30.77 30.79 30.81 30.83 30.86
##   [51] 30.89 30.93 30.97 31.01 31.06 31.11 31.17 31.22 31.28 31.34
##   [61] 31.41 31.48 31.54 31.61 31.69 31.76 31.84 31.91 31.99 32.07
##   [71] 32.15 32.24 32.32 32.40 32.49 32.57 32.66 32.75 32.84 32.92
##   [81] 33.01 33.10 33.19 33.28 33.37 33.47 33.56 33.65 33.74 33.83
##   [91] 33.92 34.02 34.11 34.20 34.29 34.39 34.48 34.57 34.66 34.76
## [101] 34.85 34.94 35.03 35.12 35.21 35.30 35.40 35.49 35.58 35.67
## [111] 35.76 35.84 35.93 36.02 36.11 36.20 36.29 36.37 36.46 36.55
## [121] 36.63
```

7.4   Modified gamma regression

Example 7.4 (Applying `glmProfile` on car insurance data)
Now we derive the optimal value for the parameter ψ_2 we have used in the left tail modified gamma regression (see Example 6.41). Therefore we apply the `glmProfile` function on the data frame `carinsur.ex` with the following specifications:

```
glmProfile(formula = (Cost/Claims)~Merit+Class,
           psi2 = seq(-1.55, -0.3, by=0.01), eta0=3.6,
           family = "Gamma",
           tail = "left",
           weights = Claims,
           start = c(3.2,0,0,0,-0.3,-0.1,-0.5,0.25),
           data = carinsur.ex)
```

Deviance Profile Plot for the Link Parameter
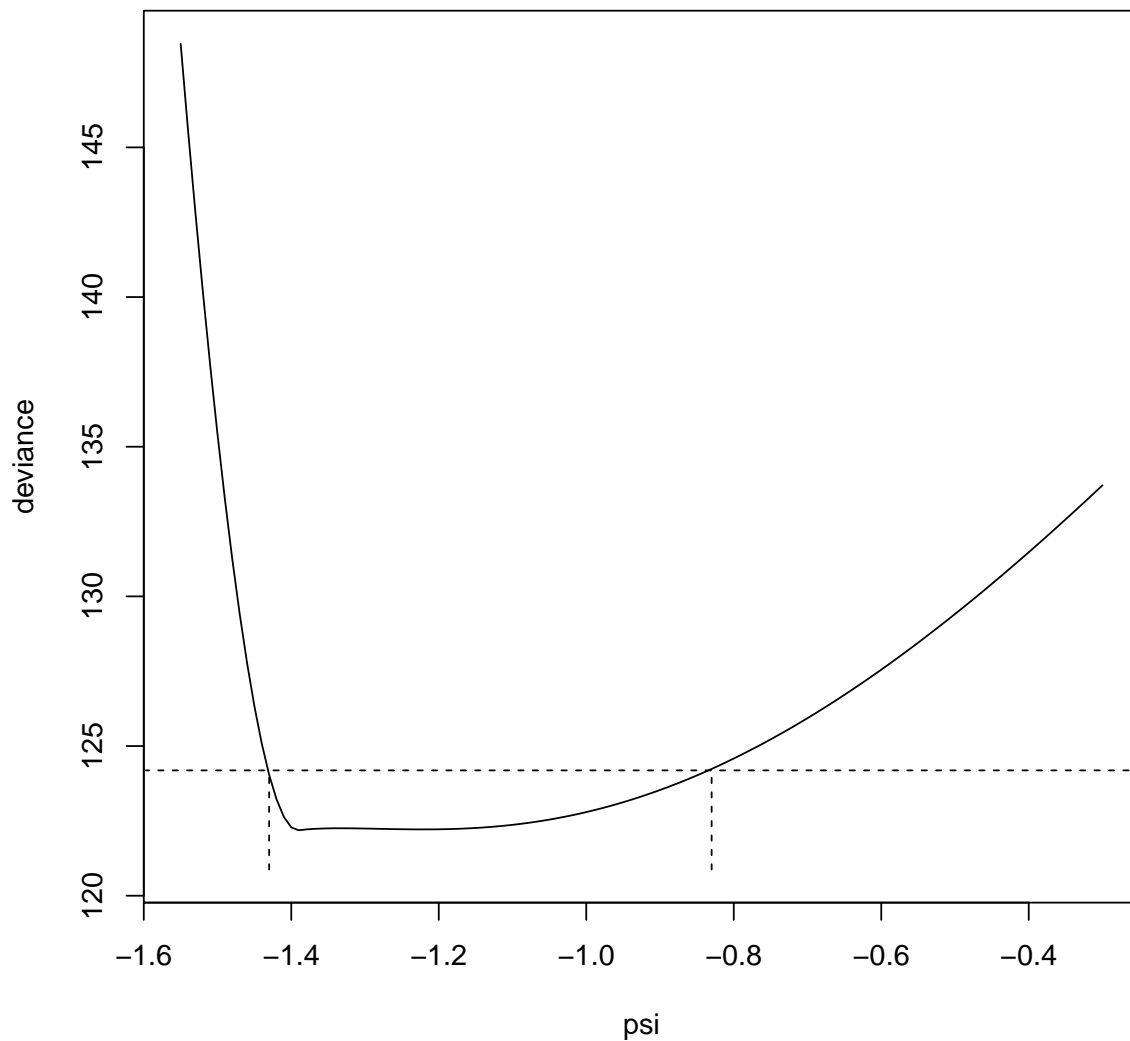
psi
Gamma GLM with left tail modified with 95% confidence interval

```
##
## minimum deviance= 122.186293700013 attained for -1.39 .
## $psi
##    [1] -1.55 -1.54 -1.53 -1.52 -1.51 -1.50 -1.49 -1.48 -1.47 -1.46
##   [11] -1.45 -1.44 -1.43 -1.42 -1.41 -1.40 -1.39 -1.38 -1.37 -1.36
##   [21] -1.35 -1.34 -1.33 -1.32 -1.31 -1.30 -1.29 -1.28 -1.27 -1.26
##   [31] -1.25 -1.24 -1.23 -1.22 -1.21 -1.20 -1.19 -1.18 -1.17 -1.16
##   [41] -1.15 -1.14 -1.13 -1.12 -1.11 -1.10 -1.09 -1.08 -1.07 -1.06
##   [51] -1.05 -1.04 -1.03 -1.02 -1.01 -1.00 -0.99 -0.98 -0.97 -0.96
##   [61] -0.95 -0.94 -0.93 -0.92 -0.91 -0.90 -0.89 -0.88 -0.87 -0.86
##   [71] -0.85 -0.84 -0.83 -0.82 -0.81 -0.80 -0.79 -0.78 -0.77 -0.76
##   [81] -0.75 -0.74 -0.73 -0.72 -0.71 -0.70 -0.69 -0.68 -0.67 -0.66
##   [91] -0.65 -0.64 -0.63 -0.62 -0.61 -0.60 -0.59 -0.58 -0.57 -0.56
## [101] -0.55 -0.54 -0.53 -0.52 -0.51 -0.50 -0.49 -0.48 -0.47 -0.46
## [111] -0.45 -0.44 -0.43 -0.42 -0.41 -0.40 -0.39 -0.38 -0.37 -0.36
## [121] -0.35 -0.34 -0.33 -0.32 -0.31 -0.30
##
## $dev
##    [1] 148.5 145.6 142.8 140.2 137.7 135.4 133.2 131.2 129.4 127.8
##   [11] 126.3 125.1 124.0 123.2 122.6 122.3 122.2 122.2 122.2 122.2
##   [21] 122.2 122.3 122.3 122.3 122.2 122.2 122.2 122.2 122.2 122.2
##   [31] 122.2 122.2 122.2 122.2 122.2 122.2 122.2 122.2 122.2 122.2
##   [41] 122.3 122.3 122.3 122.3 122.3 122.4 122.4 122.4 122.5 122.5
##   [51] 122.5 122.6 122.6 122.7 122.7 122.8 122.9 122.9 123.0 123.1
##   [61] 123.1 123.2 123.3 123.4 123.4 123.5 123.6 123.7 123.8 123.9
##   [71] 124.0 124.1 124.2 124.3 124.5 124.6 124.7 124.8 125.0 125.1
##   [81] 125.2 125.4 125.5 125.6 125.8 125.9 126.1 126.2 126.4 126.5
##   [91] 126.7 126.9 127.0 127.2 127.4 127.6 127.7 127.9 128.1 128.3
## [101] 128.5 128.6 128.8 129.0 129.2 129.4 129.6 129.8 130.0 130.2
## [111] 130.4 130.6 130.8 131.0 131.3 131.5 131.7 131.9 132.1 132.4
## [121] 132.6 132.8 133.0 133.3 133.5 133.7
```

8 Comparison

Table 8.1 sums up the key quantities we have derived in our work in the previous chapters. To clarify the improvement of the tail modified GLMs (using user-defined links) we calculated the improvement (abbreviated by "impr.") of the deviance with the following formula:

$$improvement\ in\ \% = \left(\frac{deviance\ in\ improved\ model}{deviance\ in\ ordinary\ model} - 1\right) \cdot 100\%$$

Remark 8.1 (Origin of the values)
The single values can be taken from the sections where we derived them. The values were rounded to an accuracy of two digits.

Family	Gaussian	Binomial			Poisson	Gamma
Data sets						
name	pcb.ex	beetle.ex	bys.ex	rotifer.ex	mining.ex	carinsur.ex
as seen in	Ex. 4.1	Ex. 4.2	Ex. 4.3	Ex. 4.4	Ex. 4.5	Ex. 4.6
link	identity	logit	logit	probit	log	inverse
# obs.	28	8	18	40	44	20
# covar.	1	1	3	3	2	7
p = k+1	2	2	4	4	3	8
dof	26	6	14	36	41	12
Ordinary GLMs						
as seen in	Ex. 5.1	Ex. 5.2	Ex. 5.3	Ex. 5.4	Ex. 5.5	Ex. 5.6
deviance	8.36	11.23	40.77	471.25	42.09	167.43
Tail modified GLMs						
tail	right, ψ_1	left, ψ_2	left, ψ_2	both, $\boldsymbol{\psi}$	right, ψ_1	left, ψ_2
opt. value	0.18	0.16	-1.96	$(0, -0.5)$	-0.57	-1.39
see	Ex. 6.11	Ex. 6.18	Ex. 6.20	Ex. 6.27	Ex. 6.34	Ex. 6.41
and	Ex. 7.1	Ex. 7.2	Czado (2007)		Ex. 7.3	Ex. 7.4
deviance	6.33	3.04	9.26	253.58	30.76	122.19
Comparison of the ordinary GLM with the tail modified GLM						
impr. in %	24.28	72.93	77.29	46.19	26.92	27.02

Table 8.1: Summary: improvement due to tail modifications.

Remark 8.2 (Degrees of freedom)
The difference between the number of observations and the number of parameters we have to estimate in our model is called the number of degrees of freedom of the model (often abbreviated by "dof"). Hence, we have dof = n - p (where p = k+1).

With Table 8.1 we have the possibility to compare the ordinary GLMs with the improved ones. We can see that in all of the examples we were able to reduce the deviance drastically. I.e. all of the improved GLMs have optimal $\boldsymbol{\psi}$ - values being far from $(1,1)$. This indicates that the goodness of fit of the ordinary GLMs could be optimized by using user-defined link functions in all of our examples.

9 Conclusion and outlook

This thesis addressed the implementation of parametric link families in R to fit generalized linear models. Czado (2007) described generalized linear models with parametric links and their implementation in the statistical environment of S. A crucial point was the definition of the general $h(\cdot)$ - power transformations as given in Section 6.1. We were able to define parametric link functions being either one- or two-parametric extensions of the ordinary link functions in R (see Section 6.2). The `glm` function, which is already implemented in R, can be called using a parametric link instead of an ordinary link function. This way we were able to decrease the deviance drastically in all of our examples (see Chapter 8). In Chapter 7 we determined the corresponding parameters for the tail modifying links. The function being used to determine the parameters is called `glmProfile`. It was one of the functions which were presented in Czado (2007) to fit these improved GLMs using user-defined link functions. There were also other interesting extensions given for several purposes. Now we want to provide an outlook on the theory and functions we described in this thesis by describing these extensions briefly.

- `glm.mle`: finds the maximum likelihood estimate of the joint vector $(\boldsymbol{\beta}, \boldsymbol{\psi}) \in \mathbb{R}^{p+2}$. For the single tail modifications we would expect that the joint vector is $(\boldsymbol{\beta}, \psi_1)$ (right tail modification) or $(\boldsymbol{\beta}, \psi_2)$ (left tail modification) both being $\in \mathbb{R}^{p+1}$. Nevertheless, also for single tail modifications we have to specify the vector $\boldsymbol{\psi}$ (setting $\psi_1 = 1$ for left tail modifications and $\psi_2 = 1$ for right tail modifications) in the `glm.mle` function. Either way we have to find the MLE of the joint vector $(\boldsymbol{\beta}, \boldsymbol{\psi}) \in \mathbb{R}^{p+2}$. This can be seen as an extension to our theory since we have only estimated the regression parameter $\boldsymbol{\beta} \in \mathbb{R}^p$. We pretended that somebody gave us the "correct" value of $\boldsymbol{\psi}$ to run the `glm` function with a parametric link. Hence, we did not derive estimates for $\boldsymbol{\psi}$ but rather we took the values delivering the minimum deviance by running the `glmProfile` function.

- `glm.inf`: a distortion coming from our approach is that the standard errors of the regression estimates are thus calculated as if the value of $\boldsymbol{\psi}$ was fixed. Therefore it would be useful to have a function providing the standard errors for the regression parameter $(\boldsymbol{\beta}, \boldsymbol{\psi}) \in \mathbb{R}^{p+2}$ when estimating them jointly. This is exactly what the function `glm.inf` does. It calculates the standard errors for both the model with fixed $\boldsymbol{\psi}$ only estimating $\boldsymbol{\beta}$ and the model estimating $(\boldsymbol{\beta}, \boldsymbol{\psi})$. This function also compares both standard errors and illustrates the variance inflation due to additional estimation. Moreover, the `glm.inf` function gives the link parameter $\boldsymbol{\psi}$ with its standard error and an estimated matrix containing correlations between the intercept, the covariates and the link parameters. This is an extension to our approach since we can only determine the standard errors of the regression parameters (i.e. of $\widehat{\beta}_i$ for $i \in \{1, \ldots, n\}$) by using Remark 3.28. This way we have no information about the estimation of $\boldsymbol{\psi}$. If we consider the estimation of $\boldsymbol{\psi}$ an equivalent theory about maximizing the joint likelihood of $(\boldsymbol{\beta}, \boldsymbol{\psi})$ would arise (i.e. $\max_{\boldsymbol{\beta}, \boldsymbol{\psi}} l(\boldsymbol{\beta}, \boldsymbol{\psi} \mid \boldsymbol{y})$). For the optimization also joint versions of the score equations and the fisher information matrix would appear. And as in Remark 3.28 the asymptotic distribution of $\widehat{(\boldsymbol{\beta}, \boldsymbol{\psi})}$ would be of interest.

- `glm.fitted`: serves to compare the improved tail modified models with the ordinary generalized linear models. It plots the estimated linear predictors versus the observed and the fitted means for the ordinary GLM and the GLM using a parametric link function.

We could also think out of the box and enlarge the topics presented in this thesis following the upcoming motivation. We did not use methods telling us that a linear regression model fits best to the data, i.e. yields to the most convincing predicted values. Thus, for specific data one could also compare the results of a GLM with the results of a non-linear model. Analysis of non-linear regression models is also needed in applications and described in the literature (e.g. Bates and Watts (2007)). Even if a linear regression model is appropriate we have to decide whether it makes sense to transform the response or the covariates. This is a mutual process alternating in model selection and data transformation. Another point is the choice of the covariates. In many data sets we could take a wide range of the available covariates. Often this is unrewarding. Therefore, it is necessary to penalize the complexity of the model if the improvement of the fit is not sufficient. The selection criteria uses the theory of testing hypothesis to assess the adequacy of a bigger model against the adequacy of a more compact one (see Fahrmeir and Tutz (2001)(Chapter 4)).

All in all we can conclude that the theory of generalized linear models is very interesting and wide-ranging. There are a lot of remarkable results and extensions. It is therefore not surprising that GLMs are essential models in both the theory about statistical analysis and the practical applications. One can reliably use the generalized linear models in applications due to a variety of well-developed and numerically stable routines for many statistical environments (like R for instance). The progress of delivering new packages and functions is not stagnating. Hence, the theory and applications of generalized linear models and its extensions will be a suspenseful subject also in the near future.

A   Appendix

A.1   Members of the exponential family

A.1.1   Normal distribution

The normal distribution $\mathcal{N}(\mu, \sigma^2)$ has the following density:

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

We want to rewrite the density in the form of Equation (3.1):

$$
\begin{aligned}
f(x \mid \mu, \sigma^2) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2} + \ln\left(\frac{1}{\sigma\sqrt{2\pi}}\right)\right\} \\
&= \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2} - \ln\left(\sigma\sqrt{2\pi}\right)\right\} \\
&= \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2} - \frac{1}{2}\ln\left(\sigma^2 2\pi\right)\right\} \\
&= \exp\left\{-\frac{x^2 - 2x\mu + \mu^2}{2\sigma^2} - \frac{1}{2}\ln\left(\sigma^2 2\pi\right)\right\} \\
&= \exp\left\{\frac{-x^2 + 2x\mu - \mu^2}{2\sigma^2} - \frac{1}{2}\ln\left(\sigma^2 2\pi\right)\right\} \\
&= \exp\left\{\frac{-\frac{1}{2}x^2 + x\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{1}{2}\ln\left(\sigma^2 2\pi\right)\right\} \\
&= \exp\left\{\frac{x\mu - \frac{1}{2}\mu^2}{\sigma^2} - \frac{\frac{1}{2}x^2}{\sigma^2} - \frac{1}{2}\ln\left(\sigma^2 2\pi\right)\right\} \\
&= \exp\left\{\frac{x\,\overbrace{\mu}^{:=\theta(\mu)} - \overbrace{\frac{1}{2}\mu^2}^{:=b(\theta)}}{\underbrace{\sigma^2}_{:=a(\phi)}} + \underbrace{\left[-\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \ln\left(\sigma^2 2\pi\right)\right)\right]}_{:=c(x,\phi)}\right\}
\end{aligned}
$$

$\square$

Hence, we can take $\phi = \sigma^2$ ((σ may be unknown)) and $\omega = 1$ and therefore, $a(\phi) = \frac{\sigma^2}{1} = \sigma^2 = \phi$, $\theta(\mu) = \mu$ and $b(\theta) = \frac{1}{2}\mu^2 = \frac{\theta^2}{2}$. And

$$c(x, \phi) = -\frac{1}{2}\left(\frac{x^2}{\sigma^2} + \ln\left(\sigma^2 2\pi\right)\right)$$

$$= -\frac{1}{2}\frac{x^2}{\phi} - \frac{1}{2}\ln\left(\phi 2\pi\right)$$

$$= -\ln\left(\sqrt{\phi 2\pi}\right) - \frac{x^2}{2\phi}$$

Remark A.1 (Origin of the name *generalized* linear model)
Only by ensuring that the normal distribution is a distribution of the exponential family are we able to derive the *generalized linear models* from the *linear regression model*.

Remark A.2
We are aware of the following properties for $X \sim \mathcal{N}(\mu, \sigma^2)$:

- parameters: $\mu \in \mathbb{R}$ and $\sigma^2 > 0$

- $\mathbb{E}\left[X\right] = \mu$

- $\mathbb{V}ar\left[X\right] = \sigma^2$

Remark A.3 (The standard normal distribution)
It is common to denote the density of the standard normal distribution by φ and the distribution function by Φ. Hence, we use the following notation:

$$f(x \mid 0, 1) := \varphi(x)$$
$$F(x) := \Phi(x)$$

A.1.2   Scaled binomial distribution

By the expression scaled binomial distribution we refer to the following situation. For $Y \sim Bin(n, p)$ we say $Y^* := \frac{Y}{n} \sim ScaledBin(n, p)$ follows a *scaled binomial distribution*. The ordinary binomial distribution has the following probability mass function:

$$\mathbb{P}\left(Y = k\right) = \binom{n}{k}p^k(1-p)^{n-k} \qquad\qquad k \in \{0, 1, \ldots, n\}$$

Hence, the probability mass function of the $Y^* \sim ScaledBin(n, p)$ is given by (since $k^* := \frac{k}{n}$ has another domain, i.e. $k^* \in \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$):

$$\mathbb{P}\left(Y^* = \frac{Y}{n} = k^*\right) = \mathbb{P}\left(Y = nk^*\right) = \binom{n}{nk^*}p^{nk^*}(1-p)^{n-(nk^*)}$$

Now we want to rewrite the probability mass function in the form of Equation (3.1):

$$\mathbb{P}\left(\frac{Y}{n} = k^*\right) = \binom{n}{nk^*}p^{nk^*}(1-p)^{n-(nk^*)}$$

$$= \exp\left\{\ln\left(\binom{n}{nk^*}p^{nk^*}(1-p)^{n-(nk^*)}\right)\right\}$$

$$= \exp\left\{\ln\left(\binom{n}{nk^*}\right) + (nk^*)\ln\left(p\right) + (n - (nk^*))\ln\left(1-p\right)\right\}$$

$$= \exp \left\{ \ln \left(\binom{n}{nk^*} \right) + (nk^*)\left(\ln(p) - \ln(1-p) \right) + n \ln(1-p) \right\}$$

$$= \exp \left\{ \ln \left(\binom{n}{nk^*} \right) + (nk^*) \ln \left(\frac{p}{1-p} \right) + n \ln(1-p) \right\}$$

$$= \exp \left\{ n \cdot \left(k^* \ln \left(\frac{p}{1-p} \right) + \ln(1-p) \right) + \ln \left(\binom{n}{nk^*} \right) \right\}$$

$$= \exp \left\{ \frac{1}{\frac{1}{n}} \cdot \left(k^* \ln \left(\frac{p}{1-p} \right) + \ln(1-p) \right) + \ln \left(\binom{n}{nk^*} \right) \right\}$$

$$= \exp \left\{ \frac{k^* \overbrace{\ln \left(\frac{p}{1-p} \right)}^{:=\theta(\mu)=\theta(p)} + \overbrace{\ln(1-p)}^{\rightsquigarrow b(\theta) \overset{!}{=} -\ln(1-p)}}{\underbrace{\frac{1}{n}}_{:=a(\phi)}} + \underbrace{\ln \left(\binom{n}{nk^*} \right)}_{:=c(k*,\phi))} \right\}$$

$\square$

Hence, we can take $\phi = 1$ and $\omega = n$ and therefore, $a(\phi) = \frac{1}{n}$, $\theta = \theta(p) = \ln \left(\frac{p}{1-p} \right)$ and

$$b(\theta) \overset{!}{=} -\ln(1-p) = \ln \left(\frac{1}{1-p} \right)$$

$$\text{i.e. } \frac{1}{1-p} = \frac{p}{1-p} + 1 = \exp\{\theta\} + 1$$

$$\Rightarrow b(\theta) = \ln(1 + \exp\{\theta\})$$

Remark A.4
We are aware of the following properties for $Y \sim Bin(n,p)$:

- parameters: $n \in \mathbb{N}$ and $p \in (0,1)$

- $\mathbb{E}[Y] = np$

- $\mathbb{V}ar[Y] = np(1-p)$

Remark A.5
By Remark A.4 we can derive properties for $Y^* = \frac{Y}{n} \sim ScaledBin(n,p)$:

- parameters: $n \in \mathbb{N}$ and $p \in (0,1)$

- $\mathbb{E}[Y^*] = p$

- $\mathbb{V}ar[Y^*] = \frac{p(1-p)}{n}$

Remark A.6 (Degenerated cases)
$p = 0$ and $p = 1$ are degenerated cases which will not be regarded in this thesis.

A.1.3   Poisson distribution

The Poisson distribution has the following probability mass function (for $k \in \mathbb{N}_0$):

$$\mathbb{P}\left(X = n\right) = \exp\{-\lambda\}\frac{\lambda^n}{n!}$$

Now we want to rewrite the probability mass function in the form of Equation (3.1):

$$\mathbb{P}\left(X = n\right) = \exp\{-\lambda\}\frac{\lambda^n}{n!} = \exp\left\{\ln\left(\exp\{-\lambda\}\frac{\lambda^n}{n!}\right)\right\}$$

$$= \exp\left\{\ln\left(\exp\{-\lambda\}\right) + \ln\left(\lambda^n\right) - \ln\left(n!\right)\right\}$$

$$= \exp\left\{-\lambda + n\ln\left(\lambda\right) - \ln\left(n!\right)\right\}$$

$$= \exp\left\{\frac{n\overbrace{\ln\left(\lambda\right)}^{\theta} - \overbrace{\lambda}^{b(\theta)}}{\underbrace{\frac{1}{1}}_{:=a(\phi)}} + \underbrace{\left(-\ln\left(n!\right)\right)}_{:=c(n,\phi))}\right\}$$

$\square$

Hence, we can take $\phi = 1$ and $\omega = 1$ and therefore, $a(\phi) = \frac{\phi}{\omega} = \frac{1}{1} = 1$, $\theta = \theta(\lambda) = \ln\left(\lambda\right)$ and $b(\theta) = \exp\{\theta\} = \lambda$

Remark A.7
We are aware of the following properties for $X \sim Pois(\lambda)$:

- parameter: $\lambda > 0$

- $\mathbb{E}\left[X\right] = \lambda$

- $\mathbb{V}ar\left[X\right] = \lambda$

A.1.4   Gamma distribution

The gamma distribution $\Gamma(\mu, \nu)$ has the following density:

$$f(x \mid \mu, \nu) = \frac{\nu^\nu}{\mu^\nu\Gamma(\nu)}x^{\nu-1}\exp\left\{-\frac{\nu}{\mu}x\right\}\mathbb{1}_{(0,\infty)}(x)$$

$$= \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^\nu x^{\nu-1}\exp\left\{-\frac{\nu}{\mu}x\right\}\mathbb{1}_{(0,\infty)}(x)$$

We want to rewrite the density in the form of Equation (3.1):

$$f(x \mid \mu, \nu) = \frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^\nu x^{\nu-1}\exp\left\{-\frac{\nu}{\mu}x\right\}$$

$$= \exp\left\{\ln\left(\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^\nu x^{\nu-1}\right) - \frac{\nu}{\mu}x\right\}$$

$$= \exp \left\{ -\ln\left(\Gamma(\nu)\right) + \nu\left(\ln\left(\nu\right) - \ln\left(\mu\right)\right) + \left(\nu - 1\right)\ln\left(x\right) - \frac{\nu}{\mu}x \right\}$$

$$= \exp \left\{ -\ln\left(\Gamma(\nu)\right) + \nu\left(\ln\left(\nu\right) - \ln\left(\mu\right)\right) + \nu\ln\left(x\right) - \ln\left(x\right) - \frac{\nu}{\mu}x \right\}$$

$$= \exp \left\{ \nu\left(\ln\left(\nu\right) - \ln\left(\mu\right) + \ln\left(x\right) - \frac{1}{\mu}x\right) - \ln\left(\Gamma(\nu)\right) - \ln\left(x\right) \right\}$$

$$= \exp \left\{ \nu\left(-\ln\left(\mu\right) - \frac{1}{\mu}x\right) + \nu\left(\ln\left(\nu\right) + \ln\left(x\right)\right) - \ln\left(\Gamma(\nu)\right) - \ln\left(x\right) \right\}$$

$$= \exp \left\{ \frac{x\overbrace{\left(-\frac{1}{\mu}\right)}^{:=\theta} - \overbrace{\ln\left(\mu\right)}^{:=b(\theta)}}{\underbrace{\frac{1}{\nu}}_{:=a(\phi)}} + \underbrace{\left[\nu\left(\ln\left(\nu\right) + \ln\left(x\right)\right) - \ln\left(\Gamma(\nu)\right) - \ln\left(x\right)\right]}_{:=c(x,\phi)} \right\}$$

$\square$

Hence, we can take $\phi = \frac{1}{\nu}$ (ν may be unknown) and $\omega = 1$ and therefore, $a(\phi) = \frac{\phi}{\omega} = \frac{1}{\nu} = \phi$, $\theta = \theta(\mu) = -\frac{1}{\mu}$.

$$b(\theta) = b\left(-\frac{1}{\mu}\right) \overset{!}{=} \ln(\mu)$$

$$\text{i.e.} \quad \theta = -\frac{1}{\mu} \Leftrightarrow -\theta = \frac{1}{\mu}$$

$$\Leftrightarrow \mu = -\frac{1}{\theta}$$

$$\Rightarrow b(\theta) = \ln(\mu) = \ln\left(-\frac{1}{\theta}\right) = \ln\left(\frac{1}{-\theta}\right)$$

$$\Rightarrow b(\theta) = -\ln\left(-\theta\right)$$

and

$$c(x,\phi) = \nu\left(\ln\left(\nu\right) + \ln\left(x\right)\right) - \ln\left(\Gamma(\nu)\right) - \ln\left(x\right)$$

$$= \frac{1}{\phi}\left(\ln\left(\frac{1}{\phi}\right) + \ln\left(x\right)\right) - \ln\left(\Gamma\left(\frac{1}{\phi}\right)\right) - \ln\left(x\right)$$

$$= -\frac{1}{\phi}\ln(\phi) + \frac{1}{\phi}\ln\left(x\right) - \ln\left(\Gamma\left(\frac{1}{\phi}\right)\right) - \ln\left(x\right)$$

$$= \left(\frac{1}{\phi} - 1\right)\ln\left(x\right) - \frac{\ln(\phi)}{\phi} - \ln\left(\Gamma\left(\frac{1}{\phi}\right)\right)$$

$$= \left(\frac{1-\phi}{\phi}\right)\ln\left(x\right) - \frac{\ln(\phi)}{\phi} - \ln\left(\Gamma\left(\frac{1}{\phi}\right)\right)$$

Remark A.8

We are aware of the following properties for $X \sim \Gamma(\mu, \nu)$:

- parameters: $\mu > 0$ and $\nu > 0$

- gamma - function: $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt$

- $\mathbb{E}[X] = \mu$

- $\mathbb{V}ar[X] = \frac{\mu^2}{\nu}$

Remark A.9 (Parametrization of the gamma distribution)

Many different parametrizations of the gamma distribution can be found. We will use the parametrization as given in Fahrmeir and Tutz (2001)(p. 23).

A.2   More about the general $h(\cdot)$ - power transformations

A.2.1   Right tail modification

Recall that we defined the right tail modification in Definition 6.2. In the following we want to derive certain functions corresponding to $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$. We also present their implementation in R.

$$
h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1) = \begin{cases} \eta_0 + \ln(\eta - \eta_0 + 1) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \eta & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}
$$

```r
hpsi1 <- function(psi1 = stop("Argument 'psi1' is missing"),
                  eta = stop("Argument 'eta' is missing"), eta0 = 0)
{h <- 1:length(eta)
 h[eta < eta0] <- eta[eta < eta0]
 if (any(psi1 > -1e-14 && psi1 < 1e-14)) {
    h[eta >= eta0] <- eta0 + log(eta[eta >= eta0] - eta0 + 1)
 }
 else {
    h[eta >= eta0] <- ((eta[eta >= eta0] - eta0 + 1)^psi1 - 1)/psi1
    h[eta >= eta0] <- h[eta >= eta0] + eta0
 }
 h
}
```

Domain of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$

For given $\boldsymbol{\psi} = \psi_1 \in \mathbb{R}$ the domain of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ can be expressed as a function of η:

$$
domain(h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)) = \mathbb{1}_{\{\eta \in \Psi\}} = \begin{cases} \mathbb{1}_{\{\eta > \eta_0 - 1\}} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ many\ cases & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \mathbb{1}_{\{\eta \in \mathbb{R}\}} & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}
$$

$$
\overset{\underset{take\ easier\ implementation}{=}}{{}_{\texttt{is.finite}\ is\ an\ R\ intern\ function}} \mathbb{1}_{\left\{ \texttt{is.finite}(h_{\eta_0}(\eta, \psi = \psi_1)) \right\}} \tag{A.1}
$$

Since:

for the **first** case: *$\eta \geq \eta_0$ and $\psi_1 = 0$:*

$$\eta - \eta_0 + 1 \overset{!}{>} 0 \qquad\qquad domain\ of\ \ln(\cdot)$$
$$\Leftrightarrow \eta > \eta_0 - 1$$

for the **second** case: *$\eta \geq \eta_0$ and $\psi_1 \neq 0$:*
This case requires a lot of analysis of different cases. The domain of the function for the second case is defined by the domain of $\frac{(\eta-\eta_0+1)^{\psi_1}-1}{\psi_1}$ for a specific value of ψ_1. For this value we differentiate the following cases:

- $\psi_1 = 1$: the domain is the real line

- $\psi_1 = -1$: the domain are the values for which $\eta - \eta_0 + 1 \neq 0$, i.e. $\mathbb{R} \setminus \{\eta_0 - 1\}$

- $\psi_1 \neq 1 \wedge \psi_1 \neq -1$

  - $\psi_1 > 0$
    - ψ_1 is not integer: the domain are the values for which $\eta - \eta_0 + 1$ is positive (Example: $\psi_1 = 0.5$, η_0 is arbitrary $\Rightarrow$ we have $\frac{\sqrt{\eta-\eta_0+1}-1}{0.5}$)
    - ψ_1 is an integer:
      · ψ_1 odd: the domain is the real line (Example: $\psi_1 = 3$, η_0 is arbitrary $\Rightarrow$ we have $\frac{(\eta-\eta_0+1)^3-1}{3}$)
      · ψ_1 even: the domain is the real line (Example: $\psi_1 = 2$, η_0 is arbitrary $\Rightarrow$ we have $\frac{(\eta-\eta_0+1)^2-1}{2}$)
  - $\psi_1 < 0$
    - ψ_1 is not integer: the domain are the values for which $\eta - \eta_0 + 1$ is positive (Example: $\psi_1 = -0.5$, η_0 is arbitrary $\Rightarrow$ we have $-\frac{\frac{1}{\sqrt{\eta-\eta_0+1}}-1}{0.5}$)
    - ψ_1 is an integer:
      · ψ_1 odd: the domain are the values for which $\eta - \eta_0 + 1 \neq 0$, i.e. $\mathbb{R}\setminus\{\eta_0-1\}$ (Example: $\psi_1 = -3$, η_0 is arbitrary $\Rightarrow$ we have $-\frac{\frac{1}{(\eta-\eta_0+1)^3}-1}{3}$)
      · ψ_1 even: the domain are the values for which $\eta - \eta_0 + 1 \neq 0$, i.e. $\mathbb{R}\setminus\{\eta_0-1\}$ (Example: $\psi_1 = -2$, η_0 is arbitrary $\Rightarrow$ we have $-\frac{\frac{1}{(\eta-\eta_0+1)^2}-1}{2}$)

for the **third** case: *$\eta < \eta_0$:*
$$\eta \in \mathbb{R}$$

Inverse of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ w.r.t. η

We get:

$$\eta = \left(h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)\right)^{-1}(y)$$
$$= \begin{cases} \exp\{y - \eta_0\} + \eta_0 - 1 & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ (\psi_1(y - \eta_0) + 1)^{\frac{1}{\psi_1}} + \eta_0 - 1 & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ y & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases} \quad (\text{A.2})$$

Since:

for the **first** case: $\eta \geq \eta_0$ *and* $\psi_1 = 0$:

$$y = \eta_0 + \ln(\eta - \eta_0 + 1)$$
$$\Leftrightarrow y - \eta_0 = \ln(\eta - \eta_0 + 1)$$
$$\Leftrightarrow \exp\{y - \eta_0\} = \eta - \eta_0 + 1$$
$$\Leftrightarrow \eta = \exp\{y - \eta_0\} + \eta_0 - 1$$

for the **second** case: $\eta \geq \eta_0$ *and* $\psi_1 \neq 0$:

$$y = \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}$$
$$\Leftrightarrow y - \eta_0 = \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}$$
$$\Leftrightarrow (y - \eta_0)\psi_1 = (\eta - \eta_0 + 1)^{\psi_1} - 1$$
$$\Leftrightarrow (y - \eta_0)\psi_1 + 1 = (\eta - \eta_0 + 1)^{\psi_1}$$
$$\Leftrightarrow ((y - \eta_0)\psi_1 + 1)^{\frac{1}{\psi_1}} = \eta - \eta_0 + 1$$
$$\Leftrightarrow \eta = ((y - \eta_0)\psi_1 + 1)^{\frac{1}{\psi_1}} + \eta_0 - 1$$

for the **third** case: $\eta < \eta_0$:

$$y = \eta$$

```r
hpsi1INV<-function(psi1, y, eta0 = 0)
{h <- 1:length(y)
 if (any(psi1 > -1e-14 && psi1 < 1e-14)){
 h[y >= eta0] <- eta0 - 1 + exp(y[y >= eta0] - eta0)
 }
 else {
 h[y >= eta0] <- eta0 - 1
         + ((1 + (psi1 * (-eta0 + y[y >= eta0])))^(1/psi1))
 }
 h[y < eta0] <- y[y < eta0]
 h
}
```

Derivative of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ w.r.t. η

Now we want to calculate:

$$\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1) = \begin{cases} \frac{\partial}{\partial \eta}\left(\eta_0 + \ln(\eta - \eta_0 + 1)\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\partial}{\partial \eta}\left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \frac{\partial}{\partial \eta}(\eta) & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}$$

$$= \begin{cases} \frac{1}{\eta - \eta_0 + 1} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ (\eta - \eta_0 + 1)^{(\psi_1 - 1)} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ 1 & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases} \tag{A.3}$$

```r
hpsi1DERIV<-function(psi1, eta, eta0 = 0)
{h <- 1:length(eta)
 if (any(psi1 > -1e-14 && psi1 < 1e-14)){
  h[eta >= eta0] <- 1/(eta[eta >= eta0] - eta0 + 1)
 }
 else {
  h[eta >= eta0] <- (1 - eta0 + eta[eta >= eta0])^(psi1 - 1)
 }
 h[eta < eta0] <- 1
 h
}
```

Derivative of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ w.r.t. ψ_1

Now we want to calculate:

$$\frac{\partial}{\partial \psi_1} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$$

$$= \begin{cases} \frac{\partial}{\partial \psi_1}\left(\eta_0 + \ln(\eta - \eta_0 + 1)\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\partial}{\partial \psi_1}\left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \frac{\partial}{\partial \psi_1}(\eta) & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}$$

$$= \begin{cases} \lim_{\psi_1 \to 0} \frac{\partial}{\partial \psi_1}\left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\ln(\eta - \eta_0 + 1)(\eta - \eta_0 + 1)^{\psi_1}\psi_1 - \left((\eta - \eta_0 + 1)^{\psi_1} - 1\right)}{\psi_1^2} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ 0 & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}$$

$$= \begin{cases} \lim_{\psi_1 \to 0} \frac{\partial}{\partial \psi_1}\left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\ln(\eta - \eta_0 + 1)(\eta - \eta_0 + 1)^{\psi_1}\psi_1 - (\eta - \eta_0 + 1)^{\psi_1} + 1}{\psi_1^2} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ 0 & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}$$

Remark A.10

The limit $\lim\limits_{\psi_1 \to 0} \frac{\partial}{\partial \psi_1}\left(\eta_0 + \frac{(\eta-\eta_0+1)^{\psi_1}-1}{\psi_1}\right)$ can by calculated by:

$$\lim_{\psi_1 \to 0} \frac{\partial}{\partial \psi_1}\left(\eta_0 + \frac{(\eta-\eta_0+1)^{\psi_1}-1}{\psi_1}\right)$$

$$= \lim_{\psi_1 \to 0}\left[\frac{\ln(\eta-\eta_0+1)(\eta-\eta_0+1)^{\psi_1}\psi_1 - (\eta-\eta_0+1)^{\psi_1}+1}{\psi_1^2}\right]$$

$$\overset{L'Hospital}{=} \lim_{\psi_1 \to 0}\left[\frac{\ln(\eta-\eta_0+1)\left(\ln(\eta-\eta_0+1)(\eta-\eta_0+1)^{\psi_1}\psi_1 + (\eta-\eta_0+1)^{\psi_1}\right)}{2\psi_1}\right.$$
$$\left. - \frac{\left(\ln(\eta-\eta_0+1)(\eta-\eta_0+1)^{\psi_1}\right)}{2\psi_1}\right]$$

$$= \lim_{\psi_1 \to 0}\left[\frac{\ln(\eta-\eta_0+1)^2(\eta-\eta_0+1)^{\psi_1}\psi_1}{2\psi_1}\right]$$

$$\overset{L'Hospital}{=} \lim_{\psi_1 \to 0}\left[\frac{\ln(\eta-\eta_0+1)^2\left(\ln(\eta-\eta_0+1)(\eta-\eta_0+1)^{\psi_1}\psi_1 + (\eta-\eta_0+1)^{\psi_1}\right)}{2}\right]$$

$$= \left(\frac{\ln(\eta-\eta_0+1)^2}{2}\right)\lim_{\psi_1 \to 0}\left[\ln(\eta-\eta_0+1)\underbrace{(\eta-\eta_0+1)^{\psi_1}}_{\to 1}\underbrace{\psi_1}_{\to 0} + \underbrace{(\eta-\eta_0+1)^{\psi_1}}_{\to 1}\right]$$

$$= \frac{\ln(\eta-\eta_0+1)^2}{2}$$

All in all we get:

$$\frac{\partial}{\partial \psi_1}h_{\eta_0}(\eta, \boldsymbol{\psi}=\psi_1) = \begin{cases} \frac{\ln(\eta-\eta_0+1)^2}{2} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\ln(\eta-\eta_0+1)(\eta-\eta_0+1)^{\psi_1}\psi_1-(\eta-\eta_0+1)^{\psi_1}+1}{\psi_1^2} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ 0 & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}$$

```
hpsi1DERIV1<-function(psi1, eta, eta0 = 0)
{h <- eta
 temp <- eta
 if (any(psi1 > -1e-14 && psi1 < 1e-14)){
 h[eta >= eta0] <- ((log(eta[eta >= eta0] - eta0 + 1))^2)/2
 }
 else {
  temp[eta >= eta0] <- (1 - eta0 + eta[eta >= eta0])^psi1
  h[eta >= eta0] <- ((temp[eta >= eta0] *
            log(eta[eta >= eta0] - eta0 + 1) * psi1)
          - (temp[eta >= eta0] - 1)) /(psi1^2)
 }
 h[eta < eta0] <- 0
 h
}
```

Limit $\lim_{\eta \to \infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ (for ψ_1 fixed)

Now we want to consider the limit $\lim_{\eta \to \infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ for fixed values of ψ_1. In the right tail modification the limit $\lim_{\eta \to -\infty}$ is of no interest since $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ is the identity for $\eta < \eta_0$.

$$\lim_{\eta \to \infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1) \;=\; \lim_{\eta \to \infty} \begin{cases} \eta_0 + \ln(\eta - \eta_0 + 1) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \eta & \text{otherwise (i.e. if } \eta < \eta_0) \end{cases}$$

$$= \begin{cases} \lim\limits_{\eta \to \infty} \left(\eta_0 + \ln(\eta - \eta_0 + 1)\right) & \text{if } \psi_1 = 0 \\ \lim\limits_{\eta \to \infty} \left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}\right) & \text{if } \psi_1 \neq 0 \end{cases} = \begin{cases} \eta_0 + \lim\limits_{\eta \to \infty} \ln(\eta - \eta_0 + 1) & \text{if } \psi_1 = 0 \\ \eta_0 - \frac{1}{\psi_1} + \frac{1}{\psi_1} \lim\limits_{\eta \to \infty} (\eta - \eta_0 + 1)^{\psi_1} & \text{if } \psi_1 \neq 0 \end{cases}$$

$$= \begin{cases} \infty & \text{if } \psi_1 = 0 \\ \begin{cases} \infty & \text{if } \psi_1 > 0 \\ \eta_0 - \frac{1}{\psi_1} & \text{if } \psi_1 < 0 \end{cases} & \text{if } \psi_1 \neq 0 \end{cases}$$

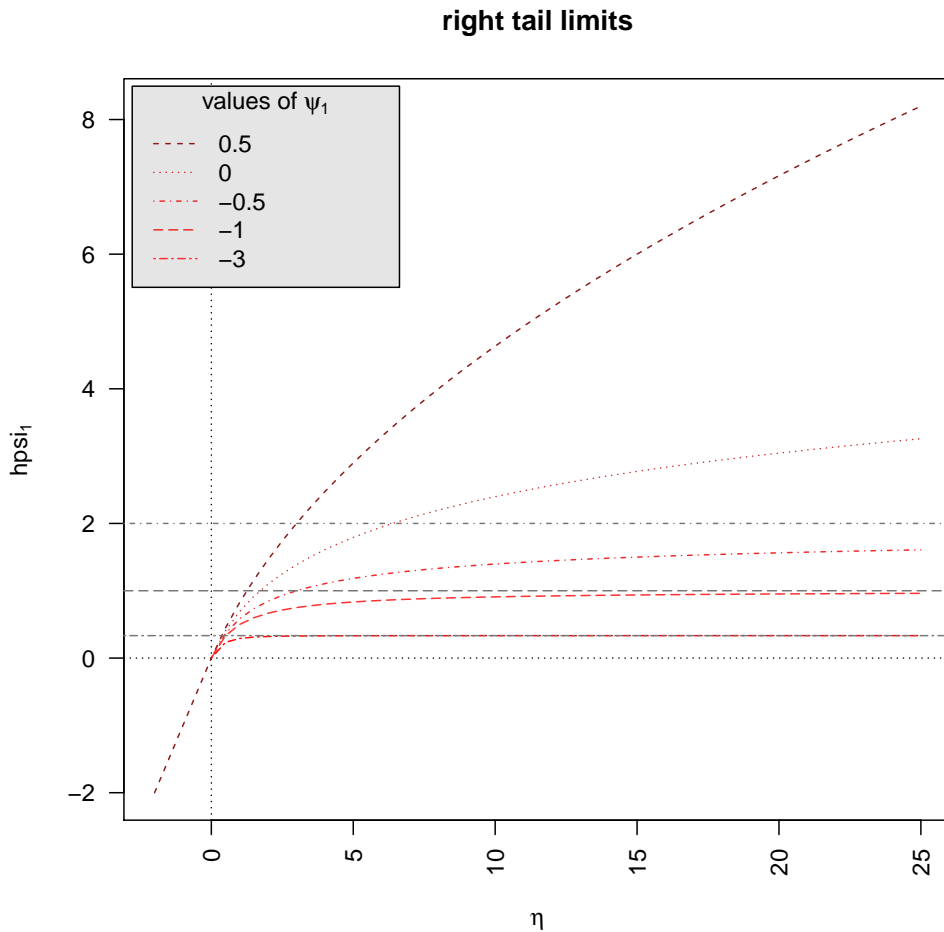Figure A.1: Limit: $\lim_{\eta \to \infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ for different values of ψ_1 and $\eta_0 = 0$. The grey dotted horizontal lines indicate the behavior as $\eta \to \infty$ in case of a finite limit.

A.2.2   Left tail modification

Recall that we defined the left tail modification in Definition 6.3. In the following we want to derive certain functions corresponding to $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$. We also present their implementation in R.

$$
h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2) = \begin{cases} \eta & \text{if } \eta \geq \eta_0 \\ \eta_0 - \ln(-\eta + \eta_0 + 1) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}
$$

```r
hpsi2 <- function (psi2 = stop("Argument 'psi2' is missing"),
                   eta = stop("Argument 'eta' is missing"), eta0 = 0)
{h <- 1:length(eta)
 h[eta >= eta0] <- eta[eta >= eta0]
 if (any(psi2 > -1e-14 && psi2 < 1e-14)) {
   h[eta < eta0] <- eta0 - log(-(eta[eta < eta0]) + eta0 + 1)
 }
 else {
   h[eta < eta0] <- -((- eta[eta < eta0] + eta0 + 1)^psi2 - 1)/psi2
   h[eta < eta0] <- h[eta < eta0] + eta0
 }
 h
}
```

Domain of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$

For given $\boldsymbol{\psi} = \psi_2 \in \mathbb{R}$ the domain of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ can be expressed as a function of η:

$$
domain(h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)) = \mathbb{1}_{\{\eta \in \Psi\}} = \begin{cases} \mathbb{1}_{\{\eta \in \mathbb{R}\}} & \text{if } \eta \geq \eta_0 \\ \mathbb{1}_{\{\eta < \eta_0 + 1\}} & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ many \ cases & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}
$$

$$
\underset{\substack{\texttt{is.finite} \ is \ an \ R \ intern \ function}}{\overset{\substack{take \ easier \ implementation}}{=}} \mathbb{1}_{\left\{ \texttt{is.finite}(h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)) \right\}} \tag{A.4}
$$

Since:

for the **first** case: $\eta \geq \eta_0$

$$
\eta \in \mathbb{R}
$$

for the **second** case: $\eta < \eta_0 \text{ and } \psi_2 = 0$

$$
-\eta + \eta_0 + 1 \overset{!}{>} 0 \qquad\qquad domain \ of \ \ln(\cdot)
$$
$$
\Leftrightarrow \eta < \eta_0 + 1
$$

for the **third** case: $\eta < \eta_0$ and $\psi_2 \neq 0$

This case requires a lot of analysis of different cases. The domain of the function for the third cases is defined by the domain of $-\frac{(-\eta+\eta_0+1)^{\psi_2}-1}{\psi_2}$ for a specific value of ψ_2. For this value we differentiate the following cases:

- $\psi_2 = 1$: the domain is the real line

- $\psi_2 = -1$: the domain are the values for which $-\eta + \eta_0 + 1 \neq 0$, i.e. $\mathbb{R} \setminus \{\eta_0 + 1\}$

- $\psi_2 \neq 1 \wedge \psi_2 \neq -1$

    - $\psi_2 > 0$
        * ψ_2 is not integer: the domain are the values for which $-\eta+\eta_0+1$ is positive (Example: $\psi_2 = 0.5$, η_0 is arbitrary $\Rightarrow$ we have $-\frac{\sqrt{-\eta+\eta_0+1}-1}{0.5}$)
        * ψ_2 is an integer:
            · ψ_2 odd: the domain is the real line (Example: $\psi_2 = 3$, η_0 is arbitrary $\Rightarrow$ we have $-\frac{(-\eta+\eta_0+1)^3-1}{3}$)
            · ψ_2 even: the domain is the real line (Example: $\psi_2 = 2$, η_0 is arbitrary $\Rightarrow$ we have $-\frac{(-\eta+\eta_0+1)^2-1}{2}$)
    - $\psi_2 < 0$
        * ψ_2 is not integer: the domain are the values for which $-\eta+\eta_0+1$ is positive (Example: $\psi_2 = -0.5$, η_0 is arbitrary $\Rightarrow$ we have $\frac{\frac{1}{\sqrt{-\eta+\eta_0+1}}-1}{0.5}$)
        * ψ_2 is an integer:
            · ψ_2 odd: the domain are the values for which $-\eta + \eta_0 + 1 \neq 0$, i.e. $\mathbb{R}\setminus\{\eta_0+1\}$ (Example: $\psi_2 = -3$, η_0 is arbitrary $\Rightarrow$ we have $\frac{\frac{1}{(-\eta+\eta_0+1)^3}-1}{3}$)
            · ψ_2 even: the domain are the values for which $-\eta + \eta_0 + 1 \neq 0$, i.e. $\mathbb{R}\setminus\{\eta_0+1\}$ (Example: $\psi_2 = -2$, η_0 is arbitrary $\Rightarrow$ we have $\frac{\frac{1}{(\eta-\eta_0+1)^2}-1}{2}$)

Inverse of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ w.r.t. η

We get:

$$\eta = \left(h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)\right)^{-1}(y)$$
$$= \begin{cases} y & \text{if } \eta \geq \eta_0 \\ \eta_0 + 1 - \exp\{\eta_0 - y\} & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \eta_0 + 1 - (\psi_2(\eta_0 - y) + 1)^{\frac{1}{\psi_2}} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases} \tag{A.5}$$

Since:

for the **first** case: $\eta \geq \eta_0$

$$y = \eta$$

for the **second** case: $\eta < \eta_0$ *and* $\psi_2 = 0$

$$y = \eta_0 - \ln(-\eta + \eta_0 + 1)$$
$$\Leftrightarrow \eta_0 - y = \ln(-\eta + \eta_0 + 1)$$
$$\Leftrightarrow \exp\{\eta_0 - y\} = -\eta + \eta_0 + 1$$
$$\Leftrightarrow \eta = \eta_0 + 1 - \exp\{\eta_0 - y\}$$

for the **third** case: $\eta < \eta_0$ and $\psi_2 \neq 0$

$$y = \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2}$$
$$\Leftrightarrow \eta_0 - y = \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2}$$
$$\Leftrightarrow (\eta_0 - y)\psi_2 = (-\eta + \eta_0 + 1)^{\psi_2} - 1$$
$$\Leftrightarrow ((\eta_0 - y)\psi_2 + 1)^{\frac{1}{\psi_2}} = -\eta + \eta_0 + 1$$
$$\Leftrightarrow \eta = \eta_0 + 1 - ((\eta_0 - y)\psi_2 + 1)^{\frac{1}{\psi_2}}$$

```
hpsi2INV<-function(psi2, y, eta0 = 0)
{h <- 1:length(y)
 h[y >= eta0] <- y[y >= eta0]
 if (any(psi2 > -1e-14 && psi2 < 1e-14)){
  h[y < eta0] <- eta0 + 1 - exp(eta0 - y[y < eta0])
 }
 else {
  h[y < eta0] <- eta0 + 1 - ((1 - (psi2 * (y[y < eta0] - eta0)))^(1/psi2))
 }
 h
}
```

Derivative of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ **w.r.t.** η

Now we want to calculate:

$$\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2) = \begin{cases} \frac{\partial}{\partial \eta}(\eta) & \text{if } \eta \geq \eta_0 \\ \frac{\partial}{\partial \eta}(\eta_0 - \ln(-\eta + \eta_0 + 1)) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \frac{\partial}{\partial \eta}\left(\eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2}\right) & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

$$= \begin{cases} 1 & \text{if } \eta \geq \eta_0 \\ -\frac{1}{-\eta + \eta_0 + 1}(-1) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ -(-\eta + \eta_0 + 1)^{(\psi_2 - 1)}(-1) & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

$$
=
\begin{cases}
1 & \text{if } \eta \geq \eta_0 \\
\frac{1}{-\eta+\eta_0+1} & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\
(-\eta + \eta_0 + 1)^{(\psi_2-1)} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0)
\end{cases}
\tag{A.6}
$$

```
hpsi2DERIV<-function(psi2, eta, eta0 = 0)
{h <- 1:length(eta)
 h[eta >= eta0] <- 1
 if (any(psi2 > -1e-14 && psi2 < 1e-14)){
  h[eta < eta0] <- 1/(eta0 + 1 - eta[eta < eta0])
 }
 else{
 h[eta < eta0] <- (eta0 + 1 - eta[eta < eta0])^(psi2 - 1)
 }
 h
}
```

Derivative of $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ w.r.t. ψ_2

Now we want to calculate:

$$
\frac{\partial}{\partial \psi_2} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)
$$

$$
=
\begin{cases}
\frac{\partial}{\partial \psi_2}(\eta) & \text{if } \eta \geq \eta_0 \\
\frac{\partial}{\partial \psi_2}(\eta_0 - \ln(-\eta + \eta_0 + 1)) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\
\frac{\partial}{\partial \psi_2}\left(\eta_0 - \frac{(-\eta+\eta_0+1)^{\psi_2}-1}{\psi_2}\right) & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0)
\end{cases}
$$

$$
=
\begin{cases}
0 & \text{if } \eta \geq \eta_0 \\
\lim\limits_{\psi_2 \to 0} \frac{\partial}{\partial \psi_2}\left(\eta_0 - \frac{(-\eta+\eta_0+1)^{\psi_2}-1}{\psi_2}\right) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\
-\frac{\left(\ln(-\eta+\eta_0+1)(-\eta+\eta_0+1)^{\psi_2}\right)\psi_2 - \left((-\eta+\eta_0+1)^{\psi_2}-1\right)}{\psi_2^2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0)
\end{cases}
$$

$$
=
\begin{cases}
0 & \text{if } \eta \geq \eta_0 \\
\lim\limits_{\psi_2 \to 0} \frac{\partial}{\partial \psi_2}\left(\eta_0 - \frac{(-\eta+\eta_0+1)^{\psi_2}-1}{\psi_2}\right) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\
\frac{(-\eta+\eta_0+1)^{\psi_2}-1-\ln(-\eta+\eta_0+1)(-\eta+\eta_0+1)^{\psi_2}\psi_2}{\psi_2^2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0)
\end{cases}
$$

Remark A.11

The limit $\lim_{\psi_2 \to 0} \frac{\partial}{\partial \psi_2} \left(\eta_0 - \frac{(-\eta+\eta_0+1)^{\psi_2}-1}{\psi_2} \right)$ can be calculated by:

$$\lim_{\psi_2 \to 0} \frac{\partial}{\partial \psi_2} \left(\eta_0 - \frac{(-\eta+\eta_0+1)^{\psi_2}-1}{\psi_2} \right)$$

$$= \lim_{\psi_2 \to 0} \left[\frac{(-\eta+\eta_0+1)^{\psi_2} - 1 - \ln(-\eta+\eta_0+1)(-\eta+\eta_0+1)^{\psi_2}\psi_2}{\psi_2^2} \right]$$

$$\overset{L'Hospital}{=} \lim_{\psi_2 \to 0} \left[\frac{\ln(-\eta+\eta_0+1)(-\eta+\eta_0+1)^{\psi_2}}{2\psi_2} \right.$$

$$\left. - \frac{\left(\ln(-\eta+\eta_0+1)^2(-\eta+\eta_0+1)^{\psi_2}\psi_2 + \ln(-\eta+\eta_0+1)(-\eta+\eta_0+1)^{\psi_2} \right)}{2\psi_2} \right]$$

$$= \lim_{\psi_2 \to 0} \left[- \frac{\ln(-\eta+\eta_0+1)^2(-\eta+\eta_0+1)^{\psi_2}\psi_2}{2\psi_2} \right]$$

$$\overset{L'Hospital}{=} \lim_{\psi_2 \to 0} \left[- \frac{\ln(-\eta+\eta_0+1)^2 \left(\ln(-\eta+\eta_0+1)(-\eta+\eta_0+1)^{\psi_2}\psi_2 + (-\eta+\eta_0+1)^{\psi_2} \right)}{2} \right]$$

$$= \left(- \frac{\ln(-\eta+\eta_0+1)^2}{2} \right) \lim_{\psi_2 \to 0} \left[\ln(-\eta+\eta_0+1) \underbrace{(-\eta+\eta_0+1)^{\psi_2}}_{\to 1} \underbrace{\psi_2}_{\to 0} + \underbrace{(-\eta+\eta_0+1)^{\psi_2}}_{\to 1} \right]$$

$$= - \frac{\ln(-\eta+\eta_0+1)^2}{2}$$

All in all we get:

$$\frac{\partial}{\partial \psi_2} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$$

$$= \begin{cases} 0 & \text{if } \eta \geq \eta_0 \\ - \frac{\ln(-\eta+\eta_0+1)^2}{2} & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \frac{(-\eta+\eta_0+1)^{\psi_2}-1-\ln(-\eta+\eta_0+1)(-\eta+\eta_0+1)^{\psi_2}\psi_2}{\psi_2^2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

```
hpsi2DERIV2<-function(psi2, eta, eta0 = 0)
{h <- eta
 temp <- eta
 h[eta >= eta0] <- 0
 if (any(psi2 > -1e-14 && psi2 < 1e-14)){
 h[eta < eta0] <- - ((log(1 + eta0 - eta[eta < eta0]))^2)/2
 }
 else {
  temp[eta < eta0] <- (1 + eta0 - eta[eta < eta0])^psi2
  h[eta < eta0] <- - ((temp[eta < eta0] *
             log( - eta[eta < eta0] + eta0 + 1) * psi2)
           - (temp[eta < eta0] - 1))/(psi2^2)
 }
 h
}
```

Limit $\lim_{\eta \to -\infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ (for ψ_2 fixed)

Now we want to consider the limit $\lim_{\eta \to -\infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ for fixed values of ψ_2. In the left tail modification the limit $\lim_{\eta \to \infty}$ is of no interest since $h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ is the identity for $\eta \geq \eta_0$.

$$\lim_{\eta \to -\infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2) = \lim_{\eta \to -\infty} \begin{cases} \eta & \text{if } \eta \geq \eta_0 \\ \eta_0 - \ln(-\eta + \eta_0 + 1) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} & \text{otherwise i.e. } \eta < \eta_0 \text{ and } \psi_2 \neq 0 \end{cases}$$

$$= \begin{cases} \lim_{\eta \to -\infty} (\eta_0 - \ln(-\eta + \eta_0 + 1)) & \psi_2 = 0 \\ \lim_{\eta \to -\infty} \left(\eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} \right) & \psi_2 \neq 0 \end{cases} = \begin{cases} \eta_0 - \lim_{\eta \to -\infty} \ln(-\eta + \eta_0 + 1) & \psi_2 = 0 \\ \eta_0 + \frac{1}{\psi_2} - \frac{1}{\psi_2} \lim_{\eta \to -\infty} (-\eta + \eta_0 + 1)^{\psi_2} & \psi_2 \neq 0 \end{cases}$$

$$= \begin{cases} -\infty & \text{if } \psi_2 = 0 \\ \begin{cases} -\infty & \text{if } \psi_2 > 0 \\ \eta_0 + \frac{1}{\psi_2} & \text{if } \psi_2 < 0 \end{cases} & \text{if } \psi_2 \neq 0 \end{cases}$$
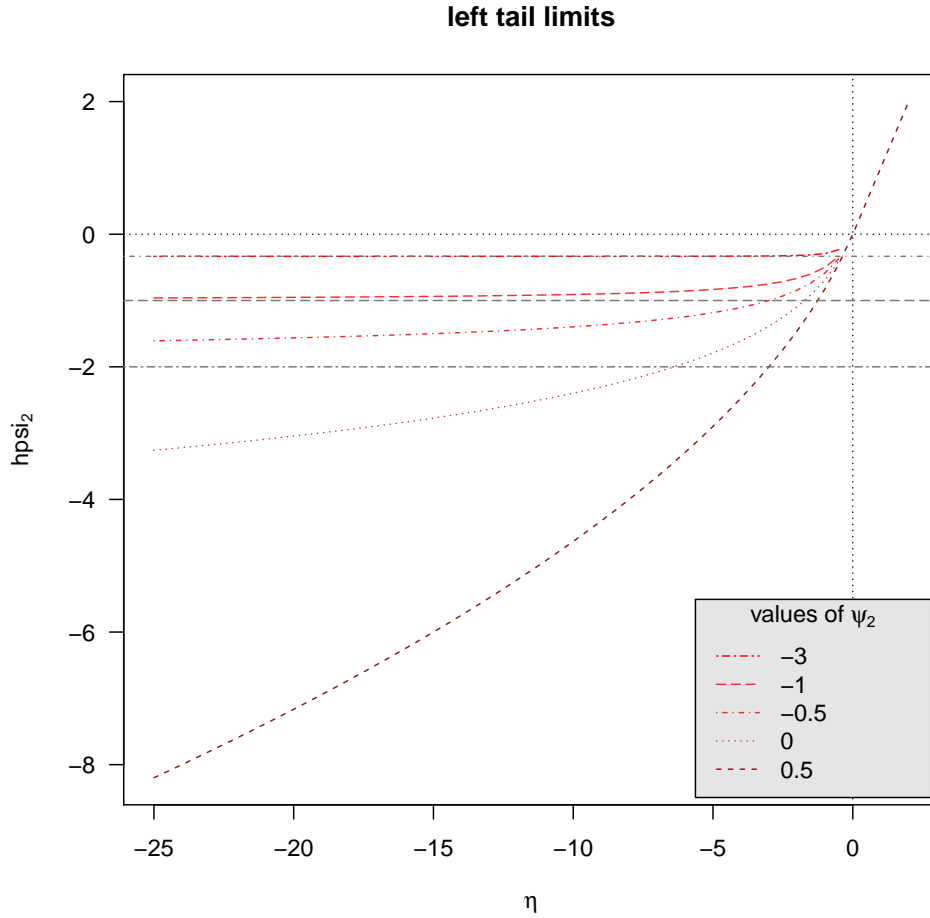

Figure A.2: Limit: $\lim_{\eta \to -\infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ for different values of ψ_2 and $\eta_0 = 0$. The grey dotted horizontal lines indicate the behavior as $\eta \to -\infty$ in case of a finite limit.

A.2.3   Both tail modification

Recall that we defined the both tail modification in Definition 6.4. In the following we want to derive certain functions corresponding to $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$. We also present their implementation in R.

$$h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2)) = \begin{cases} \eta_0 + \ln(\eta - \eta_0 + 1) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \eta_0 - \ln(-\eta + \eta_0 + 1) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

```r
hpsi12 <- function (psi1 = stop("Argument 'psi1' is missing"),
                    psi2 = stop("Argument 'psi2' is missing"),
                    eta = stop("Argument 'eta' is missing"), eta0 = 0)
{h <- 1:length(eta)
 if (any(psi1 > -1e-14 && psi1 < 1e-14)) {
    h[eta >= eta0] <- eta0 + log(eta[eta >= eta0] - eta0 + 1)
 }
 else {
    h[eta >= eta0] <- ((eta[eta >= eta0] - eta0 + 1)^psi1 - 1)/psi1
    h[eta >= eta0] <- h[eta >= eta0] + eta0
 }
 if (any(psi2 > -1e-14 && psi2 < 1e-14)) {
    h[eta < eta0] <- eta0 - log(-(eta[eta < eta0]) + eta0 + 1)
 }
 else {
    h[eta < eta0] <- -((- eta[eta < eta0] + eta0 + 1)^psi2 - 1)/psi2
    h[eta < eta0] <- h[eta < eta0] + eta0
 }
 h
}
```

Domain of $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$

For given $\boldsymbol{\psi} = (\psi_1, \psi_2) \in \mathbb{R}^2$ the domain of $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ can be expressed as a function of η:

$$
\begin{aligned}
domain(h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))) &= \begin{cases} \mathbb{1}_{\{\eta > \eta_0 - 1\}} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ many\ cases & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \mathbb{1}_{\{\eta < \eta_0 + 1\}} & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ many\ cases & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases} \\
&\underset{\substack{take\ easier\ implementation \\ \texttt{is.finite}\ is\ an\ R\ intern\ function}}{=} \mathbb{1}_{\{\texttt{is.finite}(h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2)))\}}
\end{aligned}
\tag{A.7}
$$

Remark A.12

The analysis of the different cases is the same as for the single tail modifications.

Inverse of $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ w.r.t. η

We get:

$$
\eta = \left(h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2)) \right)^{-1}(y)
$$

$$
= \begin{cases}
\exp\{y - \eta_0\} + \eta_0 - 1 & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\
((y - \eta_0)\psi_1 + 1)^{\frac{1}{\psi_1}} + \eta_0 - 1 & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\
\eta_0 + 1 - \exp\{\eta_0 - y\} & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\
\eta_0 + 1 - (1 + (\eta_0 - y)\psi_2)^{\frac{1}{\psi_2}} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0)
\end{cases}
\tag{A.8}
$$

Since:

for the **first** case: $\eta \geq \eta_0$ and $\psi_1 = 0$

$$
y = \eta_0 + \ln\left(\eta - \eta_0 + 1\right)
$$
$$
\Leftrightarrow y - \eta_0 = \ln\left(\eta - \eta_0 + 1\right)
$$
$$
\Leftrightarrow \exp\{y - \eta_0\} = \eta - \eta_0 + 1
$$
$$
\Leftrightarrow \eta = \exp\{y - \eta_0\} + \eta_0 - 1
$$

for the **second** case: $\eta \geq \eta_0$ and $\psi_1 \neq 0$

$$
y = \eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}
$$
$$
\Leftrightarrow y - \eta_0 = \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}
$$
$$
\Leftrightarrow (y - \eta_0)\psi_1 = (\eta - \eta_0 + 1)^{\psi_1} - 1
$$
$$
\Leftrightarrow (y - \eta_0)\psi_1 + 1 = (\eta - \eta_0 + 1)^{\psi_1}
$$
$$
\Leftrightarrow ((y - \eta_0)\psi_1 + 1)^{\frac{1}{\psi_1}} = \eta - \eta_0 + 1
$$
$$
\Leftrightarrow \eta = ((y - \eta_0)\psi_1 + 1)^{\frac{1}{\psi_1}} + \eta_0 - 1
$$

for the **third** case: $\eta < \eta_0$ and $\psi_2 = 0$

$$
y = \eta_0 - \ln\left(-\eta + \eta_0 + 1\right)
$$
$$
\Leftrightarrow \eta_0 - y = \ln\left(-\eta + \eta_0 + 1\right)
$$
$$
\Leftrightarrow \exp\{\eta_0 - y\} = -\eta + \eta_0 + 1
$$
$$
\Leftrightarrow \eta = \eta_0 + 1 - \exp\{\eta_0 - y\}
$$

for the **fourth** case: $\eta < \eta_0$ and $\psi_2 \neq 0$

$$
y = \eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2}
$$
$$
\Leftrightarrow \eta_0 - y = \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2}
$$
$$
\Leftrightarrow (\eta_0 - y)\psi_2 = (-\eta + \eta_0 + 1)^{\psi_2} - 1
$$

$$\Leftrightarrow (\eta_0 - y)\psi_2 + 1 = (-\eta + \eta_0 + 1)^{\psi_2}$$

$$\Leftrightarrow ((\eta_0 - y)\psi_2 + 1)^{\frac{1}{\psi_2}} = -\eta + \eta_0 + 1$$

$$\Leftrightarrow \eta = \eta_0 + 1 - ((\eta_0 - y)\psi_2 + 1)^{\frac{1}{\psi_2}}$$

```r
hpsi12INV<-function(psi1, psi2, y, eta0 = 0)
{h <- 1:length(y)
 if (any(psi1 > -1e-14 && psi1 < 1e-14)){
  h[y >= eta0] <- eta0 - 1 + exp(y[y >= eta0] - eta0)
 }
 else {
  h[y >= eta0] <- eta0 - 1 +
    ((1 + (psi1 * (y[y >= eta0] - eta0)))^(1/psi1))
 }
 if (any(psi2 > -1e-14 && psi2 < 1e-14)){
   h[y < eta0] <- eta0 + 1 - exp(eta0 - y[y < eta0])
 }
 else {
  h[y < eta0] <- eta0 + 1 -
    ((1 - (psi2 * (y[y < eta0] - eta0)))^(1/psi2))
 }
 h
}
```

Derivative of $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ w.r.t. η

Now we want to calculate:

$$\frac{\partial}{\partial \eta} h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2)) = \begin{cases} \frac{\partial}{\partial \eta}\left(\eta_0 + \ln\left(\eta - \eta_0 + 1\right)\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\partial}{\partial \eta}\left(\eta_0 + \frac{(\eta-\eta_0+1)^{\psi_1}-1}{\psi_1}\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \frac{\partial}{\partial \eta}\left(\eta_0 - \ln\left(-\eta + \eta_0 + 1\right)\right) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \frac{\partial}{\partial \eta}\left(\eta_0 - \frac{(-\eta+\eta_0+1)^{\psi_2}-1}{\psi_2}\right) & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

$$= \begin{cases} \frac{1}{\eta-\eta_0+1} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ (\eta - \eta_0 + 1)^{(\psi_1-1)} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ -\frac{1}{-\eta+\eta_0+1}(-1) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ -(-\eta + \eta_0 + 1)^{(\psi_2-1)}(-1) & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

$$= \begin{cases} \frac{1}{\eta-\eta_0+1} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ (\eta - \eta_0 + 1)^{(\psi_1-1)} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \frac{1}{-\eta+\eta_0+1} & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ (-\eta + \eta_0 + 1)^{(\psi_2-1)} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

$$\text{(A.9)}$$

```
hpsi12DERIV<-function(psi1, psi2, eta, eta0 = 0)
{h <- 1:length(eta)
 if (any(psi1 > -1e-14 && psi1 < 1e-14)){
 h[eta >= eta0] <- 1/(eta[eta >= eta0]- eta0 + 1)
 }
 else {
 h[eta >= eta0] <- (1 - eta0 + eta[eta >= eta0])^(psi1 - 1)
 }
 if (any(psi2 > -1e-14 && psi2 < 1e-14)){
 h[eta < eta0] <- 1/(eta0 + 1 - eta[eta < eta0])
 }
 else{
 h[eta < eta0] <- (1 + eta0 - eta[eta < eta0])^(psi2 - 1)
 }
 h
}
```

Derivative of $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ w.r.t. ψ_1

Now we want to calculate (compare to the calculation for the right tail modification):

$$\frac{\partial}{\partial \psi_1} h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$$

$$= \begin{cases} \frac{\partial}{\partial \psi_1}\left(\eta_0 + \ln(\eta - \eta_0 + 1)\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\partial}{\partial \psi_1}\left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ \frac{\partial}{\partial \psi_1}\left(\eta_0 - \ln(-\eta + \eta_0 + 1)\right) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ \frac{\partial}{\partial \psi_1}\left(\eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2}\right) & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

$$= \begin{cases} \lim_{\psi_1 \to 0} \frac{\partial}{\partial \psi_1}\left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}\right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\ln(\eta - \eta_0 + 1)(\eta - \eta_0 + 1)^{\psi_1}\psi_1 - \left((\eta - \eta_0 + 1)^{\psi_1} - 1\right)}{\psi_1^2} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ 0 & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ 0 & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

$$= \begin{cases} \frac{\ln(\eta - \eta_0 + 1)^2}{2} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\ \frac{\ln(\eta - \eta_0 + 1)(\eta - \eta_0 + 1)^{\psi_1}\psi_1 - (\eta - \eta_0 + 1)^{\psi_1} + 1}{\psi_1^2} & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\ 0 & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\ 0 & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0) \end{cases}$$

Remark A.13
Here we calculated the corresponding limit as in Remark A.10. Hence, we have:

$$\lim_{\psi_1 \to 0} \frac{\partial}{\partial \psi_1}\left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1}\right) = \frac{\ln(\eta - \eta_0 + 1)^2}{2}$$

```
hpsi12DERIV1<-function(psi1, psi2, eta, eta0 = 0)
{h <- eta
 temp <- eta
 if (any(psi1 > -1e-14 && psi1 < 1e-14)){
   h[eta >= eta0] <- ((log(eta[eta >= eta0] - eta0 + 1))^2)/2
 }
 else {
  temp[eta >= eta0] <- (1 - eta0 + eta[eta >= eta0])^psi1
  h[eta >= eta0] <- ((temp[eta >= eta0] *
              log(eta[eta >= eta0] - eta0 + 1)
            * psi1) - (temp[eta >= eta0] - 1))/(psi1^2)
 }
 if (any(psi2 > -1e-14 && psi2 < 1e-14)){
   h[eta < eta0] <- 0
 }
 else {
   h[eta < eta0] <- 0
 }
 h
}
```

Derivative of $h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ w.r.t. ψ_2

Now we want to calculate (compare to the calculation for the left tail modification):

$$
\frac{\partial}{\partial \psi_2} h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))
$$

$$
= \begin{cases}
\frac{\partial}{\partial \psi_2} \left(\eta_0 + \ln(\eta - \eta_0 + 1) \right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\
\frac{\partial}{\partial \psi_2} \left(\eta_0 + \frac{(\eta - \eta_0 + 1)^{\psi_1} - 1}{\psi_1} \right) & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\
\frac{\partial}{\partial \psi_2} \left(\eta_0 - \ln(-\eta + \eta_0 + 1) \right) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\
\frac{\partial}{\partial \psi_2} \left(\eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} \right) & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0)
\end{cases}
$$

$$
= \begin{cases}
0 & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\
0 & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\
\lim\limits_{\psi_2 \to 0} \frac{\partial}{\partial \psi_2} \left(\eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} \right) & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\
-\frac{\ln(-\eta + \eta_0 + 1)(-\eta + \eta_0 + 1)^{\psi_2}\psi_2 - ((-\eta + \eta_0 + 1)^{\psi_2} - 1)}{\psi_2^2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0)
\end{cases}
$$

$$
= \begin{cases}
0 & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 = 0 \\
0 & \text{if } \eta \geq \eta_0 \text{ and } \psi_1 \neq 0 \\
-\frac{\ln(-\eta + \eta_0 + 1)^2}{2} & \text{if } \eta < \eta_0 \text{ and } \psi_2 = 0 \\
\frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1 - \ln(-\eta + \eta_0 + 1)(-\eta + \eta_0 + 1)^{\psi_2}\psi_2}{\psi_2^2} & \text{otherwise (i.e. if } \eta < \eta_0 \text{ and } \psi_2 \neq 0)
\end{cases}
$$

Remark A.14

Here we calculated the corresponding limit as in Remark A.11. Hence, we have:

$$\lim_{\psi_2 \to 0} \frac{\partial}{\partial \psi_2} \left(\eta_0 - \frac{(-\eta + \eta_0 + 1)^{\psi_2} - 1}{\psi_2} \right) = -\frac{\ln(-\eta + \eta_0 + 1)^2}{2}$$

```r
hpsi12DERIV2<-function(psi1, psi2, eta, eta0)
{h <- eta
 temp <- eta
 if (any(psi1 > -1e-14 && psi1 < 1e-14)){
  h[eta >= eta0] <- 0
 }
 else {
  h[eta >= eta0] <- 0
 }
 if (any(psi2 > -1e-14 && psi2 < 1e-14)){
  h[eta < eta0] <- - ((log(1 + eta0 - eta[eta < eta0]))^2)/2
 }
 else {
  temp[eta < eta0] <- (1 + eta0 - eta[eta < eta0])^psi2
  h[eta < eta0] <- - ((temp[eta < eta0] *
             log( - eta[eta < eta0] + eta0 + 1)
          * psi2) - (temp[eta < eta0] - 1))/(psi2^2)
 }
 h
}
```

Remark A.15 (Limits $\lim_{|\eta| \to \infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = (\psi_1, \psi_2))$ (for $\boldsymbol{\psi}$ fixed))

In the sections before we studied $\lim_{\eta \to \infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_1)$ and $\lim_{\eta \to -\infty} h_{\eta_0}(\eta, \boldsymbol{\psi} = \psi_2)$ (see Subsection A.2.1 and Subsection A.2.2). For the both tail modification we can adopt the same results as derived for the single tail modifications in the corresponding cases.

Remark A.16 (Implementation of the domain)

In the implementation of the domain of the general $h(\cdot)$ - power transformations we used the R function `is.finite`. Also some of the common link functions in R use this function to describe the domain in the specification of `valideta` (see `make.link("inverse")` and `make.link("sqrt")`).

A.3   Calculation of the deviance in R

Consider again the Gaussian regression model as in Subsection 3.3.1. In Example 5.1 we saw the output of an ordinary Gaussian regression and in Example 6.10 the output of a right tail modified Gaussian regression was presented. In both examples the data set `pcb.ex` was used (see Example 4.1). In the following we will use the common formula to calculate the (unscaled) deviance in the case of a normal distribution (see McCullagh and Nelder (1983)(p. 25)):

$$D(\widehat{\boldsymbol{\mu}}, \boldsymbol{y}) = \sum_{i=1}^{n} (y_i - \widehat{\mu}_i)^2 \tag{A.10}$$

Example A.17 (Verification for the ordinary Gaussian regression)
For the ordinary Gaussian regression the value of the deviance was:

```
ordi_gauss_reg <- glm(formula = log.pcb~age.cen,
                      family = gaussian(link="identity"),
                      data = pcb.ex)
summary(ordi_gauss_reg)$deviance

## [1] 8.359
```

If we verify this value by the formula given in Equation (A.10) we get:

```
# vector of regression coefficients beta
beta <- summary(ordi_gauss_reg)$coefficients[,1]
# need g_inv for the fitted means
g_inv <- make.link("identity")$linkinv
# get the design matrix
X_design <- model.matrix(object=log.pcb~age.cen, data=pcb.ex)
# calculation of the linear predictor in vector form
eta <- X_design%*%beta
# calculation of the vector of fitted means
mu_hat <- g_inv(eta)
# get the values of the response
response <- log.pcb
# auxiliary calculation
auxiliary_calc <- (response-mu_hat)^2
# calculation of the deviance
Deviance_PCB_ord <- sum(auxiliary_calc)
# value of the deviance
Deviance_PCB_ord

## [1] 8.359
```

Example A.18 (Verification for the right tail modified Gaussian regression)
For the right tail modified Gaussian regression the value of the deviance was:

```
tamo_gauss_reg <- glm(formula = log.pcb~age.cen,
                      family = gaussian(link=psi1GAUSS(psi1=0.2,eta0=0)),
                      data = pcb.ex)
summary(tamo_gauss_reg)$deviance

## [1] 6.327
```

If we verify this value by the formula given in Equation (A.10) we get:

```
# vector of regression coefficients beta
beta <- summary(tamo_gauss_reg)$coefficients[,1]
# need g_inv for the fitted means
g_inv <- psi1GAUSS(psi1=0.2,eta0=0)$linkinv
# get the design matrix
X_design <- model.matrix(object=log.pcb~age.cen,  data=pcb.ex)
# calculation of the linear predictor in vector form
eta <- X_design%*%beta
# calculation of the vector of fitted means
mu_hat <- g_inv(eta)
# get the values of the response
response <- log.pcb
# auxiliary calculation
auxiliary_calc <- (response-mu_hat)^2
# calculation of the deviance
Deviance_PCB_tm <- sum(auxiliary_calc)
# value of the deviance
Deviance_PCB_tm

## [1] 6.327
```

A.4   Calculations for the logistic regression model

Example A.19 (Scaled quantities for the logistic regression model)
In Section 3.2 we derived unscaled quantities like Definition 3.15 or Definition 3.16. In the case of a ScaledBinom(n,p) distribution we have $a(\phi_i) = \frac{1}{n_i}$. I.e. the dispersion function depends on the observations (for $i \in \{1, \ldots, n\}$) and hence we won't succeed in finding unscaled quantities. Thus, we follow Czado et al. (2013)(Sections 4.3 and 4.4) to define scaled versions of these definitions for the logistic regression model.

Definition A.20 (Log-likelihood for the logistic regression)
Like in Equation (3.10) we calculate the log likelihood, using the quantities we derived in Subsection A.1.2.

$$l(\boldsymbol{\beta}, \phi \mid \boldsymbol{k}^*) := \sum_{i=1}^{n} \left(\frac{\theta_i k_i^* - b(\theta_i)}{a(\phi)} + c(k_i^*, \phi) \right)$$

$$\overset{Rem.\ A.21}{=} \sum_{i=1}^{n} \left(n_i \left[k_i^* \boldsymbol{x}_i^\top \boldsymbol{\beta} - \ln\left(1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\right) \right] + \ln\left(\binom{n}{nk^*} \right) \right)$$

$$= \sum_{i=1}^{n} n_i \left[k_i^* \boldsymbol{x}_i^\top \boldsymbol{\beta} - \ln\left(1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\right) \right] + \underbrace{\sum_{i=1}^{n} \ln\left(\binom{n}{nk^*} \right)}_{:=\ const.\ C\ indep.\ of\ \boldsymbol{\beta}}$$

$$= \sum_{i=1}^{n} n_i \left[k_i^* \boldsymbol{x}_i^\top \boldsymbol{\beta} - \ln\left(1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\right) \right] + C \tag{A.11}$$

Remark A.21 (Auxiliary calculation)
The quantities in Definition A.20 can be calculated by:

$$a(\phi) \overset{see\ Sec.\ A.1.2}{:=} \frac{1}{n_i}; \qquad \theta_i \overset{see\ Sec.\ A.1.2}{:=} \ln\left(\frac{p_i}{1 - p_i} \right);$$

$$p_i \overset{see\ Def.\ 3.35}{=} \frac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}};$$

$$\Rightarrow \theta_i = \ln\left(\frac{\frac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{1+\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}}{1 - \frac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{1+\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}} \right) = \ln\left(\frac{\frac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{1+\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}}{\frac{1}{1+\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}} \right) = \ln\left(\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\} \right) = \boldsymbol{x}_i^\top \boldsymbol{\beta}$$

$$\Rightarrow b(\theta_i) = \ln\left(1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\right)$$

Definition A.22 (Scaled score equations for logistic regression)
For $j \in \{1, \ldots, p\}$ we define

$$\frac{\partial l(\boldsymbol{\beta} \mid \boldsymbol{k}^*)}{\partial \beta_j} \overset{see\ Eq.\ (A.11)}{=} \sum_{i=1}^{n} n_i x_{ij} \left[k_i^* - \frac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}} \right]$$

$$\overset{p_i = \frac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{1+\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}}{=} \sum_{i=1}^{n} n_i x_{ij} \left[k_i^* - p_i \right] = \sum_{i=1}^{n} n_i x_{ij} \left[k_i^* - \mathbb{E}\left[\frac{Y_i}{n_i} \right] \right]$$

Remark A.23 (Partial derivative w.r.t. β_j)
Definition A.22 is the partial derivative of $l(\boldsymbol{\beta} \mid \boldsymbol{k}^*)$ w.r.t β_j, since:

$$\frac{\partial l(\boldsymbol{\beta} \mid \boldsymbol{k}^*)}{\partial \beta_j} \overset{see\ Eq.\ (A.11)}{=} \frac{\partial}{\partial \beta_j} \left(\sum_{i=1}^{n} n_i \left[k_i^* \boldsymbol{x}_i^\top \boldsymbol{\beta} - \ln\left(1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\right) \right] \right)$$

$$= \sum_{i=1}^{n} n_i \left[\underbrace{\frac{\partial}{\partial \beta_j} \left(k_i^* \boldsymbol{x}_i^\top \boldsymbol{\beta} \right)}_{=x_{ij} k_i^*} - \underbrace{\frac{\partial}{\partial \beta_j} \ln\left(1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}\right)}_{=\frac{1}{1+\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}} \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\} x_{ij}} \right]$$

$$= \sum_{i=1}^{n} n_i x_{ij} \left[k_i^* - \frac{\exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_i^\top \boldsymbol{\beta}\}} \right]$$

Remark A.24 (Calculation of elements of the Hessian matrix)

The (i,j) - th element of the scaled Hessian matrix is given by:

$$
\frac{\partial^2 l(\boldsymbol{\beta} \mid \boldsymbol{k}^*)}{\partial \beta_i \partial \beta_j} \overset{see\ Def.\ A.22}{=} \frac{\partial}{\partial \beta_i} \left(\sum_{z=1}^{n} n_z x_{zj} \left[k_z^* - \frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}} \right] \right)
$$

$$
= \sum_{z=1}^{n} \left(n_z x_{zj} \frac{\partial}{\partial \beta_i} \left[k_z^* - \frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}} \right] \right)
$$

Now the partial derivative $\frac{\partial}{\partial \beta_i} \left[\frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}}{1+\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}} \right]$ is given by:

$$
\frac{\partial}{\partial \beta_i} \left[\frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}} \right]
$$

$$
= \frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\} x_{zi}(1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}) - (\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\})^2 x_{zi}}{(1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\})^2}
$$

$$
= \frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\} x_{zi}}{(1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\})^2}
$$

Therefore, we get:

$$
\frac{\partial^2 l(\boldsymbol{\beta} \mid \boldsymbol{k}^*)}{\partial \beta_i \partial \beta_j}
$$

$$
= \sum_{z=1}^{n} \left(n_z x_{zj} \frac{\partial}{\partial \beta_i} \left[k_z^* - \frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}} \right] \right)
$$

$$
= \sum_{z=1}^{n} \left(n_z x_{zj} \left[-\frac{\partial}{\partial \beta_i} \left(\frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}}{1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}} \right) \right] \right)
$$

$$
= \sum_{z=1}^{n} \left(n_z x_{zj} \left[-\frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\} x_{zi}}{(1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\})^2} \right] \right)
$$

$$
= -\sum_{z=1}^{n} n_z x_{zj} x_{zi} \frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}}{(1 + \exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\})^2}
$$

$$
\overset{\frac{\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}}{\left(1+\exp\{\boldsymbol{x}_z^\top \boldsymbol{\beta}\}\right)^2}=p_z(1-p_z)}{=} -\sum_{z=1}^{n} n_z p_z (1 - p_z) x_{zi} x_{zj}
$$

With $d_i := n_i p_i (1 - p_i)$ we can define

$$
D(\boldsymbol{\beta}) := diag(d_1, \ldots, d_n) = \begin{pmatrix} d_1 & & & & \\ & d_2 & & \text{\Large 0} & \\ & & \ddots & & \\ & \text{\Large 0} & & d_{n-1} & \\ & & & & d_n \end{pmatrix} \in \mathbb{R}^{n \times n} \qquad (A.12)
$$

Using this notation we can express the *scaled Hessian matrix* through, where $X^\top$ is given as in Remark 3.19 and $\boldsymbol{x}_i$ defined as in Equation (3.6):

$$H(\boldsymbol{\beta}) = \nabla^2(\boldsymbol{\beta}) := \left(\frac{\partial^2 l(\boldsymbol{\beta} \mid \boldsymbol{k}^*)}{\partial \beta_i \partial \beta_j} \right)_{i,j \in \{1,\ldots,p\}} = -X^\top D(\boldsymbol{\beta}) X \tag{A.13}$$

Let us denote by $\boldsymbol{\mu}$ the vector $(\mu_1, \ldots, \mu_n)^\top \overset{\sim ScaledBinom}{=} (p_1, \ldots, p_n)^\top$ and $\nabla(\boldsymbol{\beta}) := \frac{\partial l(\boldsymbol{\beta}|\boldsymbol{k}^*)}{\partial \boldsymbol{\beta}} = \left(\frac{\partial l(\boldsymbol{\beta}|\boldsymbol{k}^*)}{\partial \beta_1}, \ldots, \frac{\partial l(\boldsymbol{\beta}|\boldsymbol{k}^*)}{\partial \beta_p} \right)$. Consequently, we can rewrite Definition A.22 in the following way:

$$\nabla(\boldsymbol{\beta}) = X^\top \underbrace{\begin{pmatrix} n_1 & & & & \\ & n_2 & & \text{\Large 0} & \\ & & \ddots & & \\ \text{\Large 0} & & & n_{n-1} & \\ & & & & n_n \end{pmatrix}}_{:=D_n} \underbrace{(\boldsymbol{k}^* - \boldsymbol{\mu})}_{:=\varepsilon^*} = X^\top D_n \varepsilon^*$$

Example A.25 (Ordinary logistic regression model)
Recall that we had for the logistic regression model:

$$g(\mu_i) = \ln\left(\frac{\mu_i}{1 - \mu_i} \right) = \ln\left(\frac{p_i}{1 - p_i} \right) = \boldsymbol{x}_i^\top \boldsymbol{\beta} = \eta_i(\boldsymbol{\beta}) = \eta_i \tag{A.14}$$

Therefore, we get:

$$\begin{aligned} \frac{\partial \eta_i}{\partial \mu_i} &= \frac{\partial}{\partial \mu_i} g(\mu_i) = \frac{\partial}{\partial \mu_i} \ln\left(\frac{\mu_i}{1 - \mu_i} \right) \\ &= \frac{1}{\frac{\mu_i}{1-\mu_i}} \frac{1 \cdot (1 - \mu_i) - (\mu_i \cdot (-1))}{(1 - \mu_i)^2} = \frac{1 - \mu_i}{\mu_i} \frac{1}{(1 - \mu_i)^2} \\ &= \frac{1}{\mu_i(1 - \mu_i)} \overset{\mu_i = p_i \; see \; Def. \; 3.35}{=} \frac{1}{p_i(1 - p_i)} \end{aligned} \tag{A.15}$$

For the adjusted dependent variable $Z_i^{\boldsymbol{\beta}}$ we get:

$$\begin{aligned} Z_i^{\boldsymbol{\beta}} &= \eta_i + (Y_i^* - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i} \right) \\ &\overset{\substack{see \; Eq. \; (A.14) \\ Def. \; 3.35 \; and \; Eq. \; (A.15)}}{=} \boldsymbol{x}_i^\top \boldsymbol{\beta} + (Y_i^* - p_i) \frac{1}{p_i(1 - p_i)} \\ &= \eta_i + \frac{Y_i - n_i p_i}{n_i p_i(1 - p_i)} \end{aligned} \tag{A.16}$$

Also Collett (1999)(p. 342ff., see in particular p. 346) describes this results for the logistic regression model.

Remark A.26
According to the remark in Remark 3.25 $Z_i^{\boldsymbol{\beta}}$ is the new response (and not Y_i or Y_i^* respectively) and the maximum likelihood estimates are calculated iteratively as weighted least squares estimates of the new response (see Definition 3.26).

In matrix-vector notation Equation (A.16) equals:

$$\boldsymbol{Z}^{\boldsymbol{\beta}} \overset{Eq.\ (3.5)}{=} X\boldsymbol{\beta} + \begin{pmatrix} \frac{1}{p_1(1-p_1)} & & & \\ & \frac{1}{p_2(1-p_2)} & & \boldsymbol{0} \\ & & \ddots & \\ \boldsymbol{0} & & \frac{1}{p_{n-1}(1-p_{n-1})} & \\ & & & \frac{1}{p_n(1-p_n)} \end{pmatrix} \left[\boldsymbol{Y}^* - \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_n \end{pmatrix} \right]$$

$$\overset{Def.\ 3.38}{=} X\boldsymbol{\beta} + \begin{pmatrix} \frac{1}{p_1(1-p_1)} & & & \\ & \frac{1}{p_2(1-p_2)} & & \boldsymbol{0} \\ & & \ddots & \\ \boldsymbol{0} & & \frac{1}{p_{n-1}(1-p_{n-1})} & \\ & & & \frac{1}{p_n(1-p_n)} \end{pmatrix} \left[\begin{pmatrix} \frac{Y_1}{n_1} \\ \frac{Y_2}{n_2} \\ \vdots \\ \frac{Y_{n-1}}{n_{n-1}} \\ \frac{Y_n}{n_n} \end{pmatrix} - \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_{n-1} \\ p_n \end{pmatrix} \right]$$

$$= X\boldsymbol{\beta} + \begin{pmatrix} \frac{1}{p_1(1-p_1)} & & & \\ & \frac{1}{p_2(1-p_2)} & & \boldsymbol{0} \\ & & \ddots & \\ \boldsymbol{0} & & \frac{1}{p_{n-1}(1-p_{n-1})} & \\ & & & \frac{1}{p_n(1-p_n)} \end{pmatrix}$$

$$\begin{pmatrix} \frac{1}{n_1} & & & \\ & \frac{1}{n_2} & & \boldsymbol{0} \\ & & \ddots & \\ \boldsymbol{0} & & \frac{1}{n_{n-1}} & \\ & & & \frac{1}{n_n} \end{pmatrix} \left[\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_{n-1} \\ Y_n \end{pmatrix} - \begin{pmatrix} n_1 p_1 \\ n_2 p_2 \\ \vdots \\ n_{n-1} p_{n-1} \\ n_n p_n \end{pmatrix} \right]$$

$$= X\boldsymbol{\beta} + \underbrace{\begin{pmatrix} \frac{1}{n_1 p_1(1-p_1)} & & & \\ & \frac{1}{n_2 p_2(1-p_2)} & & \boldsymbol{0} \\ & & \ddots & \\ \boldsymbol{0} & & \frac{1}{n_{n-1}p_{n-1}(1-p_{n-1})} & \\ & & & \frac{1}{n_n p_n(1-p_n)} \end{pmatrix}}_{\overset{Eq.\ (A.12)}{=}D^{-1}(\boldsymbol{\beta})} \underbrace{\left[\boldsymbol{Y} - \begin{pmatrix} n_1 p_1 \\ n_2 p_2 \\ \vdots \\ n_{n-1}p_{n-1} \\ n_n p_n \end{pmatrix} \right]}_{:=\varepsilon}$$

$$= X\boldsymbol{\beta} + \underbrace{D^{-1}(\boldsymbol{\beta})\varepsilon}_{:=\varepsilon_{adj}} \tag{A.17}$$

Remark A.27 (Unscaled and centered around mean)
Notice that ε (in Equation (A.17)) represents the unscaled and mean centered responses.

Remark A.28 (Derivation of the weights)
We brought the response $\boldsymbol{Z}^{\boldsymbol{\beta}}$ in the form of Definition 3.26 (with error term ε_{adj}). Hence, we need to determine the expectation and the variance-covariance matrix W to know what the proper weights are. We obtain:

$$\mathbb{E}\left[\boldsymbol{\varepsilon}\right] = \mathbb{E}\left[\boldsymbol{Y} - \begin{pmatrix} n_1p_1 \\ n_2p_2 \\ \vdots \\ n_{n-1}p_{n-1} \\ n_np_n \end{pmatrix}\right] \overset{Y_i \sim Bin(n_i,p_i)}{=} \boldsymbol{\mu} \begin{pmatrix} n_1p_1 - n_1p_1 \\ n_2p_2 - n_2p_2 \\ \vdots \\ n_{n-1}p_{n-1} - n_{n-1}p_{n-1} \\ n_np_n - n_np_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} = \boldsymbol{0} \in \mathbb{R}^n$$

$$W = \left(\mathbb{C}ov[\varepsilon_i, \varepsilon_j]\right)_{i,j \in \{1,\ldots,n\}}$$

$$\overset{\mathbb{E}[\varepsilon]=\boldsymbol{0}}{=} \left(\mathbb{E}\left[\varepsilon_i\varepsilon_j\right]\right)_{i,j \in \{1,\ldots,n\}} = \left(\mathbb{E}\left[(Y_i - n_ip_i)(Y_j - n_jp_j)\right]\right)_{i,j \in \{1,\ldots,n\}}$$

$$= \left(\mathbb{E}\left[Y_iY_j - n_jp_jY_i - n_ip_iY_j + n_ip_in_jp_j\right]\right)_{i,j \in \{1,\ldots,n\}}$$

$$= \left(\mathbb{E}\left[Y_iY_j\right] - n_jp_j\underbrace{\mathbb{E}\left[Y_i\right]}_{=n_ip_i} - n_ip_i\underbrace{\mathbb{E}\left[Y_j\right]}_{=n_jp_j} + n_ip_in_jp_j \right)_{i,j \in \{1,\ldots,n\}}$$

at the underbrace: $\underbrace{\phantom{-n_jp_j\mathbb{E}[Y_i]-n_ip_i\mathbb{E}[Y_j]+n_ip_in_jp_j}}_{=-n_ip_in_jp_j}$

$$= \left(\mathbb{E}\left[Y_iY_j\right] - n_ip_in_jp_j\right)_{i,j \in \{1,\ldots,n\}}$$

$$= \begin{cases} \mathbb{E}\left[Y_i\right]\mathbb{E}\left[Y_j\right] - n_ip_in_jp_j & \text{if } i \neq j, \text{ since } Y_i \text{ and } Y_j \text{ are indep.} \\ \mathbb{E}\left[Y_i^2\right] - (n_ip_i)^2 & \text{if i = j} \end{cases}$$

$$= \begin{cases} 0 & \text{if } i \neq j, \text{ since } Y_i \text{ and } Y_j \text{ are indep.} \\ \left(\underbrace{\mathbb{V}ar\left[Y_i\right]}_{\underset{Rem. A.4}{=}n_ip_i(1-p_i)} + \underbrace{\mathbb{E}\left[Y_i\right]^2}_{\underset{Rem. A.4}{=}(n_ip_i)^2} \right) - (n_ip_i)^2 & \text{if i = j} \end{cases}$$

$$= \begin{cases} 0 & \text{if } i \neq j, \text{ since } Y_i \text{ and } Y_j \text{ are indep.} \\ n_ip_i(1-p_i) & \text{if i = j} \end{cases}$$

$$\Rightarrow W = \begin{pmatrix} n_1p_1(1-p_1) & & & & \\ & n_2p_2(1-p_2) & & \mathbf{0} & \\ & & \ddots & & \\ & \mathbf{0} & & n_{n-1}p_{n-1}(1-p_{n-1}) & \\ & & & & n_np_n(1-p_n) \end{pmatrix}$$

$$\overset{Eq. (A.12)}{=} D(\boldsymbol{\beta})$$

Remark A.29
The components on the diagonal of W are the variances of Y_i for $i \in \{1,\ldots,n\}$.

Remark A.30 (Variance-covariance matrix of $\boldsymbol{\varepsilon}_{adj}$)
With the calculation above we can derive the variance-covariance matrix of $\boldsymbol{\varepsilon}_{adj}$ by:

$$W_{adj} = \mathbb{C}ov[\boldsymbol{\varepsilon}_{adj}] = \mathbb{C}ov[D^{-1}(\boldsymbol{\beta})\boldsymbol{\varepsilon}] = D^{-1}(\boldsymbol{\beta})\underbrace{\mathbb{C}ov[\boldsymbol{\varepsilon}]}_{=D(\boldsymbol{\beta})}\left(D^{-1}(\boldsymbol{\beta})\right)^{\top}$$

$$\overset{D^{-1}(\boldsymbol{\beta})}{=} \overset{sym.}{\underbrace{D^{-1}(\boldsymbol{\beta})D(\boldsymbol{\beta})}_{=\mathbb{I}_n}} D^{-1}(\boldsymbol{\beta}) = D^{-1}(\boldsymbol{\beta})$$

We used $\mathbb{C}ov[A\boldsymbol{z}] = A\mathbb{C}ov[\boldsymbol{z}]A^\top$ for $A \in \mathbb{R}^{n \times n}$ and $\boldsymbol{z} \in \mathbb{R}^n$ (see Seber (1977)(p. 11)).

Remark A.31 (Iterative values of MLE $\widehat{\boldsymbol{\beta}}$)
Given we have calculated the value $\boldsymbol{\beta^s}$, then, by Remark 3.25, we get the value $\boldsymbol{\beta^{s+1}}$ in the (s+1)-th iteration by (using above's formulas):

$$\begin{aligned}
\boldsymbol{\beta^{s+1}} &= (X^\top W_{adj}^{-1}X)^{-1}X^\top W_{adj}^{-1}\boldsymbol{Z^{\beta^s}} \\
&= (X^\top D(\boldsymbol{\beta^s})X)^{-1}X^\top D(\boldsymbol{\beta^s})\boldsymbol{Z^{\beta^s}}
\end{aligned}$$

Therefore, the final maximum likelihood estimates $\widehat{\boldsymbol{\beta}}$ fulfill:

$$\widehat{\boldsymbol{\beta}} = (X^\top D(\widehat{\boldsymbol{\beta}})X)^{-1}X^\top D(\widehat{\boldsymbol{\beta}})\boldsymbol{Z^{\widehat{\beta}}}$$

Example A.32 (Tail modified logistic regression model)
Recall that we had for the tail modified logistic regression (see Remark 6.15):

$$g(\mu_i, \boldsymbol{\psi}) = (h_{\eta_0}(\eta_i, \boldsymbol{\psi}))^{-1}\left(\ln\left(\frac{\mu_i}{1-\mu_i}\right)\right) = \eta_i$$

Therefore, we get:

$$\begin{aligned}
\frac{\partial \eta_i}{\partial \mu_i} &= \frac{\partial}{\partial \mu_i}g(\mu_i) = \frac{\partial}{\partial \mu_i}\left[(h_{\eta_0}(\eta_i, \boldsymbol{\psi}))^{-1}\left(\ln\left(\frac{\mu_i}{1-\mu_i}\right)\right)\right] \\
&= \frac{1}{\left(\frac{\partial}{\partial \eta_i}h_{\eta_0}\right)\left[(h_{\eta_0}(\eta_i, \boldsymbol{\psi}))^{-1}\left(\ln\left(\frac{\mu_i}{1-\mu_i}\right)\right)\right]}\underbrace{\frac{\partial}{\partial \mu_i}\left(\ln\left(\frac{\mu_i}{1-\mu_i}\right)\right)}_{\underset{see\ Eq.\ (A.15)}{=}\frac{1}{p_i(1-p_i)}}
\end{aligned} \quad (A.18)$$

We also have that (see Table 3.2):

$$b''(\theta_i) = \frac{1}{(1+\exp\{\theta_i\})^2} = b'(\theta_i)(1 - b'(\theta_i)) = p_i(1 - p_i)$$

For the adjusted dependent variable $Z_i^{\boldsymbol{\beta}}$ we get:

$$\begin{aligned}
Z_i^{\boldsymbol{\beta}} &= \eta_i + (Y_i^* - \mu_i)\left(\frac{\partial \eta_i}{\partial \mu_i}\right) \\
&\overset{\underset{Def.\ 3.35}{Def.\ 3.4}}{\underset{Eq.\ (A.18)}{=}} \boldsymbol{x}_i^\top \boldsymbol{\beta} + (Y_i^* - p_i)\frac{1}{\left(\frac{\partial}{\partial \eta_i}h_{\eta_0}\right)\left[(h_{\eta_0}(\eta_i, \boldsymbol{\psi}))^{-1}\left(\ln\left(\frac{\mu_i}{1-\mu_i}\right)\right)\right]}\frac{1}{p_i(1-p_i)}
\end{aligned}$$

The weights are given by (see Definition 3.27):

$$
W_i^{\boldsymbol{\beta}} = \left[\overbrace{p_i(1-p_i)}^{=b''(\theta_i)} \overbrace{\left(\frac{1}{\left(\frac{\partial}{\partial \eta_i} h_{\eta_0}\right) \left[(h_{\eta_0}(\eta_i, \boldsymbol{\psi}))^{-1} \left(\ln\left(\frac{\mu_i}{1-\mu_i}\right) \right) \right]} \frac{1}{p_i(1-p_i)} \right)^2}^{=\left(\frac{\partial \eta_i}{\partial \mu_i}\right)^2} \right]^{-1}
$$

$$
= \left[\frac{1}{p_i(1-p_i)} \frac{1}{\left(\left(\frac{\partial}{\partial \eta_i} h_{\eta_0}\right) \left[(h_{\eta_0}(\eta_i, \boldsymbol{\psi}))^{-1} \left(\ln\left(\frac{\mu_i}{1-\mu_i}\right) \right) \right] \right)^2} \right]^{-1}
$$

$$
= p_i(1-p_i) \left(\left(\frac{\partial}{\partial \eta_i} h_{\eta_0}\right) \left[(h_{\eta_0}(\eta_i, \boldsymbol{\psi}))^{-1} \left(\ln\left(\frac{\mu_i}{1-\mu_i}\right) \right) \right] \right)^2
$$

Remark A.33 (Partial derivatives w.r.t. $\boldsymbol{\eta_i}$)
The derivatives $\frac{\partial}{\partial \eta_i} h_{\eta_0}$ are calculated in Section A.2.

A.5   Abbreviations and Notation

A.5.1   Abbreviations

The abbreviations in the following table are commonly used:

Symbol	Abbreviation	Explanation
f	*pmf*	probability mass function (for discrete distributions)
f	*pdf*	probability distribution function (for continuous distributions)
F	*cdf*	cumulative distribution function
	GLM	generalized linear model
	GLMs	generalized linear models
	MLE	maximum likelihood estimator or estimate
	IWLS	Iterative weighted least squares
	i.i.d.	independent and identically distributed
	w.r.t.	with respect to
	w.l.o.g.	without loss of generality
	Res. or *res.*	restriction
	s.t.	such that or so that

A.5.2   Notation

Throughout the thesis we used the following notation:

Symbol	Explanation	Represents
A $\\$ Y	capital letter (majuscule)	a matrix, i.e. $A \in \mathbb{R}^{m \times n}$ $(m, n \in \mathbb{N})$ depending on the context Y can also describe a random variable
x $\\$ y	small letter (minuscule)	a real value, i.e. $x \in \mathbb{R}$ depending on the context y can also describe a realization of a random variable, i.e. $Y = y$ means that the random variable Y takes the value y in this specific case
$\boldsymbol{x}$	small letter in bold print	a vector of real numbers, i.e. $\boldsymbol{x} = (x_1, x_2, \ldots, x_{n-1}, x_n) \in \mathbb{R}^n$
$\boldsymbol{Y}$	capital letter in bold print	a vector of random variables, i.e. $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_{n-1}, Y_n) \in \mathbb{R}^n$
$\widehat{\boldsymbol{\beta}}$	hat over letter in bold print	depending on the context $\widehat{\beta}$ can represent a random vector (referred to as *estimator*) or a specific outcome of the estimator regarding certain data (referred to as *estimates*)

References

Bailey, R. A. and L. J. Simon (1959). An Actuarial Note on the Credibility of Experience of a Single Private Passenger Car. *PCAS XLVI 19*.

Bailey, R. A. and L. J. Simon (1960). Two studies in automobile insurance ratemaking. *Astin Bulletin 1* (4), pp. 192 − 217.

Bates, D. M. and D. G. Watts (2007). *Nonlinear Regression Analysis and Its Applications*. John Wiley & Sons, Inc.

Bliss, C. I. (1935). The calculation of the dosage-mortality curve. *Annals of Applied Biology, 22*, pp. 134 − 167.

Collett, D. (1999). *Modelling Binary Data*. Chapman & Hall/CRC.

Czado, C. (1992). On Link Selection in Generalized Linear Models. In L. Fahrmeir, B. Francis, R. Gilchrist, and G. Tutz (Eds.), *Advances in GLIM and Statistical Modelling*, Lecture Notes in Statistics, Volume 78, pp. 60 − 65. Springer, New York.

Czado, C. (2007). Fitting Generalized Linear Models with Parametric Link in S. Technical report, York University, North York, Ontario, Canada.

Czado, C., N. Krämer, and E. C. Brechmann (2013). Generalized linear models with applications. Unpublished, as of November 15, 2013.

Czado, C. and T. Schmidt (2011). *Mathematische Statistik*. Springer.

Dobson, A. J. (1999). *An Introduction to Generalized Linear Models*. Chapman & Hall/CRC.

Fahrmeir, L. and G. Tutz (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models* (2nd ed.). Springer, New York.

Hardin, J. W. and J. M. Hilbe (2007). *Generalized Linear Models and Extensions* (2nd ed.). Stata Press.

Higgins, J. E. and G. G. Koch (1977). Variable Selection and Generalized Chi-square Analysis of Categorical Data applied to a Large Cross-sectional Occupational Health Survey. *International Statistical Review 45*, pp. 51 − 62.

Ligges, U. (2006). R Help Desk - Accessing the Sources. *RNews Volume 6/4 (October)*, pp. 43 − 45.

McCullagh, P. and J. A. Nelder (1983). *Generalized Linear Models*. Chapman and Hall.

Myers, R. H. (1990). *Classical and Modern Regression with Applications* (2nd ed.). PWS-Kent Publishing Company.

Myers, R. H., D. C. Montgomery, and G. G. Vining (2002). *Generalized Linear Models: With Applications in Engineering and the Sciences*. John Wiley & Sons, Inc.

Seber, G. A. F. (1977). *Linear Regression Analysis*. John Wiley & Sons, Inc.

Ulbrich, M. and S. Ulbrich (2012). *Nichtlineare Optimierung*. Springer Basel AG.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Chapman & Hall/CRC.

Xie, Y. (2013). *Dynamic Documents with R and knitr*. CRC Press.