

10 Poisson, negative binomial, binomial and gamma GLMs in R-INLA

In this chapter we explain how to apply Poisson, negative binomial (NB), Bernoulli, binomial, and gamma generalised linear models (GLM) in R-INLA. Using R-INLA for such models is certainly overkill as it is more convenient to use the `glm` function, but it prepares us for things that come later in this book.



Prerequisite for this chapter: You need to be familiar with GLMs; see for example Zuur et al. (2007; 2009a; 2013) or Hilbe (2014).

10.1 Poisson and negative binomial GLMs in R-INLA

10.1.1 Introduction

In this section we use data from Timi et al. (2008), who analysed abundances of parasites infecting the body surface and gills of the Brazilian sand perch *Pinguipes brasiliensis*. Fish samples were collected in three main areas of the Argentine Sea. We also have information on sex, length, and weight of the fish, and we would like to know which of these variables are useful for explaining the numbers of parasites in a fish. The data were also used in Zuur et al. (2016a) to compare models with a Poisson, negative binomial, and generalised Poisson distribution.

We import the data from the text file `Turcoparasitos.txt`.

```
> TP <- read.table(file = "Turcoparasitos.txt",
                  header = TRUE,
                  dec = ".")
```

A detailed data exploration was carried out following the protocol described in Zuur et al. (2010). We made Cleveland dotplots to investigate the presence of potential outliers. None were present. We also made multi-panel scatterplots and calculated variance inflation factors for the covariates. Not surprisingly, there was high collinearity between fish length, fish weight, and sagittal length. We selected fish length (abbreviated as `LT` in the data file) for the analysis. Figure 10.1 shows a scatterplot of the number of fish versus length for each of the three locations; we have enough observations per location to fit a model with an interaction between length and location (location represents temperature and we expect that the length effect on the number of parasites changes with temperature). Although the three lines in the graph clearly show a difference, we need to apply a model with an interaction term to prove that there is an important difference between the locations.

The R code for these steps of the data exploration are given on the website for this book.

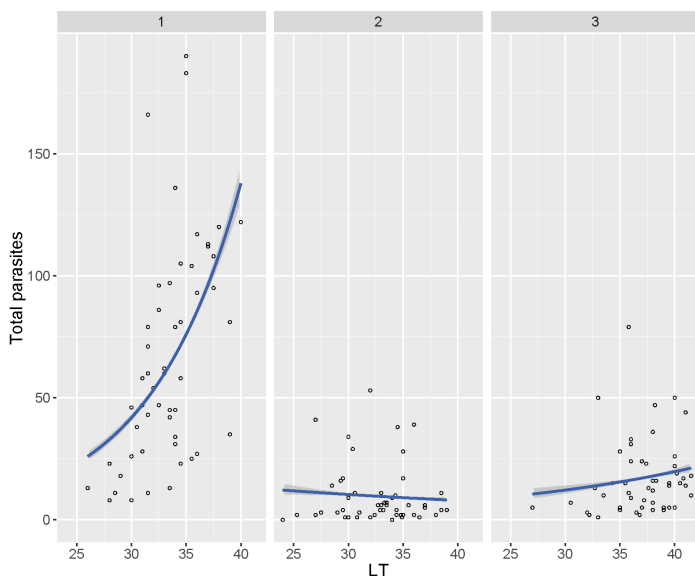


Figure 10.1. Scatterplot of total number of parasites per fish plotted versus length of the fish. Each panel corresponds to a location. In each panel a curve obtained from a Poisson GLM was added.

10.1.2 Poisson GLM in R-INLA

To investigate whether the relationship between total numbers of parasites and length differs per location, while taking into account a sex effect, we apply the following Poisson GLM.

$$\begin{aligned}
 TP_i &\sim \text{Poisson}(\mu_i) \\
 E(TP_i) &= \mu_i \quad \text{and} \quad \text{var}(TP_i) = \mu_i \\
 \log(\mu_i) &= \text{Intercept} + \text{Sex}_i + LT_i + \text{Location}_i + LT_i \times \text{Location}_i
 \end{aligned}
 \tag{10.1}$$

TP_i is the total number of parasites in fish i , where $i = 1, \dots, 155$. LT_i is the length of fish i . The other variables are self-explanatory. Before we can execute the model we define the two categorical covariates as factors in R.

```

> TP$fSex <- factor(TP$SEX)
> TP$fLoc <- factor(TP$Location)

```

We are now ready to execute the model in R-INLA. The `dic = TRUE` argument in the `control.compute` option ensures that we obtain a DIC, which we can use for model selection.