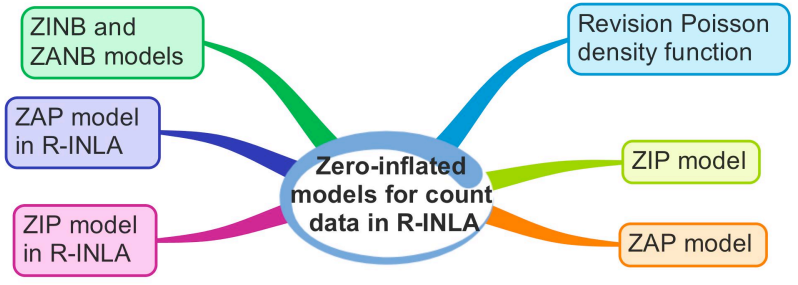
18 Zero-inflated models for count data in R-INLA

In this chapter we will analyse count data with an excessive number of zeros. Various types of models exist for the analysis of such data. Here, we will apply so-called zero-inflated Poisson (ZIP) models and zero-altered Poisson (ZAP) models. The latter ones are also called hurdle models. We also discuss the negative binomials versions of these models, namely the zero-inflated negative binomial (ZINB) model and the zero-altered negative binomial (ZANB) model.

The flowchart below gives a visual outline of this chapter. We will start with a short revision of the Poisson density function and then provide a theoretical explanation of ZIP and ZAP models. Zero-inflated models and zero-altered models are useful tools, but at times they can also be very confusing. It is important to fully comprehend what these models do. We therefore explain them using simulated data. In the second part of the chapter we show how to implement these models in R-INLA using a real data set.



The benefit of applying these models in R-INLA (and not, for example, with the `pscl` or `glmmTMB` packages) is that we can easily extend them to allow for spatial or spatial-temporal dependency. We will show how to do this in Chapter 19.

Prerequisite for this chapter: We assume that you are familiar with the material explained in Chapters 10 and 13 of Volume I (applying Poisson GLM with spatial correlation in R-INLA).

Before we explain models that are especially designed for a response variable with a large number of zeros, we would like to emphasise that a covariate might also be able to model the large number of zeros. Hence, one should never start the statistical analysis with a zero-inflated model, but first try an ordinary generalised linear model (GLM).