

# Model-based Quantile Regression for Discrete Data

Tullia Padellini<sup>1</sup> and Håvard Rue<sup>2</sup>

<sup>1</sup>Dipartimento di Scienze Statistiche – Sapienza Università di Roma

<sup>2</sup>CEMSE Division – King Abdullah University of Science and Technology

April 12, 2018

## Abstract

Quantile regression is a class of methods voted to the modelling of conditional quantiles. In a Bayesian framework quantile regression has typically been carried out exploiting the Asymmetric Laplace Distribution as a working likelihood. Despite the fact that this leads to a proper posterior for the regression coefficients, the resulting posterior variance is however affected by an unidentifiable parameter, hence any inferential procedure beside point estimation is unreliable.

We propose a model-based approach for quantile regression that considers quantiles of the generating distribution directly, and thus allows for a proper uncertainty quantification. We then create a link between quantile regression and Generalized Linear Models by mapping the quantiles to the parameter of the response variable, and we exploit it to fit the model with R-INLA. We extend it also to the case of discrete responses, where there is no 1-to-1 relationship between quantiles and distribution's parameter, by introducing continuous generalizations of the most common discrete variables (Poisson, Binomial and Negative Binomial) to be exploited in the fitting.

## Introduction

Quantile regression is a supervised technique aimed at modeling the quantiles of the conditional distribution of some response variable. With respect to “standard” regression, which is concerned with modeling the conditional mean, quantile regression is especially useful when the tails of the distribution are of interest, as for example when the focus is on extreme behavior rather than average, or when it is important to assess whether or not covariates affect uniformly different levels of the population.

Even though the idea dates back to Galton (1883) (as noted in Gilchrist (2008)), quantile regression was formally introduced only relatively recently by Koenker and Bassett (1978). Since then, the use of quantiles in regression problems has seen an impressive growth and has been thoroughly explored in both the parametric (see Yue and Rue (2011), Wang et al. (2017)), and non-parametric framework (see Yu and Jones (1998), Takeuchi et al. (2005), Li et al. (2007)) with applications ranging from the Random Forest quantile regression of Meinshausen (2006) to D-vine copulas for quantiles in Kraus and Czado (2017), through quantile regression in graphical models as in Ali et al. (2016).

Despite these many different flavors of quantile regression, as noted by Lum and Gelfand (2012), all quantile-based models can be grouped into just two categories: *Conditional Quantile Models*, where the estimation procedure is carried out separately for each quantile of interest and *Joint Quantile Models*, where multiple quantiles of interest are estimated simultaneously. Modelling quantiles jointly requires stronger assumptions on both

covariates and responses, does not allow for linear modeling of the quantiles and it is computationally intensive even when using rough approximations. Its advocates, such as Reich et al. (2011) and Tokdar and Kadaney (2012) stress the fact that joint modeling results in ordered quantile curves, hence it is noticeable immune to quantile crossing, which is a paradoxical phenomenon occurring when the quantile curves are not an increasing function of the quantile level  $\alpha$ . As we believe quantile crossing should be interpreted as a flag that the model is not correct or that data is insufficient rather than an issue to be solved, in the following we will refer exclusively to Conditional Quantile Models.

One of the most significant developments in the quantile regression literature has been the introduction of the Asymmetric Laplace Distribution (ALD) as a working likelihood Yu and Moyeed (2001). From a frequentist point of view, the use of the ALD gave rise to a class of likelihood based method for fitting quantile models and has been instrumental in introducing random effects in linear and non linear quantile regression models; see for example Geraci and Bottai (2007), Geraci and Bottai (2014), Geraci (2017) or Marino and Farcomeni (2015) for a more comprehensive review.

The introduction of the ALD has been even more critical in the Bayesian framework, where the likelihood is a required in inferential procedure Yu and Moyeed (2001). As a result fully bayesian versions of quantile regression, such as Yue and Rue (2011), as well as Bayesian versions of regularized methods, such as Quantile Bayesian Lasso and Quantile Bayesian Elastic Net, have been developed in the last couple of years, exple being Alhamzawi et al. (2011) or Li et al. (2010).

Extensions of the Asymmetric Laplace Distribution such as the Asymmetric Laplace Process Lum and Gelfand (2012), broadened quantile regression to spatially dependent data. Despite their popularity however, ALD based methods are not always satisfactory, especially in terms uncertainty quantification. The use of the ALD introduces an identifiable parameter in the posterior variance, hence any inference beside point estimation is precluded.

In this work we propose a new approach to quantile regression, based on the direct modelling of the quantiles of the generating distribution. Our proposal allows to extend the GLMM framework to quantile modeling by reformulating quantile regression in terms of link functions. This formulation not only recast quantile regression in a much more cohesive setting and overcomes the natural fragmentation that derives from the vastness of the quantile regression literature, but it is also key to an efficient and ready to use fitting procedute, as the connection allows to estimate the model using R-INLA Rue et al. (2009) Rue et al. (2017).

We focus on the delicate case of discrete responses, where the quantiles cannot be direct modeled. The literature on quantiles for counts (or discrete data in general) is largely based on the works of Machado and Santos Silva (2005), which proposes a jittering procedure to make discrete observations continuous by adding noise. In the Bayesian framework count responses have also been modeled through the ALD. For example Lee and Neocleous (2010) couples the jittering with the ALD; alternatively, Congdon (2017) proposes a two-stage regression that uses ALD based quantile regression to model the continuous rate parameter of a Poisson distribution.

This work is structured as follows. In Section 1, after giving a brief refresher of what quantile regression is and what are the issues with the ALD, we introduce our model based approach for quantile regression. In Section 2 we analyze the case where the response variable is discrete, by introducing a continuous version of the most popular distributions for counts. Finally Section 3 shows an application of our proposed method in disease mapping, using the Scottish Lip Cancer data.

	Loss Function $L(Y, r(X))$	Regression Function $r^*(X)$
Quadratic Loss	$(Y_i - r(Y_i))^2$	$\mathbb{E}[Y_i X_i]$
0 – 1 Loss	$\mathbb{1}\{Y_i \neq r(X_i)\}$	$\text{Mode}(Y_i X_i)$
Absolute Loss	$ Y_i - r(X_i) $	$\text{Median}(Y_i X_i)$
Check Loss	$(Y_i - r(X_i))(\alpha - \mathbb{1}\{Y_i - r(X_i) < 0\})$	$\mathbb{Q}_\alpha(Y_i X_i)$

Table 1: Most common loss functions and corresponding regression functions.

## 1 Quantile Regression

The goal of regression methods is to explain a random variable  $Y_i$  as a function of covariates and/or random effects  $X_i$ ; in formulas

$$Y_i = r(X_i) + \varepsilon_i \quad (1)$$

where  $\varepsilon_i$  is an error term which takes into account the randomness of the  $Y_i$ , while  $r(\cdot)$ , the *regression function*, represents the deterministic part of the relation between the response and the covariates. The regression function  $r(X_i)$  is a summary of the conditional distribution of  $Y_i|X_i$ , chosen to minimize the expected loss (or risk) occurring when we neglect the error term to explain  $Y_i$ , or, in other words,  $r(X_i)$  must be such that the deterministic term  $r(X_i)$  is, on average, “close” to the random variable  $Y_i$ . If the loss is taken to be the quadratic loss, i.e.

$$L(Y_i, r(X_i)) = (Y_i - r(X_i))^2 \quad (2)$$

for example, then the regression function minimizing the risk is the conditional mean  $\mathbb{E}[Y_i|X_i]$ . A different loss function results in a different interpretation of the deterministic term of the regression, as shown in Table 1.

The choice of the check (or pinball) loss  $\rho_\alpha(x) = x(\alpha - \mathbb{1}\{x < 0\})$ , a tilted version of the absolute value, results in the regression function being the conditional quantiles.

$$\begin{aligned} r^*(X_i) &= \arg \min_{r(X_i)} \mathbb{E}[L(Y_i, r(X_i))] \\ &= \arg \min_{r(X_i)} \mathbb{E}[\rho_\alpha(Y_i - r(X_i))] \\ &= \mathbb{Q}_\alpha(Y_i|X_i). \end{aligned}$$

No loss function is uniformly better than the others, but each has different strengths. The advantage of the check loss over the quadratic loss (hence of quantile regression over mean regression), for example, is that it gives a more complete picture of the distribution of  $Y_i|X_i$ , it is more robust with respect to outliers, it allows for dealing with censored data without additional assumptions, and most importantly it allows to model extreme behavior.

As in mean regression, quantile regression models can be parametric, semi-parametric or non parametric altogether. In the first and most basic formulation of Koenker and Bassett (1978), quantile linear regression, the quantile of level  $\alpha$  of the conditional distribution  $Y_i|X_i$ , can be modeled as:

$$\mathbb{Q}_\alpha(Y_i|X_i) = X_i^t \beta_\alpha \quad (3)$$

where the notation  $\beta_\alpha$  highlights the dependence of the regression coefficients to the quantile level. Given a sample  $\{(Y_i, X_i)\}_{i=1}^n$ , estimate for the regression coefficient  $\beta_\alpha$  can be found by minimizing the empirical risk:

$$\begin{aligned} \hat{\beta}_\alpha &= \arg \min_{\beta_\alpha} \widehat{\mathbb{E}}[\rho_\alpha(Y_i - X_i^t \beta_\alpha)] \\ &= \arg \min_{\beta_\alpha} \sum_{i=1}^n \rho_\alpha(Y_i - X_i^t \beta_\alpha). \end{aligned} \quad (4)$$

This is a standard linear programming (LP) problem and can be trivially solved by means of *simplex method* or *interior point methods*.

## 1.1 Asymmetric Laplace Distribution

Quantile regression as defined by the optimum problem in Equation (4) does not require any distributional assumption for the response variable, and is thus an intrinsically non parametric (in the sense of model-free) method. When adopting a bayesian approach however, this lack of generating model assumption may be a problem, as it implies that there is no likelihood. Several different approaches have been proposed to tackle this issue: for example Gelfand and Kottas (2002) proposes a Dirichlet mixture models, Yang and He (2012) considers empirical likelihood while Reich and Smith (2013) advocates a semiparametric model for the whole quantile process. Despite these more recent developments, the one method dominating the literature is still using the Asymmetric Laplace Distribution (ALD) as a working likelihood, as suggested by Yu and Moyeed (2001).

A random variable  $X$  has an Asymmetric Laplace Distribution with parameters  $\alpha$ , scale  $\sigma$  and center  $\mu$ , i.e.  $X \sim \text{ALD}(\alpha, \sigma, \mu)$ , if it has density

$$f_X(x) = \frac{\alpha(1-\alpha)}{\sigma} \exp \left\{ -\frac{\rho_\alpha(x-\mu)}{\sigma} \right\}. \quad (5)$$

By assuming that the conditional distribution for the response variable  $Y_i|X_i$  is an ALD with parameters  $\alpha$  taken to be the quantile level we are interested in, and  $\mu_i = X_i^t \beta$  (or a more complicated function of  $X_i^t \beta$  if we want to move beyond linear quantile regression) the resulting likelihood is:

$$L(\beta; \mathcal{D}_n) \propto \exp \left\{ -\frac{\sum \rho_\alpha(Y_i - X_i^t \beta)}{\sigma} \right\}. \quad (6)$$

It is immediate to see that the Maximum Likelihood Estimator (MLE) corresponds to the quantile regression estimator in Equation (4), hence the ALD translates quantile regression in the bayesian framework, without making true distributional assumption on the response variable; it is thus not the generating model but a working model (or likelihood). Although its non differentiability restrict the class of algorithms to be exploited in the fitting, a drawback that may be overcome by smoothing the check loss in the exponent (see Yue and Rue (2011) or Fasiolo et al. (2017)), nevertheless the popularity of the ALD in the Bayesian framework stems mostly from its computational attractiveness. A random variable  $Y \sim \text{ALD}(\alpha, \sigma, \mu)$  admits in fact the following decomposition:

$$Y = \sigma(\theta_1 V + \theta_2 Z \sqrt{V}) \quad (7)$$

where  $\theta_1 = (1 - 2\sigma)/(\sigma - \sigma^2)$ ,  $\theta_2^2 = 2/(\sigma - \sigma^2)$ ,  $Z \sim N(0, 1)$  and  $V \sim \text{Exp}(1)$ , with  $Z$  and  $V$  independent. This can be used to recast quantile regression of Equation (3) as the following hierarchical model

$$\begin{aligned} Y_i | X_i, V_i &\sim N(X_i^t \beta + \theta_1 \sigma V_i, \theta_2^2 \sigma^2 V_i) \\ \sigma V_i &\sim \text{Exp}(\sigma) \end{aligned}$$

which allows for an easy implementation of most MCMC algorithms, see Kozumi and Kobayashi (2011). As shown in Yu and Moyeed (2001), most common choices of prior distributions on the regression coefficients  $\beta$ , including the bayesian Lasso and Elastic net, yield proper posteriors, making quantile regression ever so popular among bayesians; see Alhamzawi and Yu (2014) or Li et al. (2007) to name a few examples.

## 1.2 Model Based Quantile Regression

The major drawback of adopting any working likelihood rather than a generating distribution is that the validity of posterior inference is not automatically guaranteed by Bayes Theorem. This is especially true when the working likelihood is chosen to be the Asymmetric Laplace Distribution; in this case, as pointed out by Yang et al. (2016), the scale parameter  $\sigma$  of the ALD affects the posterior variance, despite not having any impact on the quantile itself. A random variable  $X \sim ALD(\alpha, \sigma, \mu)$ , in fact, is such that  $\mathbb{P}(X \leq \mu) = \alpha$  does not depend on  $\sigma$  Yue and Rue (2011). Bayesian inference based on the ALD could be corrected by using an adjusted variance estimator, as shown in Yang et al. (2016), however the importance of such a result is mostly theoretical. The definition of the adjusted variance in fact involves once again the unknown parameter  $\sigma$  and it is only asymptotically valid. Our alternative is to reject altogether the use of a working likelihood in favor of the true generating model. We propose a *model-based quantile regression*, which exploits the shape of the conditional distribution to link the covariates of interest to the distribution parameter.

This approach is general enough to be exploited in frequentist analyses as well, however it is especially appealing in the bayesian framework, since it grants the identifiability of all parameters of the posterior, and thus allows for meaningful and exact posterior inference, which, at this time, is something that is still missing in the literature.

Given that  $Y_i|X_i$  has distribution  $F(y_i; \theta_i)$ , where  $\theta_i$  is canonical parameter of the distribution, our procedure can be formalized in two steps.

- **Modeling step:** the quantile of  $Y_i|X_i$ ,  $q_i^\alpha = \mathbb{Q}_\alpha(Y_i|X_i)$  is modeled as

$$q_i^\alpha = g(\eta_i^\alpha) \quad (8)$$

where  $g$  is an invertible function chosen by the modeler and  $\eta_i^\alpha$  is the linear predictor, which depends on the level  $\alpha$  of the quantile. No restriction is placed on the linear predictor, which can include fixed as well as random effect. Our approach is thus flexible enough to include parametric or semi parametric models, where the interest may lay in assessing the difference in the impact that the covariate may have at different levels of the distribution, as well as fully non parametric models, where the focus is shifted towards prediction instead.

- **Mapping step:** the quantile  $q_i^\alpha$  is univocally mapped to the parameter  $\theta_i$  as

$$\theta_i = h(q_i^\alpha, \alpha) \quad (9)$$

where the function  $h$  must be invertible to ensure the identifiability of the model and explicitly depends on the quantile level  $\alpha$ . The map  $h$  gives us a first interpretation of model-based quantile regression as a reparametrization of the generating likelihood function  $F(y_i; \theta_i)$ , in terms of its quantiles, i.e.  $F(y_i; q_i^\alpha = h^{-1}(\theta_i, \alpha))$ .

By linking the quantiles of the generating distribution to its canonical parameter  $\theta_i$ , we are indirectly modeling  $\theta_i$  as well, hence we are implicitly building a connection between quantile regression and Generalized Linear (Mixed) Models (GLMM), which are also concerned with the modeling of  $\theta_i$ . The modeling and mapping steps in fact, can be considered as a way to define a link function, in the GLMM sense, as the composition  $\theta_i = h(g(\eta_i))$ , and this allows us to rephrase quantile regression as a new link function in a standard GLMM problem. Drawing a path from GLMM to quantile regression is instrumental in the fitting however the pairing is only formal: coefficients and random effect have a completely different interpretation.

From a computational standpoint, the main advantage of coupling quantile regression to GLMM is that this new class of models can be implemented in R-INLA Rue et al.

(2017), which allows for both flexibility in the model definition and efficiency in their fitting. R-INLA is an R package that implements the INLA (Integrated Nested Laplace Approximation, see Rue et al. (2009)) method for approximating marginal posterior distributions for hierarchical bayesian models, where the latent structure is assumed to be a Gaussian Markov Random Field, Rue and Held (2005), which is especially convenient for computations due to its naturally sparse representation. The class of model that can be formulated in a hierarchical fashion is broad enough to include parametric and semi parametric models; the INLA approach is thus extremely flexible and provides a unified fitting framework for mixed and additive models in all their derivations.

## 2 Discrete Data

Quantile regression was originally defined for continuous responses and extending it to the case of discrete variable has proven to be challenging. In the model-free case, inference is limited by the fact that the non-differentiable objective function in Equation (4) together with the points of positive mass of one of the variables involved in the optimization problem, makes it impossible to derive an asymptotic distribution for the sample quantiles. In the case of model-based quantile regression, dealing with discrete distribution is non trivial since it is difficult to define both the model  $g$  and the map  $h$  as defined in Section 1.2. As far as the modeling step is concerned, most common model choices, such as the log model for count data or logit for binary responses are typically continuous and are not well suited to represent the conditional quantiles, which are intrinsically discrete. More tragically, as the quantile function is discrete, it is not possible to define an injective map  $h$ , which means that it is not possible to define a unique  $\theta$  generating each quantile, as can be seen in Figure 1.

The core idea for dealing with discrete data is to approximate them with continuous distributions that retain some of the structure of the discrete random variable. To date, the most famous strategy to do so is jittering, which consists in adding uniform noise  $U_i \sim \text{Uniform}(0, 1)$  to the response variable  $Y_i$  and then model quantiles of the “continuous”  $Z_i = Y_i + U_i$ , as shown in Machado and Santos Silva (2005).

Our strategy for modeling is also to assume that the discrete responses observed are generated by a continuous distribution and rather than modeling the quantiles for the discrete distribution we model directly the quantiles of its continuous counterpart. However, what distinguish our proposal is that the continuous distributions we model are defined by interpolating the cumulative distribution function (c.d.f.) of the original discrete variable, so that they can be considered as generalization of the original distribution. Inspired by Iliencko (2013), we focus in particular on discrete distributions whose c.d.f. can be written as

$$F_X(x; \theta) = \mathbb{P}(X \leq x) = k(\lfloor x \rfloor, \theta) \quad (10)$$

where  $k$  is a continuous function in the first argument. The continuous interpolation is then defined by removing the floor operator, so that the function  $k(x, \theta)$  is the c.d.f. of  $X'$ , a continuous version of  $X$ . By definition of floor we have that

$$F_X(x) = k(\lfloor x \rfloor, \theta) = k(x, \theta) = F_{X'}(x) \quad (11)$$

for all integer  $x$ , and since the two c.d.f.s are the same at the integer values of  $x$ , this can be seen as a continuous generalization of the original variable.

Modeling quantiles of the continuous counterparts of discrete distribution is necessary in the Bayesian framework, where the distribution is needed as a likelihood, but it has a clear advantage also in the frequentist setting, since it allows for a better assessment of variability. Confidence intervals for the regression coefficient in fact heavily rely on the

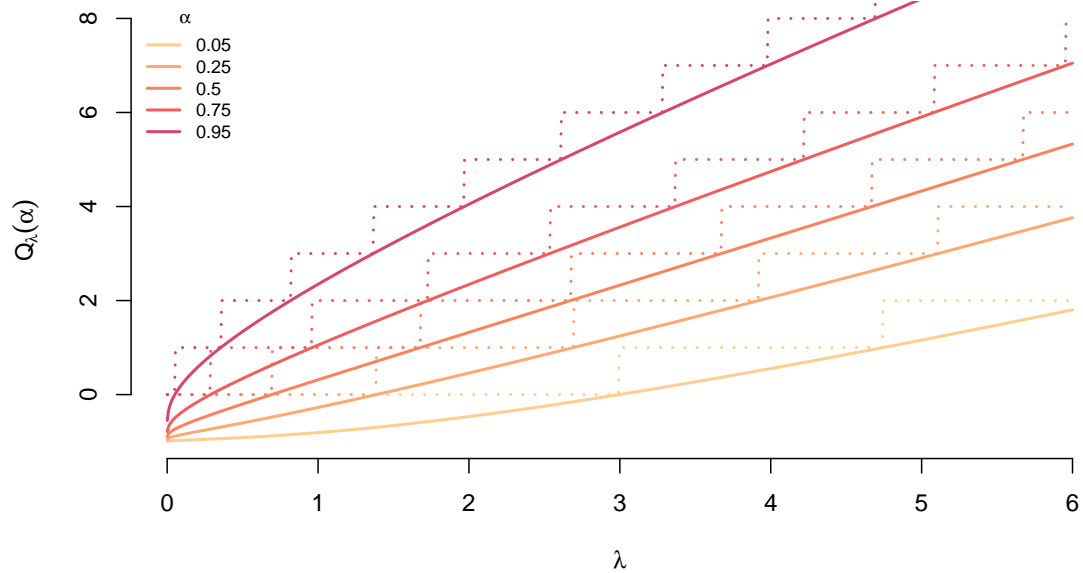
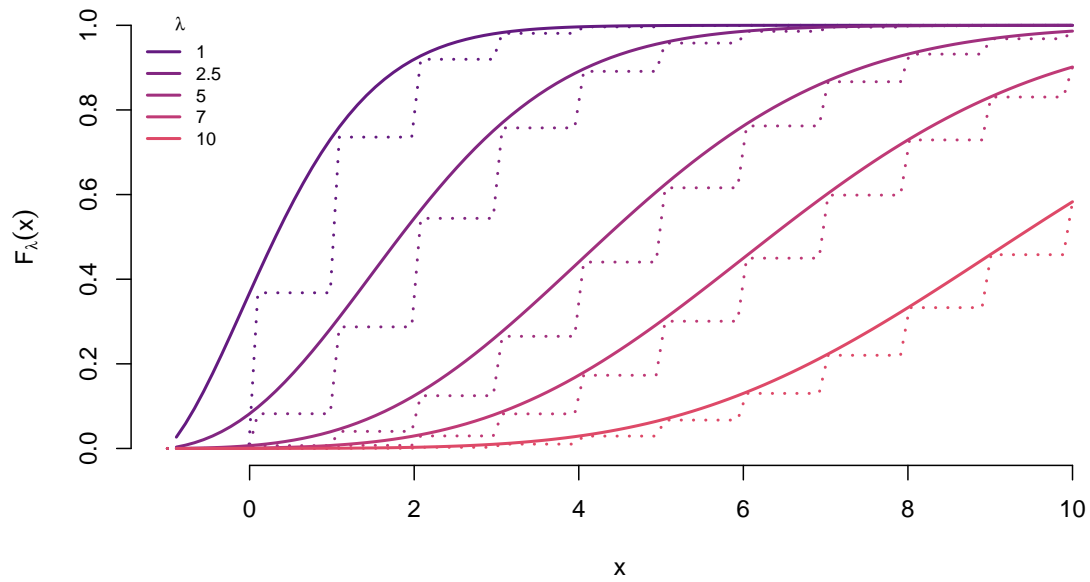


Figure 1: Top: c.d.f. of the discrete (dashed line) and continuous (continuous line) Poisson distribution for several values of  $\lambda$ . Bottom: quantile function of the discrete (dashed line) and continuous (continuous line) Poisson distribution.

asymptotic normality of the sample quantiles, which is guaranteed only when the distribution generating the data is continuous, while sample quantiles of discrete distribution in general are not asymptotically normal. Even in the jittering approach, where the data are made “continuous”, it is cumbersome to determine the variance associated to the estimates for the coefficients; the asymptotic normality of the sample quantiles is granted in fact as the sample size as well as the number of repetition of the jittering procedure go to infinity, and thus variance estimation require computationally intensive re-sampling procedures, moreover there is no clear relation between the quantiles of  $Z_i = Y_i + U_i$ , and the quantiles of the  $Y_i$  directly.

## 2.1 Continuous Poisson

The class of distribution defined by Equations (10) and (11) is broad enough to include the three distribution most frequently encountered in applications: Poisson, Binomial and Negative Binomial. We explore in detail the Poisson case, the other two are trivial extensions.

The starting point for defining a continuous version of the Poisson distribution is to rewrite the cumulative density function for the Poisson as the ratio of Incomplete and Regular Gamma function:

$$X \sim \text{Poisson}(\lambda) \quad F_X(x) = \mathbb{P}(X \leq x) = \frac{\Gamma(\lfloor x \rfloor + 1, \lambda)}{\Gamma(\lfloor x \rfloor + 1)} \quad x \geq 0 \quad (12)$$

where

$$\Gamma(x, \lambda) = \int_{\lambda}^{\infty} e^{-s} s^{x-1} ds$$

is the upper incomplete Gamma function. Extending the Poisson distribution to the continuous case from this formulation is just a matter of removing the floor operator, that is

$$X' \sim \text{Continuous Poisson}(\lambda) \quad F_{X'}(x) = \mathbb{P}(X' \leq x) = \frac{\Gamma(x + 1, \lambda)}{\Gamma(x + 1)} \quad x > -1. \quad (13)$$

where the domain has been extended from  $x \geq 0$  to  $x > -1$  in order to avoid mass at 0. The Continuous Poisson defined in Equation (13) is similar to that of Ilienکو (2013), with the noticeable difference that our definition of Continuous Poisson is shifted by 1, so that the Discrete Poisson  $X$  is a monotonic left continuous function of the Continuous Poisson  $X'$ ; more specifically Continuous and Discrete versions of the Poisson are related by

$$X = \lceil X' \rceil. \quad (14)$$

By integration by parts we have

$$\frac{\Gamma(x + 1, \lambda)}{\Gamma(x + 1)} - \frac{\Gamma(x, \lambda)}{\Gamma(x)} = \lambda^x e^{\lambda} / \Gamma(x + 1) \quad (15)$$

which is enough to show that:

$$\begin{aligned} \mathbb{P}(\lceil X' \rceil = x) &= \mathbb{P}(X' \in (x - 1, x]) = F_{X'}(x) - F_{X'}(x - 1) \\ &= \frac{\Gamma(x + 1, \lambda)}{\Gamma(x + 1)} - \frac{\Gamma(x, \lambda)}{\Gamma(x)} = \lambda^x e^{\lambda} / \Gamma(x + 1) \\ &= \mathbb{P}(X = x). \end{aligned} \quad (16)$$

$$(17)$$

Following Ilienکو (2013), we have that  $F_{X'}(x)$  is a well defined c.d.f., in the sense that it is non-decreasing in  $x$ , is right-continuous and it satisfies:

$$\lim_{x \rightarrow \infty} F_{X'}(x) = 1 \quad \text{and} \quad \lim_{x \rightarrow -\infty} F_{X'}(x) = 0.$$



By assuming that discrete responses are generated by a Continuous Poisson discrete, it is possible to extend quantile regression to count data. If  $Y_i|\eta_i \sim \text{Continuous Poisson}(\lambda_i)$ , in fact, we can specify the link function  $g$  and the parameter map  $h$ . More specifically we have:

$$g : \quad q_i^\alpha = g(\eta_i^\alpha) = \exp\{\eta_i^\alpha\} \quad (18)$$

$$h : \quad \lambda_i = h(q_i^\alpha) = \frac{\Gamma^{-1}(q_i^\alpha + 1, 1 - \alpha)}{\Gamma(q_i^\alpha + 1)}. \quad (19)$$

When units  $i$  are subjected to different exposures  $E_i$ , there are two ways of encoding it into the model:

- by including them in the model as offset, discounting the quantiles directly and considering  $q_i^\alpha/E_i$

$$\begin{aligned} q_i^\alpha &= \exp\{\eta_i^\alpha + \log(E_i)\} = E_i \exp\{\eta_i^\alpha\} \\ \lambda_i &= \frac{\Gamma^{-1}(q_i^\alpha + 1, 1 - \alpha)}{\Gamma(q_i^\alpha + 1)} \end{aligned} \quad (20)$$

- by adjusting the global parameter and consider  $\lambda_i/E_i$

$$\begin{aligned} q_i^\alpha &= \exp\{\eta_i^\alpha\} \\ \lambda_i &= E_i \frac{\Gamma^{-1}(q_i^\alpha + 1, 1 - \alpha)}{\Gamma(q_i^\alpha + 1)}. \end{aligned} \quad (21)$$

While in Poisson mean regression these two approaches yield the same results, as

$$\lambda_i = E_i \exp\{\eta_i\} \iff \lambda_i/E_i = \exp \eta_i \quad (22)$$

in Poisson quantile regression this is not true. In general

$$\frac{\Gamma^{-1}(E_i \exp\{\eta_i^\alpha\}^\alpha + 1, 1 - \alpha)}{\Gamma(E_i \exp\{\eta_i^\alpha\}^\alpha + 1)} \neq E_i \frac{\Gamma^{-1}(\exp\{\eta_i^\alpha\}^\alpha + 1, 1 - \alpha)}{\Gamma(\exp\{\eta_i^\alpha\}^\alpha + 1)} \quad (23)$$

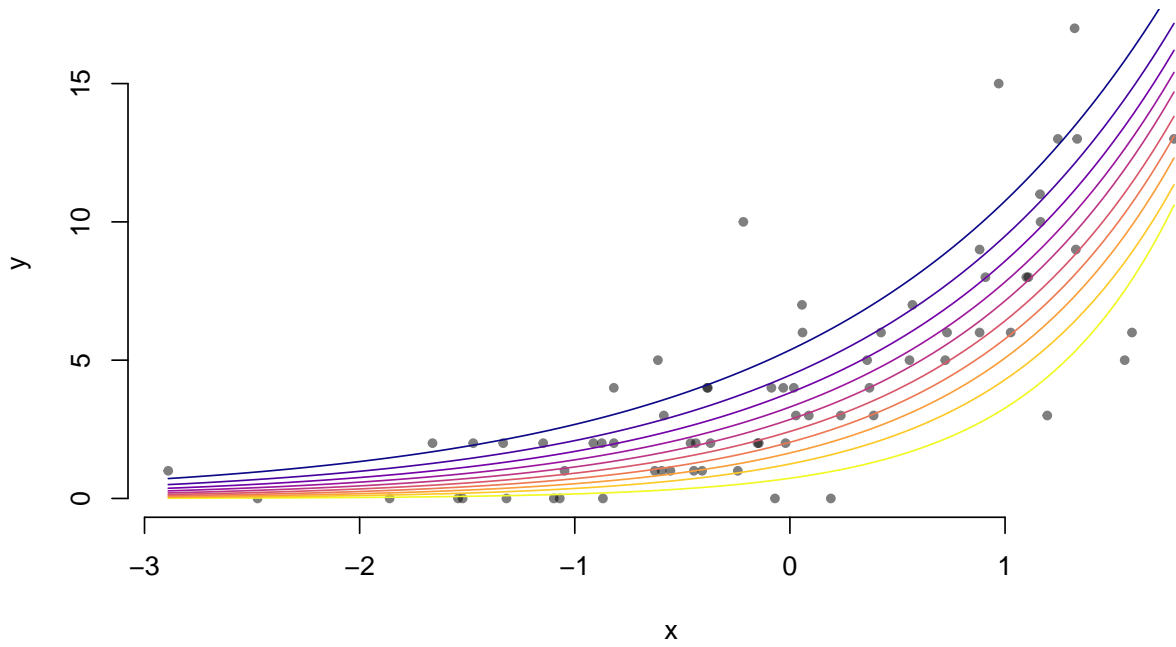
and, besides the trivial case  $E_i = 1$ , it is not obvious to determine whether there are values of  $E_i$  for which the equality would hold since there is no closed form solution for  $\Gamma^{-1}$ . A case could be made for both modeling strategies, the former being a “quantile-specific” model while the latter being more of a global model, and choosing between them depends on the application.

As can be seen in Figure 1, the quantiles of the two distributions are not the same, and the regression model returns fitted quantiles for the Continuous Poisson. However, fitted quantiles of the discrete distribution can be obtained by exploiting quantile equivariance, since we defined the continuous Poisson so that its discrete counterpart is obtained through is a monotonic left continuous function. Let  $Y_i|\eta_i \sim \text{Poisson}(\lambda_i)$  and  $Y_i'|\eta_i \sim \text{Continuous Poisson}(\lambda_i)$ , then we have

$$\mathbb{Q}_\alpha(Y_i'|\eta_i) = \mathbb{Q}_\alpha(\lceil Y_i \rceil|\eta_i) = \lceil \mathbb{Q}_\alpha(Y_i|\eta_i) \rceil. \quad (24)$$

Figure 2 shows a simulated toy example, with  $n = 70$  observations generated from a Poisson  $\lambda_i = \exp(1 + X_i)$ , where the covariates  $X_i$  are simulated from a standard Normal. With respect to Machado and Santos Silva (2005), our method seems to be less affected by the phenomenon of quantile crossing, which compromises the interpretation of the quantile estimates with the jittering approach even in such a trivial example.

### Model Based Quantile Regression



### Jittering Based Quantile Regression

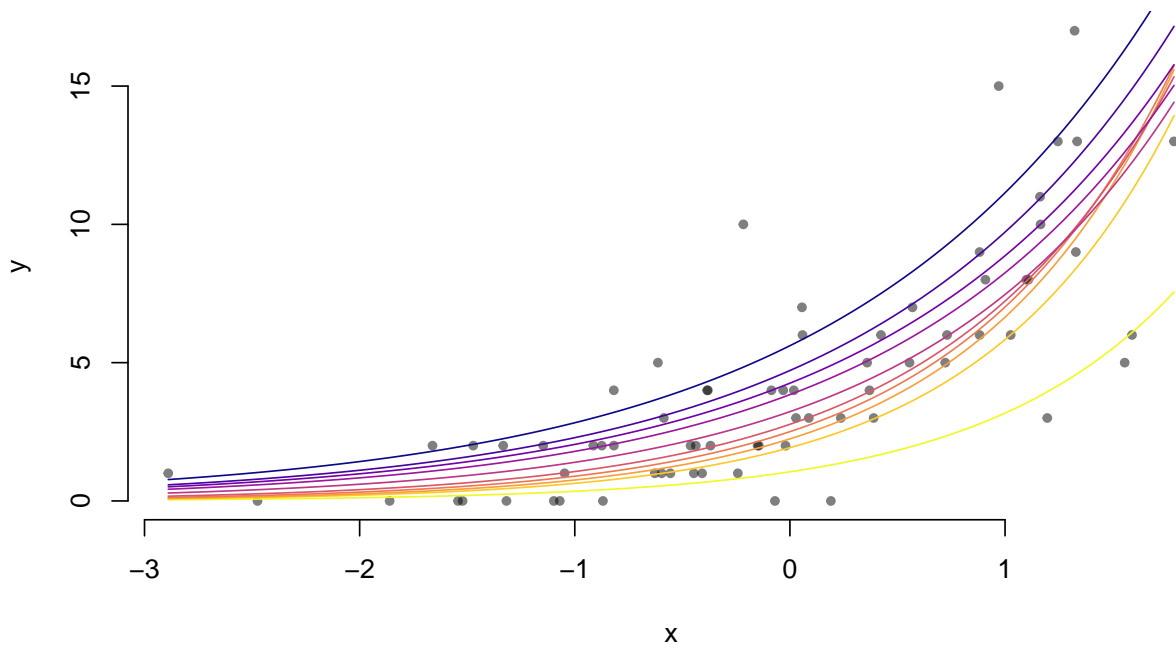


Figure 2: Quantile curves estimated with model based quantile regression (top) and jittering (bottom).

## 2.2 Continuous Count distributions

The Binomial and the Negative Binomial distribution can also be trivially extended to the continuous case. Their c.d.f. can in fact be written as:

$$Y \sim \text{Binomial}(n, p) \quad F_Y(y) = I_{1-p}(n - \lfloor y \rfloor, \lfloor y \rfloor + 1) \quad (25)$$

$$Z \sim \text{Negative Binomial}(r, p) \quad F_Z(z) = I_{1-p}(r, \lfloor z \rfloor + 1) \quad (26)$$

where  $I_x(a, b)$  is the *regularized incomplete Beta function* defined as:

$$I_x(a, b) = \frac{B(a, b, x)}{B(a, b)} \quad \text{with} \quad B(a, b, x) = \int_x^1 t^a (1-t)^{b-1} dt \quad (27)$$

Again the extension of these two random variables to the continuous case can be obtained by removing the floor operator:

$$Y' \sim \text{Continuous Binomial}(n, p) \quad F_{Y'}(y) = I_{1-p}(n - y, y + 1) \quad (28)$$

$$Z' \sim \text{Continuous Negative Binomial}(r, p) \quad F_{Z'}(z) = I_{1-p}(r, z + 1) \quad (29)$$

It is obvious that these two continuous distributions, exactly like the continuous Poisson defined before, result in interpolation of both the c.d.f. and quantile functions of the discrete original, analogously to what can be seen in Figure 1.

The advantage of this method of interpolating the c.d.f.s of discrete random variables is that the behavior of the resulting continuous random variables mimic that of their discrete counterparts.

In the discrete case in fact it is well known that the Poisson distribution is the limiting case of both the Binomial and the Negative Binomial when the probability of observing one event goes to 0 and that Binomial and Negative Binomial are also entwined in a 1-to-1 relation. The same relations are preserved in the continuous case, hence the two classes of distribution have similar meaning.

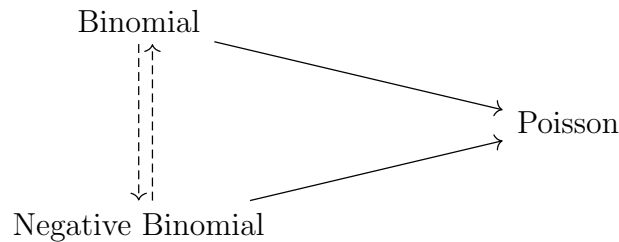


Figure 3: Diagram summarizing the connection between Binomial, Negative Binomial distribution. Continuous lines indicate asymptotic relations, dashed lines denote a finite sample relation.

**Poisson and Binomial** Let  $X$  be a Continuous Poisson with parameter  $\lambda$ ,  $Y$  be a Continuous Binomial with parameters  $n$  and  $p$ . Then by following Iliencko (2013) we have that for  $n \rightarrow \infty$  and  $p \rightarrow 0$  so that  $np \rightarrow \lambda$

$$F_Y(x) = \frac{B(x+1, N-x, p)}{B(x+1, N-x)} \rightarrow \frac{\Gamma(x+1, \lambda)}{\Gamma(x+1)} = F_X(x). \quad (30)$$

Analogously to its discrete version, the Continuous Poisson can thus be interpreted as an approximation for a Binomial-like distribution in the case of rare events.

**Poisson and Negative Binomial** Let  $X$  be a Continuous Poisson with parameter  $\lambda$ ,  $Z$  be a Continuous Negative Binomial with parameters  $r$  and  $p$ . Then it follows trivially from Equation (30) that for  $r \rightarrow \infty$  and  $p \rightarrow 0$  so that  $rp \rightarrow \lambda$  we have

$$F_Z(x) = \frac{B(x+1, r, p)}{B(x+1, r)} \longrightarrow \frac{\Gamma(x+1, \lambda)}{\Gamma(x+1)} = F_X(x). \quad (31)$$

From a modeling perspective, this motivates the choice of the Continuous Negative Binomial instead of the Continuous Poisson in cases where there is over-dispersion, i.e. the assumption of mean = variance is clearly violated.

**Binomial and Negative Binomial** Let  $Z$  be a Continuous Negative Binomial with parameters  $r$  and  $p$  and  $Y$  be a Continuous Binomial with parameters  $s+r$  and  $1-p$ , then

$$\begin{aligned} F_Z(s) &= 1 - I_p(s+1, r) \\ &= 1 - I_p((s+r) - (r-1), (r-1) + 1) \\ &= 1 - \mathbb{P}(Y \leq r-1) \\ &= \mathbb{P}(Y \geq r) \end{aligned} \quad (32)$$

which justifies the interpretation of the Continuous Negative Binomial as the waiting time until the arrival of the  $r$ -th success in a Binomial-like experiment.

### 3 An application - Disease Mapping

We conclude by showing an application to Lip Cancer data in Scotland, a famous case study in the disease mapping literature, most noticeably in Wakefield (2007) and Cressie (1993), consisting of lip cancer incidence among men over the period 1975–1980, aggregated over 56 counties. One of the goals of disease mapping is to identify which areas correspond to a high risk; standard risk measure, such as the ratio between observed and expected cases in each area, the Standardized Mortality (or Morbidity) Ratio (SMR)  $SMR_i = Y_i/E_i$ , is not reliable here due to the high variability of expected cases  $E_i$  (Figure 4), hence is advisable to introduce a random effect model that exploits the spatial structure to obtain more stable estimates of the risk. Assuming that, conditionally on covariates  $X_i$  and random effects  $b_i$ , the observations are generated by a Poisson distribution

$$Y_i | X_i, b_i \sim \text{Poisson}(\lambda_i) \quad (33)$$

we adopt the following model for the conditional quantile of level  $\alpha$

$$\mathbb{Q}_\alpha(Y_i | X_i, b_i) = E_i \theta_{i,\alpha} = E_i \exp\{\eta_i\}. \quad (34)$$

We opted for the *quantile-level* approach for handling exposures  $E_i$  in order to ease interpretation; as we discount each quantile for the exposures, in fact, the parameter  $\theta_{i,\alpha}$  can be considered the relative risk of unit  $i$  at level  $\alpha$  of the population. The linear predictor  $\eta_i$  can be decomposed into

$$\eta_i = \beta_0 + \beta_1 X_i + b_i \quad (35)$$

where  $\beta_0$  represent the overall risk,  $X_i$  is the percentage of people engaged in outside activity divided by 10, as common in the literature Wakefield (2007), and  $b_i$  consists in the sum of an unstructured random effect capturing overdispersion and measurement errors and spatially structured random effect. In order to avoid the confounding between the two components of the random effect and to avoid scaling issues we adopt for  $b$  the modified version of the Besag–York–Mollier (BYM) model introduced in Simpson et al. (2017):

$$b_i = \frac{1}{\tau_b} \left( \sqrt{1 - \phi} v_i + \sqrt{\phi} u_i \right). \quad (36)$$

Both random effects are normally distributed, and in particular

$$v \sim N(0, \mathbf{I}) \quad (37)$$

$$u \sim N(0, Q_u^{-1}) \quad (38)$$

so that  $b \sim N(0, Q_b^{-1})$  with  $Q_b^{-1} = \tau_b^{-1}(1 - \phi)\mathbf{I} + \phi Q_u^{-1}$ , a weighted sum of the precision matrix for the  $\mathbf{I}$  and the precision matrix representing the spatial structure  $Q_u$ , scaled in the sense of Sørbye and Rue (2014).

We assign priors on the precision  $\tau_b$  and the mixing parameter  $\phi$  using the penalized complexity (PC) approach, as defined in Simpson et al. (2017) and detailed in Riebler et al. (2016) in the special case of disease mapping. Estimated coefficients shown in Table 2 show that the impact of the covariate is stronger in the upper tail of the distribution, as we could expect.

Despite regression being a key tool for disease mapping, the use of quantile regression

	Mean	1st Quartile	2nd Quartile	3rd Quartile
$\beta_0$	-0.1179 (0.1137)	-0.4814 (0.1382)	-0.2098 (0.1220)	0.0228 (0.1113)
$\beta_1$	0.2619 (0.1138)	0.2506 (0.1382)	0.2588 (0.1220)	0.2632 (0.1113)
$\tau_b$	4.3799 (1.3815)	3.4609 (1.1950)	4.0826 (1.3374)	4.5075 (1.4091)
$\phi$	0.9279 (0.0815)	0.9093 (0.0971)	0.9226 (0.0859)	0.9278 (0.0815)

Table 2: Posterior mean estimates of model parameters (and corresponding standard deviations).

instead of mean regression is still unexplored, with exceptions in Congdon (2017) and Chambers et al. (2014). This is somehow surprising, since the focus of disease mapping is on extreme behaviors of the population, for which using quantiles, that provide insights on the tails of the distributions, would seem a more natural choice than considering means. The relative risk  $\theta_{i,\alpha}$  can be directly used to detect “high risk” areas. Following Congdon (2017), the  $i^{\text{th}}$  area region is considered at “high risk” if  $[\theta_{i,0.05}, \theta_{i,0.95}] > 1$ , where 1 represents an increase in the risk, otherwise it is assumed to be “low risk”. Mean regression methods for identifying “high risk” areas are also based on relative risk  $\theta_i$ , although defined in a different way, i.e.

$$\mathbb{E}[Y_i | X_i, b_i] = E_i \theta_i = E_i \exp\{\eta_i\}. \quad (39)$$

Posterior probability of an increase in the risk are then used to assess whether an area has high risk or not, so that the  $i^{\text{th}}$  area is considered to be of high risk if  $\mathbb{P}(\theta_i > 1 | Y_1, \dots, Y_n) > t$  where  $t$  is a threshold value depending on the application (in this case we chose  $t = 0.9$ ).

The difference between the two methods is that in the former high risk areas are those where the risk increases for every level of the population, i.e. for those who are very sensible and those who are less sensible to the disease, while the latter considers only the mean level, which is a synthetic measure for the whole population but it may be subject to compensation. Figure 3 shows the critical areas identified by quantile and mean regression. The similarity of the results of our method with those corresponding to a more traditional approach, as well as to previous analyses, most noticeably Wakefield (2007), reassures us that our method yields reasonable results. At the same time, the minor discrepancies between the two maps is also encouraging, as the two methods have different definitions of high risk; different results correspond in fact to different insights on the disease risk and the non-overlap between quantile-based exceedance probability-based methods testifies that there is information to be gained from our approach.

## References

- R. Alhamzawi and K. Yu. Bayesian Lasso-mixed quantile regression. *Journal of Statistical Computation and Simulation*, 84(4):868–880, 2014.
- R. Alhamzawi, U. Ub, K. Yu, and D. F. Benoit. Bayesian adaptive Lasso quantile regression Bayesian adaptive Lasso quantile regression. *Statistical Modelling*, (July), 2011.
- A. Ali, J. Z. Kolter, and R. J. Tibshirani. The multiple quantile graphical model. In *Advances in Neural Information Processing Systems*, pages 3747–3755, 2016.
- R. Chambers, E. Dreassi, and N. Salvati. Disease Mapping via Negative Binomial Regression M-quantiles. *Statistics in Medicine*, 33(27):4805–4824, 2014.
- P. Congdon. Quantile Regression for Area Disease Counts : Bayesian Estimation using Generalized Poisson Regression. 44(0):92–103, 2017.
- N. Cressie. *Statistics for Spatial Data*. Wiley Online Library, 1993.
- M. Fasiolo, Y. Goude, R. Nedellec, and S. N. Wood. Fast calibrated additive quantile regression. *arXiv preprint arXiv:1707.03307*, 2017.
- F. Galton. *Inquiries into human faculty and its development*. JM Dent and Company, 1883.
- A. E. Gelfand and A. Kottas. A Computational Approach for Full Nonparametric Bayesian Inference Under Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics*, 11(2):289–305, 2002.
- M. Geraci. Nonlinear quantile mixed models. *arXiv preprint arXiv:1712.09981*, 2017.
- M. Geraci and M. Bottai. Quantile regression for longitudinal data using the asymmetric Laplace distribution. *Biostatistics*, 8(1):140–154, 2007.
- M. Geraci and M. Bottai. Linear quantile mixed models. *Statistics and Computing*, 24(3):461–479, 2014.
- W. Gilchrist. Regression revisited. *International Statistical Review*, 76(3):401–418, 2008.
- A. Iliencko. Continuous counterparts of Poisson and binomial distributions and their properties. *arXiv preprint arXiv:1303.5990*, 2013.
- R. Koenker and G. Bassett. Regression Quantiles. *Econometrica*, 46(1):33–50, 1978.
- H. Kozumi and G. Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11):1565–1578, 2011.
- D. Kraus and C. Czado. D-vine copula based quantile regression. *Computational Statistics and Data Analysis*, 110:1–18, 2017.
- D. Lee and T. Neocleous. Bayesian quantile regression for count data with application to environmental epidemiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59:905–920, 2010.
- Q. Li, R. Xiy, and N. Lin. Bayesian regularized quantile regression. *Bayesian Analysis*, 5(3):533–556, 2010.
- Y. Li, Y. Liu, and J. Zhu. Quantile regression in reproducing kernel Hilbert spaces. *Journal of the American Statistical Association*, 102(477):255–268, 2007.

- K. Lum and A. E. Gelfand. Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Analysis*, 7(2):235–258, 2012.
- J. A. Machado and J. M. Santos Silva. Quantiles for counts. *Journal of the American Statistical Association*, 100(472):1226–1237, 2005.
- M. F. Marino and A. Farcomeni. Linear quantile regression models for longitudinal experiments: An overview. 73(2):229–247, 2015.
- N. Meinshausen. Quantile Regression Forests. *Journal of Machine Learning Research*, 7: 983–999, 2006.
- B. J. Reich and L. B. Smith. Bayesian quantile regression for censored data. *Biometrics*, 69(3):651–660, 2013.
- B. J. Reich, M. Fuentes, and D. B. Dunson. *Journal of the American Statistical Association*, 106(493):6–20, 2011.
- A. Riebler, S. H. Sørbye, D. Simpson, and H. Rue. An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25(4):1145–116, 2016.
- H. Rue and L. Held. *Gaussian Markov Random Fields Theory and Applications*. CRC press, 2005.
- H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 71(2):319–392, 2009.
- H. Rue, A. Riebler, S. H. Sørbye, J. B. Illian, D. P. Simpson, and F. K. Lindgren. Bayesian Computing with INLA : A Review. *Annual Review of Statistics and Its Application*, 4: 395–421, 2017.
- D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28, 2017.
- S. H. Sørbye and H. Rue. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, 8:39–51, 2014.
- I. Takeuchi, Q. V. Le, T. Sears, and A. J. Smola. Nonparametric Quantile Regression. *Journal of Machine Learning Research*, 7(May), 2005.
- S. T. Tokdar and J. B. Kadaney. Simultaneous linear quantile regression: A semiparametric bayesian approach. *Bayesian Analysis*, 7(1):51–72, 2012.
- J. Wakefield. Disease mapping and spatial regression with count data. *Biometrics*, 8(2): 158–183, 2007.
- H. J. Wang, I. W. McKeague, and M. Qian. Testing for marginal linear effects in quantile regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(2):433–452, 2017.
- Y. Yang and X. He. Bayesian empirical likelihood for quantile regression. *Annals of Statistics*, 40(2):1102–1131, 2012.
- Y. Yang, H. J. Wang, and X. He. Posterior Inference in Bayesian Quantile Regression with Asymmetric Laplace Likelihood. *International Statistical Review*, 84(3):327–344, 2016.

- K. Yu and M. C. Jones. Local Linear Quantile Regression Local Linear Quantile Regression. *Journal of the American Statistical Association*, 93(441):228–237, 1998.
- K. Yu and R. Moyeed. Bayesian quantile regression. *Statistics & Probability Letters*, 54: 437–447, 2001.
- Y. R. Yue and H. Rue. Bayesian inference for additive mixed quantile regression models. *Computational Statistics and Data Analysis*, 55(1):84–96, 2011.



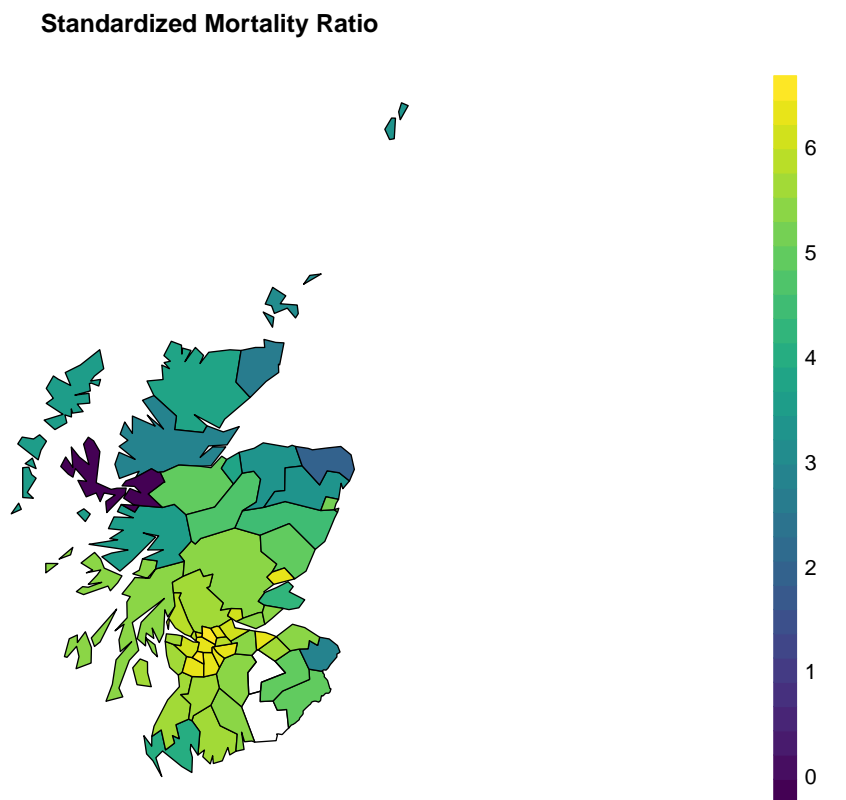
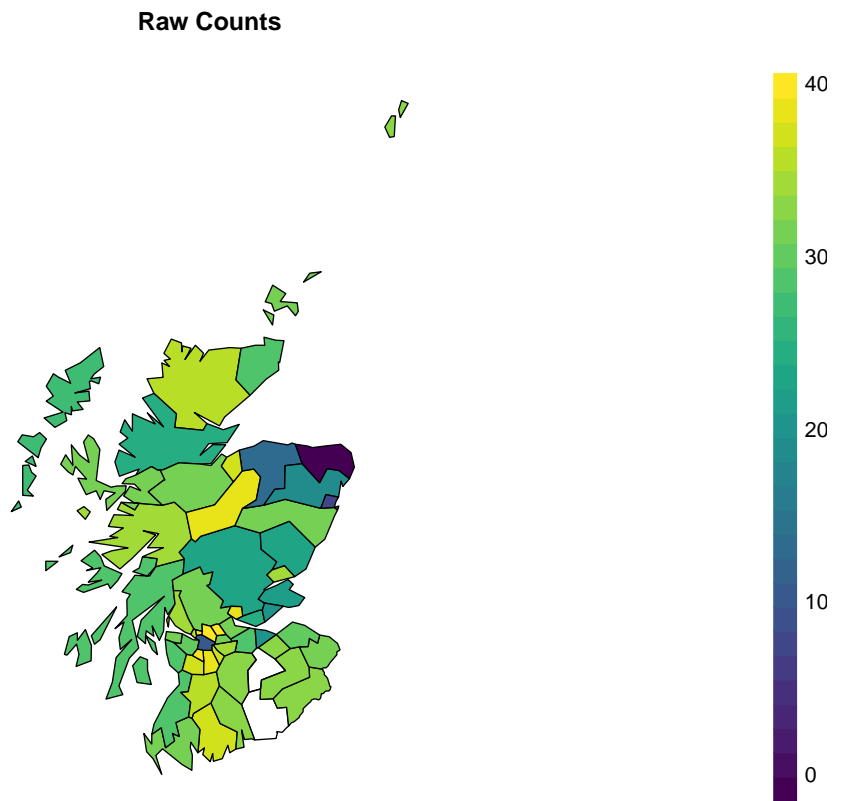
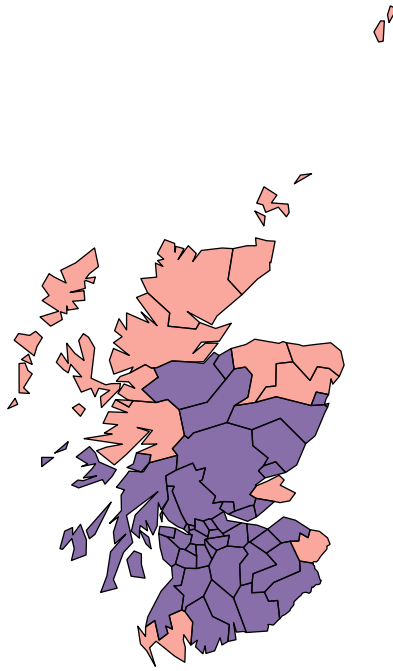


Figure 4: Incidence (top) and Standardized Mortality Ratio (bottom) of lip cancer in Scotland.

### Detected by Quantile Regression

- Low Risk
- High Risk



### Detected by Exceedance probability

- Low Risk
- High Risk

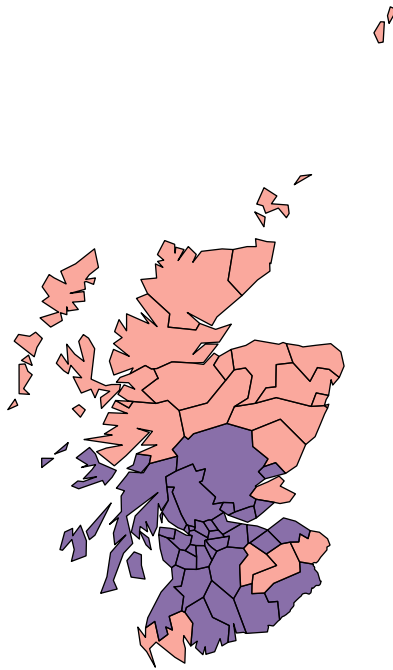


Figure 5: Areas at high risk of lip cancer according to quantile regression (top) and exceedance probability (bottom).