Ecological Informatics 75 (2023) 102085

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

Embracing asymmetry in nature: How to account for skewness in ecological data

Perry J. Williams[a,*], Xinyi Lu[b], Henry R. Scharf[c], Mevin B. Hooten[d]

[a] *Department of Natural Resources and Environmental Science, University of Nevada, Reno, 1664 N Virginia St, Reno 89557, NV, USA*
[b] *Colorado Cooperative Fish and Wildlife Research Unit, Colorado State University, 711 Oval Drive, Fort Collins 80523, CO, USA*
[c] *Department of Mathematics and Statistics, San Diego State University, 5500 Campanile Dr, San Diego 92182, CA, USA*
[d] *Department of Statistics and Data Sciences, The University of Texas at Austin, 110 Inner Campus Drive, Austin 71782, TX, USA*

ARTICLE INFO

Keywords:
Asymmetric Laplace distribution
Asymmetric normal distribution
Count data
Expectile regression
Henslow's sparrow
Quantile
Skewness

ABSTRACT

Asymmetric regression is an alternative to conventional linear regression that allows us to model the relationship between predictor variables and the response variable while accommodating skewness. Advantages of asymmetric regression include incorporating realistic ecological patterns observed in data, robustness to model misspecification and less sensitivity to outliers. Bayesian asymmetric regression relies on asymmetric distributions such as the asymmetric Laplace (ALD) or asymmetric normal (AND) in place of the normal distribution used in classic linear regression models. Asymmetric regression concepts can be used for process and parameter components of hierarchical Bayesian models and have a wide range of applications in data analyses. In particular, asymmetric regression allows us to fit more realistic statistical models to skewed data and pairs well with Bayesian inference. We first describe asymmetric regression using the ALD and AND. Second, we show how the ALD and AND can be used for Bayesian quantile and expectile regression for continuous response data. Third, we consider an extension to generalize Bayesian asymmetric regression to survey data consisting of counts of objects. Fourth, we describe a regression model using the ALD, and show that it can be applied to add needed flexibility, resulting in better predictive models compared to Poisson or negative binomial regression. We demonstrate concepts by analyzing a data set consisting of counts of Henslow's sparrows following prescribed fire and provide annotated computer code to facilitate implementation. Our results suggest Bayesian asymmetric regression is an essential component of a scientist's statistical toolbox.

1. Introduction

Observations in nature rarely meet the assumptions associated with basic parametric models taught in applied statistics courses (Fowler, 1990). Researchers seek to describe patterns and learn about mechanisms that give rise to observed data, but they often have to contend with 1) measurement error, 2) lack of conditional independence, 3) overdispersion, and 4) asymmetry — characteristics that are not accounted for using the standard toolbox. Among the remedial approaches for such cases, hierarchical models can be used to address all of the aforementioned issues (Clark, 2003; Cressie et al., 2009; Hobbs and Hooten, 2015; Hooten and Hefley, 2019). Such models are the subject of many studies in nature, and while the first three issues are commonly discussed (e.g., Winship et al., 2012; Hefley et al., 2017; Ver Hoef and Boveng, 2007), explicit approaches to deal with the fourth issue of asymmetry (i.e., skewness) are less well-known and thus our focus herein.

In what follows, we adopt a parametric model-based perspective that is prominent in likelihood and Bayesian analyses. That is, we take the well-accepted approach of building generative models for observed data that are based on conditional probability distributions for data, processes, and parameters (Berliner, 1996). The conventional approach to construct statistical models first involves the choice of a probability distribution with support that matches the data. For example, if the data y_i, for $i = 1, \ldots, n$, are counts (i.e., non-negative integers), then a Poisson distribution might be chosen as a model that could have given rise to them conditional on the Poisson intensity parameter λ_i, which may vary with i. The choice of a Poisson data model comes with an associated conditional mean, variance, kurtosis (i.e., heavy- or light-tailedness), and skewness (along with higher order moments that describe features

* Corresponding author.
E-mail address: perryw@unr.edu (P.J. Williams).

https://doi.org/10.1016/j.ecoinf.2023.102085
Received 29 November 2022; Received in revised form 22 March 2023; Accepted 23 March 2023
Available online 21 April 2023
1574-9541/

of the probability distribution). Therefore, the chosen data model inherits a set of properties automatically, making it very convenient to specify parametric statistical models quickly.

It is widely known that we should check the assumptions of statistical models to determine if they could have given rise to our data (Conn et al., 2018). For example, dispersion is among the most common characteristics checked in parametric statistical models. One measure of dispersion is the variance (i.e., the second central moment). If model checking reveals that our chosen distribution is not able to provide enough variance to match that present in the data, we may select a so-called overdispersed distribution to represent the data instead (e.g., a negative binomial distribution instead of a Poisson distribution, Ver Hoef and Boveng, 2007; Lindén and Mäntyniemi, 2011). Alternatively, we may choose to insert an additional stochastic level of hierarchy to the model to provide extra dispersion via nested randomness.

While it is common to adjust model specifications to address measurement error, dependence, and dispersion, the skewness implied by chosen probability distributions is often accepted without investigation (with few exceptions, e.g., Komori et al., 2016). The skewness of a probability distribution is defined as the standardized 3rd moment. For example, a Poisson distribution with mean (and variance) λ has skewness $\lambda^{-\frac{1}{2}}$ (Johnson et al., 2005). Thus, for large intensities, the Poisson distribution increases in mean and variance, but decreases in skewness such that the distribution becomes symmetric as $\lambda \rightarrow \infty$. By comparison, had we chosen the negative binomial distribution such that $y_i \sim \text{NegBinom}(\lambda_i, N)$, the skewness would be

$$\lambda^{-\frac{1}{2}} \frac{\frac{N+2\lambda}{N+\lambda}}{\left(\frac{N}{N+\lambda}\right)^{\frac{1}{2}}}, \tag{1}$$

which converges to $\lambda^{-\frac{1}{2}}$ as $N \rightarrow \infty$ and the negative binomial converges to the Poisson distribution. The skewness associated with the negative binomial distribution (1) is always larger than that of the Poisson distribution when $N < \infty$.

Critical evaluation of skewness provides an opportunity to contribute to our understanding of the natural world (Kozłowski and Gawelczyk, 2002; Wool et al., 1980). We suggest skewness should be considered in the construction of parametric statistical models for ecological data just as the mean, variance, and kurtosis often are considered. We discuss asymmetric probability distributions that address skewness and can be used to model continuous-valued response variables (as well as processes and parameters in hierarchical models). We then focus on the asymmetric Laplace distribution (ALD) and show how it can be used in a regression context. Fixing the asymmetry parameter in the ALD to a preselected skewness has ties to quantile regression and we discuss advantages and disadvantages of doing so. We provide an example model for count data that uses the ALD as a process model and apply it to perform asymmetric regression to improve our understanding of population response to prescribed fire for a threatened bird species (Henslow's sparrow).

2. Asymmetric continuous distributions

A variety of distributions are commonly used to model discrete-valued data (e.g., binomial, Poisson) and they each have associated skewness. For continuous-valued variables that are bounded below by zero, most probability distributions are skewed. However, for those that are unbounded, the normal (or Gaussian) distribution is the most commonly specified in statistical models by far. The normal distribution has useful properties (such as marginal, conditional, and joint forms) and arises naturally in large sums and averages. However, the normal distribution is conditionally symmetric regardless of its location or dispersion.

An alternative to the normal distribution is the asymmetric normal distribution (AND; Waldmann et al., 2017). For observation y, suppose that $y \sim \text{AND}(\mu, \sigma^2, \tau)$, where the AND probability density function is specified as

$$[y|\mu, \sigma^2, \tau]_{\text{AND}} \equiv \frac{2}{\sqrt{\sigma^2 \pi}} \left(\sqrt{\frac{1}{1-\tau}} + \sqrt{\frac{1}{\tau}} \right)^{-1} \exp\left(-\frac{(y-\mu)^2}{\sigma^2} |\tau - I_{\{y<\mu\}}| \right), \tag{2}$$

and the parameters μ and σ^2 represent central tendency and dispersion (but not necessarily mean and variance). When the asymmetry parameter $\tau = 0.5$, then μ is the mean of the distribution and it is symmetric.

By comparison, the asymmetric Laplace distribution (ALD) has probability density function

$$[y|\mu, \sigma^2, \tau]_{\text{ALD}} \equiv \frac{\tau(1-\tau)}{\sigma} \exp\left(-\frac{y-\mu}{\sigma} (\tau - I_{\{y<\mu\}}) \right), \tag{3}$$

where the parameter μ is the median of the distribution when $\tau = 0.5$ (Yu and Zhang, 2005). The ALD is notably more sharply peaked than the normal and is commonly used as a prior in Bayesian models where lasso regularization is performed (Park and Casella, 2008; Hooten and Hobbs, 2015).

The parameter τ controls skewness in both the AND and ALD whereby, as τ varies from 0.5, the distribution becomes more skewed in one direction or the other. Fig. 1 compares the normal, AND, and ALD probability density functions with parameters set at $\mu = 1$ and $\sigma = 0.1$, and $\tau = 0.1$. Notice that the skewness parameter $\tau < 0.5$ skews the asymmetric distributions to the right and it has a much more pronounced effect on the ALD than it does on the AND.

3. Regression using asymmetric models

The linear regression model

$$y_i \sim \text{N}(\mathbf{x}_i' \boldsymbol{\beta}, \sigma^2), \tag{4}$$

for $i = 1, \ldots, n$ observations, is most commonly used to describe how the mean of a continuous response variable $\text{E}(y_i|\boldsymbol{\beta}, \sigma^2)$ changes with respect to predictor variables of interest $\mathbf{x}_i$,

$$\text{E}(y_i|\boldsymbol{\beta}, \sigma^2) = \mathbf{x}_i' \boldsymbol{\beta}. \tag{5}$$

When the response variable has a skewed distribution, the Box-Cox transformation or power transformation may be used to remove skewness. However, sometimes transformation alone is not enough, or asymmetry itself is of interest. Therefore, it is necessary to seek distributions that model skewness explicitly.

An alternative to the implicit assumption of symmetry in (4) is to specify a distribution for y_i that can accommodate asymmetry, such as the ALD or AND. Generalizing traditional symmetric regression by including an unknown skewness parameter, τ, allows researchers to account for asymmetry in the response.

The ALD can be formally incorporated into a linear regression model by assuming

$$y_i \sim \text{ALD}(\mathbf{x}_i' \boldsymbol{\beta}, \sigma, \tau), \tag{6}$$

in place of the assumption of normality in (4), where $\mathbf{x}_i' \boldsymbol{\beta}$ and σ represent the location and scale parameters, respectively, as in (4), but now, the additional parameter $\tau \in (0, 1)$ represents the skewness of the response distribution.

Use of the ALD in regression modeling is advantageous in that it yields more robust inference when the true distribution of y_i is not normal, and there are only minor losses in efficiency compared to the normal distribution when the data truly are normal (Huber, 1981). Another key advantage of (3) is that it has connections to classical quantile regression (we return to this point in the next section).

Similarly, if a more rounded distribution is preferred to model the

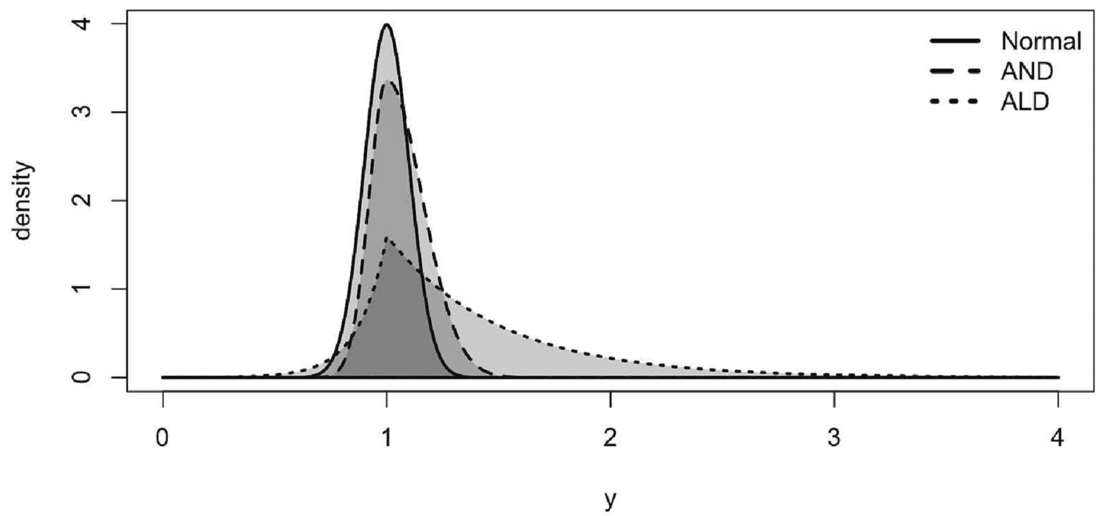


Fig. 1. Comparison of normal probability density function (solid line) with the AND (dashed line) and ALD (dotted line) using $\mu = 1, \sigma = 0.1$, and $\tau = 0.1$.

response variables in regression, then the AND provides an alternative to ALD regression. In AND regression, we use the AND distribution from (2) with $\mu_i = \mathbf{x}_i'\boldsymbol{\beta}$ such that

$$y_i \sim \text{AND}(\mathbf{x}_i'\boldsymbol{\beta}, \sigma, \tau), \quad (7)$$

where again, the parameters $\boldsymbol{\beta}, \sigma$, and τ are estimated from the data.

3.1. Example: plant trait relationships

Leger et al. (2019) described a study where they found that phenotypic traits of a perennial grass (*Elymus elymoides*) were predictive of its performance in an arid ecosystem in the Western United States (U.S.). We examined their data to characterize the relationship between seed weight (mg) and the logarithm of plant biomass (in log mg, averaged across cultivars). To analyze these data, we fit three Bayesian regression models while considering the plant performance measured in terms of log plant biomass as the response variable (y_i for $i = 1, \ldots, n$ and $n = 35$ sites) and seed weight as the predictor variable (x_i). The first model was a conventional Gaussian regression model (4), the second model was an ALD regression (6), and the third model was an AND regression (7). In all three of the models we used the same prior distributions for $\boldsymbol{\beta}$ and σ ($\boldsymbol{\beta} \sim \text{N}(\mathbf{0}, 1000 \cdot \mathbf{I})$ and $\sigma \sim \text{Uniform}(0, 100)$). For the ALD and AND regressions, we used a uniform prior for the skewness parameter ($\tau \sim \text{Uniform}(0, 1)$).

We fit the three regression models using Markov Chain Monte Carlo (MCMC) algorithms and 200,000 iterations. We computed posterior predictive p-values for each model using both skewness and kurtosis statistics (Conn et al., 2018) as well as the deviance information criterion (DIC Spiegelhalter et al., 2002; Hooten and Hobbs, 2015) to compare the predictive performance of the models (Table 1). We also computed marginal posterior means and 95% equal-tailed credible intervals for all model parameters (Table 1). The results obtained by fitting the regression models to the plant trait data suggest the normal assumption is not appropriate for these data; a common finding when modeling data observed in nature (Box, 1976). While the posterior predictive p-value for kurtosis does not indicate the tails of the distribution are too heavy or light, the p-value for skewness suggests that the data are more skewed than that allowed by the normal (Table 1). These results indicate that a skewed distribution is more appropriate for describing the relationship between seed weight and biomass in this plant species.

In contrast with the normal model, neither asymmetric models indicated issues with skewness or kurtosis assumptions for these data (Table 1). However, the DIC score for model selection indicated better

Table 1
Posterior mean and 95% credible interval of model parameters, predictive p-values, and DIC.

	Normal	ALD	AND
$E(\beta_0\|\mathbf{y})$ (CI)	0.113 (−0.493, 0.719)	−0.435 (−0.799, −0.026)	−0.295 (−0.775, 0.292)
$E(\beta_1\|\mathbf{y})$ (CI)	0.301 (0.107, 0.494)	0.346 (0.227, 0.453)	0.327 (0.173, 0.470)
$E(\sigma\|\mathbf{y})$ (CI)	0.526 (0.414, 0.678)	0.111 (0.414, 0.678)	0.315 (0.121, 0.543)
$E(\tau\|\mathbf{y})$ (CI)	—	0.205 (0.078, 0.371)	0.133 (0.012, 0.421)
skewness p-value	0.023	0.692	0.175
kurtosis p-value	0.219	0.770	0.312
DIC	55.736	43.412	47.766

predictive performance with the ALD regression model. This is confirmed by the posterior predictive distributions (PPDs) shown in Fig. 2. Notice that the PPD resulting from fitting the normal regression model to the plant trait data only slightly underestimates the heaviness of the upper tail of the distribution, but is too conservative on the lower tail because it is symmetric. By comparison, the PPD resulting from the AND model (Fig. 2c) captured the central tendency but resulted in an upper tail that was slightly too light which is related to the somewhat small posterior predictive p-value for skewness (Table 1). Finally, the ALD posterior predictive distribution (Fig. 2b) captures the central tendency, variation, and skewness of the data well, which confirms the DIC results and supports the utility of asymmetric statistical models in practice.

Overall, our results indicate that the ALD regression predicts the data better than the other models and does not indicate lack of fit. Thus, we should use the ALD model for inference and prediction for these data. In this case, our ALD regression model suggests a positive relationship between seed weight and log biomass while accounting for asymmetry by reducing τ so that the model is positively skewed to accommodate the data.

4. The ALD and Bayesian quantile regression

Quantile regression seeks to infer the relationship between predictors and a selected quantile (or quantiles) of the response data (e.g., Koenker and Bassett, 1978), and relaxes the assumption made in most parametric regression models that this relationship has the same linear

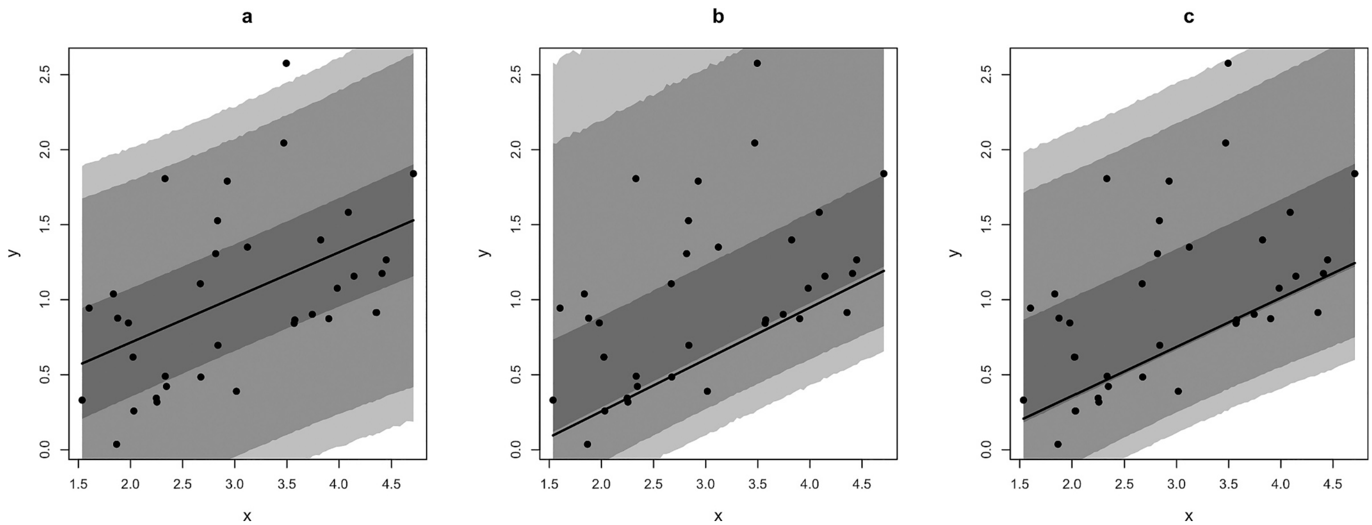


Fig. 2. Comparison of posterior predictive distribution resulting from a.) normal, b.) ALD, and c.) AND regression models applied to perennial plant trait data (black points). In the background, 99% (light shading), 95% (medium shading), and 50% (dark shading) equal-tailed point-wise credible intervals. The point-wise maximum *a posteriori* prediction is shown as solid black line.

pattern for each quantile. Quantile regression analyses have been used effectively by scientists as an alternative to, or in supplement of, classic linear regression models for 20 years (Cade and Noon, 2003). The use of quantile regression concepts to infer characteristics of quantiles in data have provided novel, and otherwise unobtainable, insight on natural processes such as climate change (Koenker and Schorfheide, 1994), species distributions (Cade et al., 1999), and predator–prey relationships (Scharf et al., 1998).

There are at least three general approaches to quantile regression models. They include distribution-free or semiparametric approaches (e.g., Koenker and Bassett, 1978), maximum likelihood approaches (e.g., Gilchrist, 2000), and approaches based on the ALD (e.g., Yu and Moyeed, 2001). Our methods fall under the ALD approach, but we discuss general properties of quantile regression, below.

The traditional approach to quantile regression requires the user to pre-specify τ and then estimate $\boldsymbol{\beta}(\tau)$ by solving the minimization problem

$$\boldsymbol{\beta}(\tau) = \min_{\boldsymbol{\beta}\in\mathbb{R}} \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\boldsymbol{\beta})(\tau - I_{\{y_i<\mathbf{x}_i'\boldsymbol{\beta}\}}). \tag{8}$$

The term "quantile regression" is used because the optimization is consistent with assuming only that the τ-quantile of the distribution of $y_i|\boldsymbol{\beta}(\tau)$ is linear in $\mathbf{x}_i$ (i.e., $q_\tau(y_i|\boldsymbol{\beta}) = \mathbf{x}_i'\boldsymbol{\beta}$). Quantile regression therefore relies on weaker assumptions than classical regression. For instance, the variance of y_i is not required to be homoskedastic.

The loss function in (8) penalizes outliers less than squared error loss (Williams and Hooten, 2016), thus quantile regression is more robust to outliers than the classic linear regression model described in (4) (Koenker and Bassett, 1978; Huber, 1981). In comparison to (4), regression quantile estimates do not depend on an assumed form of error distribution (Koenker and Bassett, 1978). There are several non-Bayesian applications of quantile regression for continuous data. There is also well developed software (e.g., the `quantreg` package in R) to facilitate implementation.

Quantile regression for continuously valued response data and fixed quantile(s) of interest, τ, was extended to the Bayesian framework by Yu and Moyeed (2001). The authors developed an approximate method for Bayesian inference that relies on fitting a collection of separate parametric models based on the ALD (3) with pre-selected skewness values. Maximizing the likelihood of the ALD is directly related to the minimization function in (8) (Koenker and Machado, 1999). Thus, a likelihood function based on the ALD is a natural component of a Bayesian quantile regression model (Yu and Moyeed, 2001). Inference about model parameters can be obtained using conventional Bayesian methods for each separate model fit (e.g., MCMC; Appendix A).

The Bayesian implementation of quantile regression has advantages over traditional approaches including accounting for parameter uncertainty, incorporating *a priori* information, and the ability to fit more realistic hierarchical models (Hooten and Hefley, 2019). Additionally, the Bayesian framework is particularly useful for modeling discrete-response data. However, for those interested in asymptotic data settings and frequency interpretations of Bayesian inference, Bayesian quantile regression may not be consistent (i.e., converge to true value under large n; Yang et al., 2016). Consistent asymptotic properties have been established in the non-Bayesian quantile regression framework, however, limitations include difficulties optimizing the regression parameters of discrete data (Florios and Skouras, 2008) and building confidence intervals and covariance matrices for parameter estimates (Abrevaya and Huang, 2005; Machado and Silva, 2005).

In general, quantile regression is a useful technique for exploratory data analysis. In addition, the use of an ALD likelihood allows for quantile regression-like inference when a single percentile τ is pre-selected in that it provides insights about the relationships between response and predictors in the tails of the data. Similarly, the AND likelihood can be used for expectile regression and provides a different type of insight about the tails of the response distribution. However, quantile and expectile regression models are not generative in the sense that they cannot give rise to the observed data. Asymmetric regression techniques, on the other hand, are generative when the skewness parameter τ is estimated, as in our previous example.

5. Bayesian asymmetric and quantile regression for counts

Count data are typically modeled by generalizing classic linear regression such that

$$\begin{aligned} y_i &\sim [y_i|\lambda_i,\kappa], \\ \lambda_i &= g(\mathbf{x}_i'\boldsymbol{\beta}) \end{aligned} \tag{9}$$

(Nelder and Wedderburn, 1972), where $[a|b]$ represents the probability mass function (PMF; e.g., Poisson, negative binomial) of a conditional on b, λ_i is the mean of the PMF, κ is a dispersion parameter, when appropriate, and g is an appropriate link function (e.g., the exponential function).

Discrete data, including binary data and count data, do not have continuous quantiles, precluding the use of traditional quantile regression. Machado and Silva (2005) extended quantile regression for count data in a non-Bayesian framework using a specific form of jittering (Stevens, 1950). In this approach, the response data are transformed stochastically by adding uniformly distributed noise to the count variable $z_i = y_i + u_i$, for $u_i \sim \text{Uniform}(0,1)$. The quantile function for the transformed response variable z_i is

$$q_\tau(z_i|\boldsymbol{\beta}(\tau),\sigma^2) = \exp(\mathbf{x}_i'\boldsymbol{\beta}(\tau)) + \tau. \tag{10}$$

Addition of τ in (10) is a result of $\mathbf{z}$ being bounded below by τ due to the addition of the uniform random variable. Machado and Silva (2005) showed that (10) results in the continuous variable z_i with conditional quantiles that have a one-to-one relationship with the conditional quantiles of the counts, permitting z_i to provide a basis for inference. Standard inference methods used for y_i are asymptotically valid for z_i (Machado and Silva, 2005). The relationship between the quantiles of z and y is (Machado and Silva, 2005)

$$q_\tau(z_i|\boldsymbol{\beta}(\tau),\sigma^2) = q_\tau(y_i|\boldsymbol{\beta}(\tau),\sigma^2) + \frac{\tau - \sum_{y=0}^{q_\tau(y_i|\boldsymbol{\beta}(\tau),\sigma^2)-1}[Y=y|\boldsymbol{\beta}(\tau),\sigma^2]}{[Y=q_\tau(y_i|\boldsymbol{\beta}(\tau),\sigma^2)|\boldsymbol{\beta}(\tau),\sigma^2]}. \tag{11}$$

A Bayesian hierarchical asymmetric regression model for ecological count data is

$$\begin{aligned}
z_i &= \begin{cases} \log(y_i+u_i-\tau), & y_i+u_i>\tau \\ \log(\eta), & y_i+u_i\leqslant\tau \end{cases}, \\
z_i &\sim \text{ALD}(\mathbf{x}_i'\boldsymbol{\beta},\sigma,\tau), \\
u_i &\sim \text{Uniform}(0,1), \\
\boldsymbol{\beta} &\sim \text{Normal}(\boldsymbol{\mu}_\beta,\boldsymbol{\Sigma}_\beta), \\
\sigma &\sim \text{Uniform}(0,u_\sigma).
\end{aligned} \tag{12}$$

In the absence of prior information about the skewness in the process z_i, a natural prior distribution for τ is

$$\tau \sim \text{Uniform}(0,1). \tag{13}$$

One approach to Bayesian quantile regression for count data is to adopt the above model for a collection of fixed values of τ, inferring parameter values through multiple simultaneous model fits, as is done with continuous data.

The mixture model at the top of (12) is required for count data that contain zeros because $P(u_i - \tau \leqslant 0|y=0) > 0$, and log(0) is undefined. In previous applications of quantile regression for count data, either $y_i \geqslant 1, \forall i \in 1,\dots,n$, precluding the necessity to chose η (e.g., Lee and Neocleous, 2010), or η has been selected by the investigator to be a "suitably small positive number" (Machado and Silva, 2005, p.1231). To date, no definition of "suitably small" has been provided for quantile regression of counts. To preserve the ordering of z_i in relation to the count data y_i, η must fall between 0 and $1 + u^* - \tau$, where $u^* = \min\{u_i : y_i + u_i > 1, \text{ and } y_i + u \leqslant 2\}$. The choice of η affects inference (larger values place more emphasis on information contained in zeros than smaller values), and the magnitude of the effect depends on the amount of zeros in the data (Appendix B). Instead of choosing one value for η, as recommended by Machado and Silva (2005) another potential approach is to impute values by sampling $\eta^{(k)} \sim \text{Uniform}(0, 1 + u_i^{(k)*} - \tau)$ for each $k = 1,\dots,K$ MCMC iteration. Thus, uncertainty in the choice of η is considered in parameter estimates during analysis. We demonstrate this process and provide R code in Appendix B.

The hierarchical framework in (12) could also be used to account for false negative observations. For example, if replicate counts of species $c_{i,t}$ are taken at site i during time $t = 1,\dots,T$, and assuming y_i is the latent, true abundance at site i, we can estimate the relationship between the observed counts, true abundance, and detection probability p, by assuming $c_{i,t} \sim \text{Binomial}(y_i, p)$.

6. Bayesian ALD for improved prediction and inference

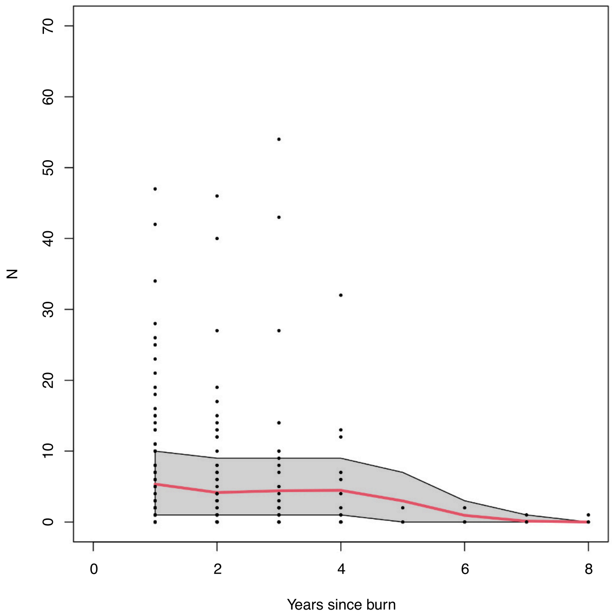
The hierarchical model in (12) used for quantile regression assumes an investigator is interested in one or more specific quantiles of interest, and therefore selects values of τ *a priori* before fitting the models. However, as we showed in our previous asymmetric regression example, using the ALD (or AND) as a likelihood allows us to estimate τ as an unknown parameter in an asymmetric regression model. When the skewness parameter, τ, in the ALD is treated as unknown and estimated from the data, regression for count data using the ALD represents a flexible alternative to Poisson regression that can account for, and provide inference on, additional skewness in the data. The approach is similar to specifying a negative binomial distribution with unknown dispersion parameter for the response but accounts for additional skewness rather than overdispersion in the data. Further, if a symmetric model, such as the symmetric normal distribution in Eq. (4), is fit to asymmetric data, parameter estimates may fail to recover the true values of the parameter. We evaluated this behavior in Appendix C. Specifically, we found that if asymmetric data are generated using the AND, and fit using Eq. (4), 95% credible intervals of the intercept parameter did not cover the value of the parameter that was used to generate data when the asymmetry parameter (τ) was < 0.4 or > 0.6. However, accounting for asymmetry permitted 95% credible invervals of parameter estimates to reliably cover true values of the parameter for all values of τ (Appendix C). As we demonstrate in the following application to counts of Henslow's sparrows following a prescribed burn, the ALD regression approach has the potential to offer greater predictive power compared to both Poisson and negative binomial regression while still retaining an interpretable functional form.

7. Application: population response to prescribed fire

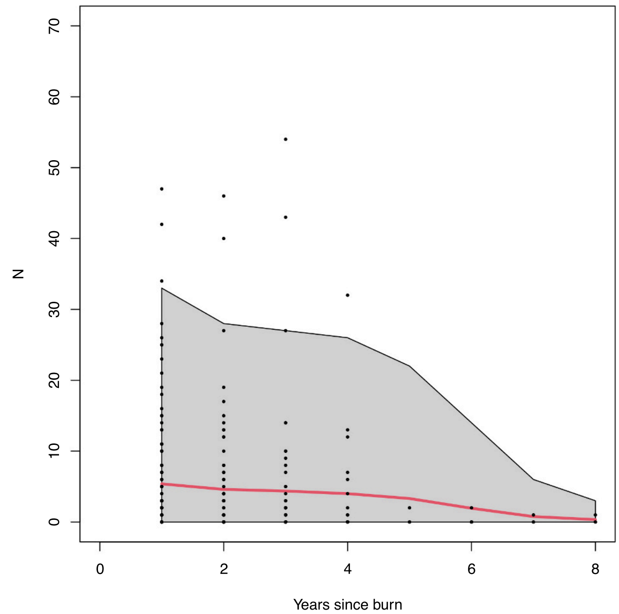
We consider data consisting of counts of Henslow's sparrows following prescribed fire (Fig. 3) collected at Big Oaks National Wildlife Refuge in southeastern Indiana, USA. The Henslow's sparrow has experienced substantial range-wide population declines the last half-century (Herkert et al., 2002). Big Oaks National Wildlife Refuge has one of the largest remaining breeding populations of Henslow's sparrows in the world. Henslow's sparrow population declines appear to be associated with loss or degradation of their preferred habitat. Preferred habitat consists of grasslands with tall and dense grass, a thick litter layer, and sparse or no woody vegetation (Herkert et al., 2002). Woody encroachment is a major driver of Henslow's sparrow habitat loss (Herkert et al., 2002). Prescribed fire is the primary management tool used to prevent or delay woody encroachment at Big Oaks National Wildlife Refuge. After a prescribed fire has been implemented, biologists collect count data to evaluate the population response to the fire, and to guide future management actions (Williams and Hooten, 2016). Count data have been collected for at least eight years post fire at some sites. Henslow's sparrows do not typically use a grassland the same year it is burned (year zero) due to sparse vegetation, but counts are usually higher one year after a burn (year 1) compared to pre-burn counts.

7.1. Estimating skewness using ALD regression

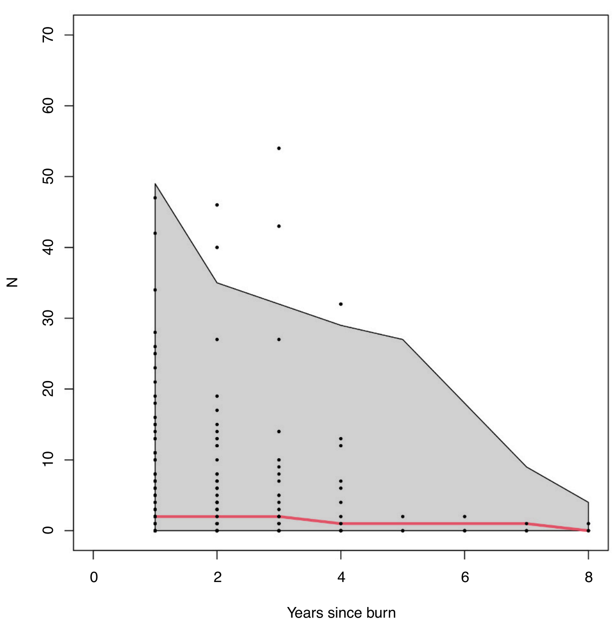
We first explored the Henslow's sparrow data in Fig. 3 to identify some summary statistics. We calculated the mean and median counts of Henslow's sparrows conditional on the covariate years since burn (Appendix A). We found that mean Henslow's sparrow counts, conditional on years since burn, were always higher than the median Henslow's sparrow counts (Appendix A). In symmetric distributions, the median equals the mean. These results suggested that our data were asymmetric and may be better represented with an asymmetric model. We fit three statistical models to the data presented in Fig. 3. The first two models represent conventional approaches, and the third is a novel ALD-based approach. Specifically, we assumed y_i were realizations from a Poisson



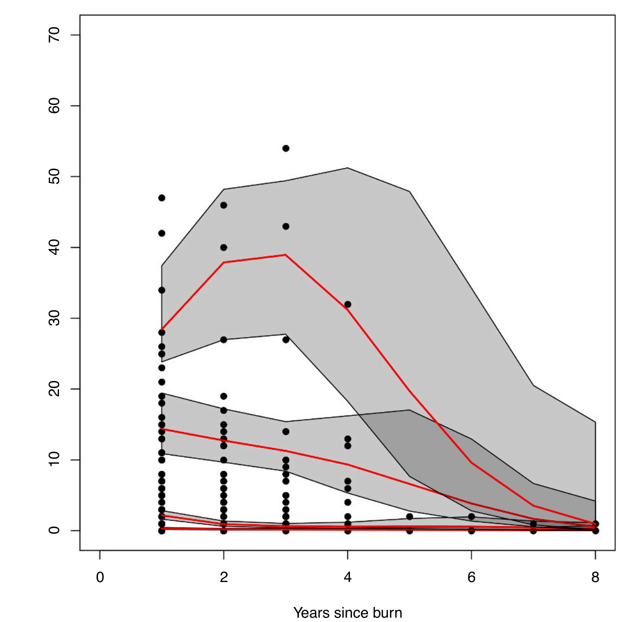
(a) Poisson



(b) Negative binomial



(c) ALD regression, τ estimated



(d) Quantile regression, τ selected

Fig. 3. Henslow's sparrow counts collected from 305 plots at Big Oaks National Wildlife Refuge, Indiana, USA, between 1996 and 2007. Fig. 3a shows the estimated Poisson regression process model, Fig. 3b shows the estimated negative binomial process model, Fig. 3c shows the estimated ALD process model with estimated quantile, τ, and Fig. 3d shows the estimated quantile regression process model with τ selected to equal (from bottom to top) 0.1, 0.5, 0.9, and 0.975 (red line = estimated posterior mean of the process model, gray polygon = 95% credible intervals).

distribution in the first model (Fig. 3a), and negative binomial distribution in the second model (Fig. 3b). The third model we fit to the data was the hierarchical ALD model in (12), where we treated τ as an unknown parameter to be estimated, with prior distribution $\tau \sim \text{Uniform}(0, 1)$.

The Poisson distribution assumes the variance equals the mean. The negative binomial distribution and ALD relax this assumption and allow more flexibility using an additional over-dispersion/skewness parameter, κ/τ. We used an exponential inverse link function for f. We used three basis vectors for $\mathbf{x}_1, \mathbf{x}_2$, and $\mathbf{x}_3$, generated from basis spline functions with 3 degrees of freedom over the range 0–8 (see Appendix A). Basis vectors permit flexible modeling of non-linear processes (Hefley et al., 2017). We used exchangeable normal priors with mean equal to zero and variance equal to 10 for coefficients in $\boldsymbol{\beta}$, and for the negative binomial distribution, a log-normal prior distribution for κ with mean and variance equal to log(2) and log(3), respectively. For the ALD model, we set the hyperparameter for the prior distribution of the scale parameter to $u_\sigma = 10$.

We fit each model to the Henslow's sparrow data using a MCMC algorithm written in R. The full algorithm for each model is provided in Appendix A. We evaluated convergence by visually inspecting trace plots. We assessed model fit using Bayesian p-values, and compared model predictive ability using the widely applicable Bayesian information criterion (WAIC; Watanabe, 2013; Hooten and Hobbs, 2015; Conn et al., 2018).

The estimated process models are provided in Fig. 3. All models we fit suggested that, on average, Henslow's sparrow counts decreased three years after prescribed fire, suggesting periodic prescribed fire is important to improve Henslow's sparrow habitat through time (Fig. 3). The Poisson model exhibited lack of fit (posterior predictive p-value ≈ 1.00), and had poor predictive ability relative to the other models (WAIC = 3,572). This was due to the over-dispersed nature of the data with respect to the Poisson regression model. The negative binomial distribution improved the model fit compared to the Poisson regression model (posterior predictive p-value = 0.45) by explicitly accounting for over-dispersion with the additional parameter, κ. The posterior mean of $\kappa = 0.35$ (95% CI (0.27, 0.42); Appendix A). Estimated predictive ability also increased with the negative binomial distribution compared to the Poisson distribution (WAIC = 1,470). The ALD regression model showed no lack of model fit (posterior predictive p-value = 0.33) and was the best predictive model (WAIC = 1,378). The estimated posterior mean of the skewness parameter was $E(\tau|\mathbf{y}) = 0.55$ (95% CI = (0.42, 0.70)). Values greater than 0.5 suggest a distribution with more mass associated with lower counts while values less than 0.5 suggest a distribution with more mass associated with higher counts. An estimated skewness parameter greater than 0.5 was expected for Henslow's sparrows given the empirical distribution of the observed data (Fig. 3a). The added flexibility of estimating τ in the asymmetric regression model allowed us to predict better the skewed (toward 0) nature of the data observed in (Fig. 3a) than would be predicted by either the Poisson or negative binomial process model. The skewness in the ALD also permits enough mass in the upper tail of the distribution to permit large counts of animals, a common observation in wildlife ecology that is seldom addressed.

7.2. *Inference on tails of the data: Bayesian quantile regression*

Scientists are often interested in the effect of prescribed fire (or other disturbances) on wildlife resource use. However, wildlife response to disturbance can be limited by latent, unobservable environmental factors, potentially unrelated to fire frequency. A population response to fire can be dampened, or eliminated if other limiting factors dominate.

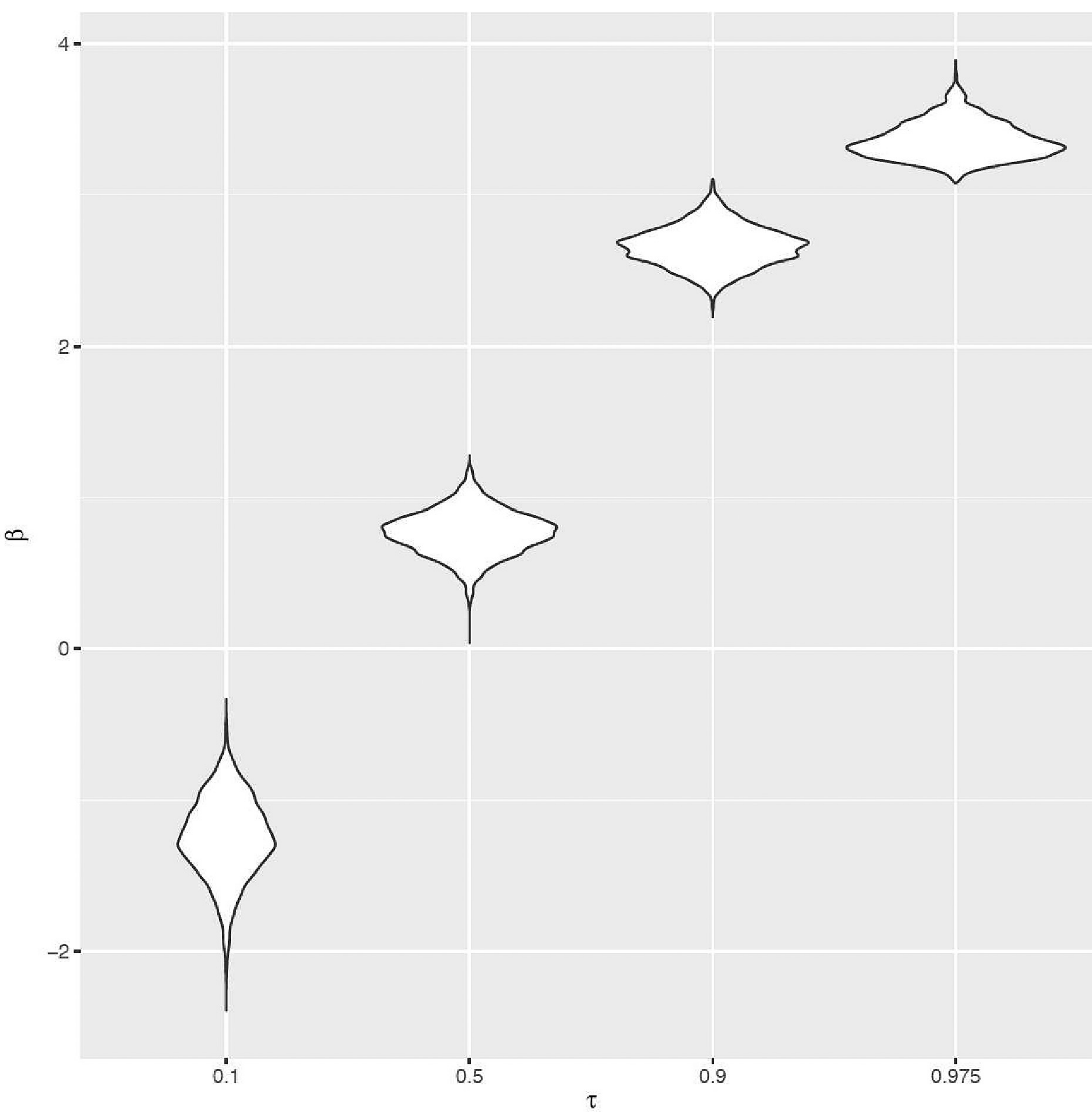


Fig. 4. Estimated marginal posterior distributions for $\beta_{1,\tau}$ for quantiles $\tau = 0.1, 0.5, 0.9, 0.975$ in the Henslow's sparrow response-to-fire model from data collected at Big Oaks National Wildlife Refuge, southeastern Indiana. Estimated differences in these parameters drive differences in the process model in Fig. 3d.

Addressing latent limiting environmental factors on population response to covariates in a cohesive modeling framework was an original motivation for applying quantile regression to ecological data (e.g., Cade and Noon, 2003). We incorporate these concepts in our application, and show how they can be extended to count data.

A common approach to making inference on regression quantiles is to choose several quantiles of interest, and examine how coefficient parameter estimates relate to different quantiles of the response distribution (e.g., Cade et al., 1999; Cade and Noon, 2003). For example, a predictor variable such as *time since burn* might have no effect on the response variable in habitats that have other, unmeasurable, limiting factors, but have a large effect on the response variable at sites where Henslow's sparrows were previously limited by dense vegetation and woody encroachment. These parameter estimates can be viewed/summarized using *forest plots* (Fig. 4).

Estimated process models fit with four selected quantiles, $\tau = 0.1, 0.5, 0.9, 0.975$, for the Henslow's sparrow data are shown in Fig. 3d. The process model for the top quantile selected, $\tau = 0.975$ reflects how the "best sites" (i.e., the sites with the highest counts of birds, conditional on the covariate *years since burn*) respond to prescribed fire. The best sites see an overall increase in the number of birds 2–3 years following prescribed fire, compared to decreases after one year following prescribed fire. There is more uncertainty in the upper tail of the response distribution due to less data in those quantiles (Fig. 3d). Similarly, the process model for the bottom quantile selected, $\tau = 0.1$, describes how the "worst sites" respond to prescribed fire. The worst sites see little change, or even a decrease in use, 2–4 years following a prescribed fire, compared to one year following a prescribed fire (Fig. 3d). We can also select the median response to prescribed fire by setting $\tau = 0.5$. Abundance at median sites decrease slightly two years following prescribed fire compared to one year after fire (Fig. 3d). Importantly, the effects at each quantile level need not be equal (Fig. 4), as they would be by definition in any conventional generalized linear model, including the cases of Poisson, negative binomial, and ALD response distributions.

8. Discussion

Asymmetric data present analysis challenges when using common statistical models. We demonstrated how incorporating the ALD and AND into familiar model structures permits scientists to appropriately model asymmetric data and improve predictive power while aiding natural interpretation. Developing models for asymmetric data in a hierarchical Bayesian framework facilitates the specification of a variety of asymmetric regression models and can be used to account for random effects, such as those that include spatial and temporal structure. The methods we introduced represent useful and readily implementable extensions to scientists' standard model-building toolbox.

We also demonstrated that the AND and the ALD can be used to perform expectile and quantile regression, even when the data are discrete counts. Quantifying different potential responses to a disturbance event, and the uncertainty associated with those responses, is critical to understand basic ecological mechanisms and how they vary in space and time. We demonstrated how Bayesian quantile regression can be used for count data through an application to Henslow's sparrows by examining the different response to prescribed fire at a variety of quantiles including the best sites (i.e., $\tau = 0.975$), the worst sites ($\tau = 0.1$), and the median ($\tau = 0.5$). Sites within our study area responded differently to prescribed fire. Some sites responded favorably, and others did not appear to benefit (Fig. 3d); presumably due to unmeasured local conditions. Quantifying the differential response to disturbance provides a formal framework for exploratory data analysis and fosters future investigation into the spatial, temporal, biotic, and abiotic mechanisms responsible for different responses to disturbance. By identifying and comparing sites in various quantiles, and their response to natural disturbance, we can begin to examine factors limiting the natural response, and improve management actions.

Compared to non-Bayesian approaches to quantile regression, the Bayesian framework provides for uncertainty quantification both in parameter estimates (Fig. 4) and the process of interest (e.g., counts of Henslow's sparrows Fig. 3d). Formally quantifying uncertainty in the process of interest provides a basis for targeted monitoring to optimally reduce future uncertainty (*sensu*Williams et al., 2018), with the ultimate goal of increased scientific understanding. In the case of Henslow's sparrows, a monitoring program could be developed to identify the mechanism responsible for why sites in the bottom quantiles do not respond as favorably to sites in the top quantiles.

In addition to estimating response to fire, conditional on user-selected quantiles, we discussed an asymmetric regression model for count data using the ALD. We demonstrated that the ALD, modified for count data, provides a better predictive model (assessed using WAIC) than other commonly used statistical models (i.e., Poisson, negative binomial).

Unlike its non-Bayesian counterpart, inference for Bayesian quantile regression proceeds via a collection of sub-models, each specified using an ALD at the level of the response with a fixed and selected skewness parameter (3). However even when errors in real, continuous-valued data are not distributed according to an ALD, empirical and theoretical results have shown posterior consistency using a misspecified ALD likelihood under general conditions (Yu and Moyeed, 2001; Sriram et al., 2013; Benoit and Van den Poel, 2017).

Methods and software exist for applying Bayesian quantile regression to binary data. Although these methods may be appealing for scientists who collect binary data, the methods have not been fully explored. We refer readers to Scharf et al. (2022) for a more in-depth discussion of regression methods for binary data, including quantile regression, illustrated using ecological applications.

Statistical theory permits fitting quantile regression models to count data y_i transformed by adding a Uniform(0,1) random variable, u_i such that $z_i = \log(y_i + u_i - \tau)$. When $y_i = 0$ and $u_i < \tau, z_i$ is undefined. This result is not pertinent if there are no zeros in the data, as in Lee and Neocleous (2010). When zeros exist, however, the investigator must artificially select $0 < \eta$, and the selected value of η potentially affects inference (Appendix B). The Bayesian approach at least allows us to integrate over the potential range of values of η by composition sampling. We recommend that, if investigators apply Bayesian quantile regression to count data containing zeros, they examine the sensitivity of choosing η and how it affects inference, as in Appendix B.

9. Conclusion

Box (1979) stated "now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations". Nearly all data in nature are asymmetric. We present parametric models to estimate and address asymmetry in natural observation. We have shown by doing so results in better predictive models of natural processes. We conclude that Bayesian asymmetric regression is an essential component of a scientist's statistical toolbox to be used regularly alongside Gaussian, Laplace, and Poisson regression models.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

Joseph Robb, Brian Winters, Jason Lewis, Teresa Vanosdol and staff at Big Oaks National Wildlife Refuge, U.S. Fish and Wildlife Service collected and provided data on Henslow's Sparrow counts and burn histories. Kevin Shoemaker and Paul Conn provided valuable comments and insights about this work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, athttps://doi.org/10.1016/j.ecoinf.2023.102085.

References

Abrevaya, J., Huang, J., 2005. On the bootstrap of the maximum score estimator. Econometrica 73 (4), 1175–1204.
Benoit, D.F., Van den Poel, D., 2017. bayesqr: A bayesian approach to quantile regression. J. Stat. Softw. 76 (i07), 1–32.
Berliner, L., 1996. Hierarchical Bayesian time series models. In: Hanson, K., Silver, R. (Eds.), Maximum Entropy and Bayesian Methods. Kluwer Academic Publishers, pp. 15–22.
Box, G.E., 1976. Science and statistics. J. Am. Stat. Assoc. 71 (356), 791–799.
Box, G.E., 1979. Robustness in the strategy of scientific model building. In: Robustness in statistics. Elsevier, pp. 201–236.
Cade, B.S., Noon, B.R., 2003. A gentle introduction to quantile regression for ecologists. Front. Ecol. Environ. 1 (8), 412–420.
Cade, B.S., Terrell, J.W., Schroeder, R.L., 1999. Estimating effects of limiting factors with regression quantiles. Ecology 80 (1), 311–323.
Clark, J.S., 2003. Uncertainty and variability in demography and population growth: a hierarchical approach. Ecology 84 (6), 1370–1381.
Conn, P.B., Johnson, D.S., Williams, P.J., Melin, S.R., Hooten, M.B., 2018. A guide to bayesian model checking for ecologists. Ecol. Monogr. 88, 526–542.
Cressie, N., Calder, C.A., Clark, J.S., Hoef, J.M.V., Wikle, C.K., 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecol. Appl. 19 (3), 553–570.
Florios, K., Skouras, S., 2008. Exact computation of max weighted score estimators. J. Econom. 146 (1), 86–91.
Fowler, N., 1990. The 10 most common statistical errors. Bull. Ecol. Soc. Am. 71 (3), 161–164.
Gilchrist, W., 2000. Statistical modelling with quantile functions. Chapman and Hall/CRC.
Hefley, T.J., Broms, K.M., Brost, B.M., Buderman, F.E., Kay, S.L., Scharf, H.R., Tipton, J. R., Williams, P.J., Hooten, M.B., 2017. The basis function approach for modeling autocorrelation in ecological data. Ecology 98 (3), 632–646.
Herkert, J.R., Vickery, P.D., Kroodsma, D.E., 2002. Henslow's sparrow: Ammodramus henslowii. In: Poole, A. (Ed.), The Birds of North America. Cornell Laboratory of Arnithology, Ithaca. Retrieved August 21, 2018.
Hobbs, N.T., Hooten, M.B., 2015. Bayesian Models: A Statistical Primer for Ecologists. Princeton University Press.
Hooten, M.B., Hefley, T.J., 2019. Bringing Bayesian Models to Life. CRC Press.
Hooten, M.B., Hobbs, N., 2015. A guide to bayesian model selection for ecologists. Ecol. Monogr. 85 (1), 3–28.
Huber, P., 1981. Robust Statistics. John Wiley, New York, New York, USA.
Johnson, N.L., Kemp, A.W., Kotz, S., 2005. Univariate Discrete Distributions. John Wiley, New York, New York, USA.
Koenker, R., Bassett, G., 1978. Regression quantiles. Econometrica 46 (1), 33–50.
Koenker, R., Machado, J.A., 1999. Goodness of fit and related inference processes for quantile regression. J. Am. Stat. Assoc. 94 (448), 1296–1310.
Koenker, R., Schorfheide, F., 1994. Quantile spline models for global temperature change. Clim. Change 28 (4), 395–404.
Komori, O., Eguchi, S., Ikeda, S., Okamura, H., Ichinokawa, M., Nakayama, S., 2016. An asymmetric logistic regression model for ecological data. Methods Ecol. Evol. 7 (2), 249–260.
Kozłowski, J., Gawelczyk, A.T., 2002. Why are species' body size distributions usually skewed to the right? Funct. Ecol. 16 (4), 419–432.
Lee, D., Neocleous, T., 2010. Bayesian quantile regression for count data with application to environmental epidemiology. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 59 (5), 905–920.
Leger, E.A., Atwater, D.Z., James, J.J., 2019. Seed and seedling traits have strong impacts on establishment of a perennial bunchgrass in invaded semi-arid systems. J. Appl. Ecol. 56 (6), 1343–1354.
Lindén, A., Mäntyniemi, S., 2011. Using the negative binomial distribution to model overdispersion in ecological count data. Ecology 92 (7), 1414–1421.
Machado, J.A.F., Silva, J.S., 2005. Quantiles for counts. J. Am. Stat. Assoc. 100 (472), 1226–1237.
Nelder, J.A., Wedderburn, R.W.M., 1972. Generalized linear models. J. R. Stat. Soc. Ser. A (Gen.) 135, 370–384.
Park, T., Casella, G., 2008. The bayesian lasso. J. Am. Stat. Assoc. 103 (482), 681–686.
Scharf, F.S., Juanes, F., Sutherland, M., 1998. Inferring ecological relationships from the edges of scatter diagrams: comparison of regression techniques. Ecology 79 (2), 448–460.
Scharf, H.R., Lu, X., Williams, P.J., Hooten, M.B., 2022. Constructing flexible, identifiable and interpretable statistical models for binary data. Int. Stat. Rev. 90 (2), 328–345.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. J. R. Stat. Soc.: Ser. B (Stat. Methodol.) 64 (4), 583–639.
Sriram, K., Ramamoorthi, R., Ghosh, P., et al., 2013. Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. Bayesian Anal. 8 (2), 479–504.
Stevens, W., 1950. Fiducial limits of the parameter of a discontinuous distribution. Biometrika 37 (1/2), 117–129.
Ver Hoef, J.M., Boveng, P.L., 2007. Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? Ecology 88 (11), 2766–2772.
Waldmann, E., Sobotka, F., Kneib, T., 2017. Bayesian regularisation in geoadditive expectile regression. Stat. Comput. 27 (6), 1539–1553.
Watanabe, S., 2013. A widely applicable bayesian information criterion. J. Mach. Learn. Res. 14 (Mar), 867–897.
Williams, P.J., Hooten, M.B., 2016. Combining statistical inference and decisions in ecology. Ecol. Appl. 26 (6), 1930–1942.
Williams, P.J., Hooten, M.B., Womble, J.N., Esslinger, G.G., Bower, M.R., 2018. Monitoring dynamic spatio-temporal ecological processes optimally. Ecology 99 (3), 524–535.
Winship, A.J., Jorgensen, S.J., Shaffer, S.A., Jonsen, I.D., Robinson, P.W., Costa, D.P., Block, B.A., 2012. State-space framework for estimating measurement error from double-tagging telemetry experiments. Methods Ecol. Evol. 3 (2), 291–302.
Wool, D., et al., 1980. On ecological inference from kurtosis and skewness of morphological characters. Res. Popul. Ecol. 22 (2), 263–272.
Yang, Y., Wang, H.J., He, X., 2016. Posterior inference in Bayesian quantile regression with asymmetric Laplace likelihood. Int. Stat. Rev. 84 (3), 327–344.
Yu, K., Moyeed, R.A., 2001. Bayesian quantile regression. Stat. Probab. Lett. 54 (4), 437–447.
Yu, K., Zhang, J., 2005. A three-parameter asymmetric laplace distribution and its extension. Commun. Stat.–Theory Methods 34 (9–10), 1867–1879.