Count Transformation Models: The cotram Package

Sandra Siegfried and **Torsten Hothorn**
Universität Zürich

Abstract

The **cotram** package offers a ready-to-use R implementation of count transformation models, providing a simple but flexible approach for the regression analysis of count responses arising from various, and possibly complex, data-generating processes. In this unified maximum-likelihood framework count models can be formulated, estimated, and evaluated easily. Specific models in the class can be flexibly customised by the choice of the link function and the parameterisation of the transformation function. Interpretation of explanatory variables in the linear predictor is possible at the scales of the discrete odds ratio, hazard ratio, or reverse time hazard ratio, or as conditional mean of transformed counts. The implemented methods for the model class further provide simple tools for model evaluation. The package simplifies the use of transformation models for modelling counts, while ensuring appropriate settings for count data specifically. Extension to the formulated models can be made by the inclusion of response-varying effects, strata-specific transformation functions, or offsets, based on the underlying infrastructure of the **tram** and **mlt** R add-on packages, which further ensure the correct handling of the likelihood for censored or truncated observations.

Keywords: conditional distribution function, conditional quantile function, count regression, deer-vehicle collisions, transformation model.

1. Introduction

Count transformation models are a novel model class, offering a flexible and data-driven approach to regressing count data. The diverse set of models in the class, as proposed and discussed in Siegfried and Hothorn (2019), are tailored to analyse count responses from various underlying data-generating processes in a unified maximum-likelihood framework. The R add-on package **cotram** features the implementation of the proposed model class, providing a simple and user-friendly interface to fit and evaluate count transformation models. The package is built using the general infrastructure of the R add-on packages **tram** (Hothorn 2019b) and **mlt** (Hothorn 2018, 2019a) for likelihood-based inference and further extensions to the implemented model specifications.

Count transformation models arise from the direct modelling of the conditional discrete distribution function capturing changes governed by a linear predictor $\boldsymbol{x}^\top\boldsymbol{\beta}$. The models in the class can be represented by the general formulation of the conditional distribution function for any y

$$F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y \mid \boldsymbol{x}) = \mathbb{P}(Y \leq y \mid \boldsymbol{x}) = F_Z\left(h_Y\left(\lfloor y \rfloor\right) - \boldsymbol{x}^\top\boldsymbol{\beta}\right), \quad y \in \mathbb{R}^+ \tag{1}$$

with specific models originating from the choice of the different link functions F_Z^{-1}. The model class includes models with a logit, complementary log-log (cloglog), log-log, and probit link and thus offers interpretability of the linear predictor at various scales. The framework allows evaluating and interpreting the models in a discrete way, while using a computationally attractive, low-dimensional, continuous representation. The models are designed to simultaneously estimate the transformation function h_Y and the regression coefficients $\boldsymbol{\beta}$ optimising the exact discrete log-likelihood. Simultaneous estimation of the parameters (developed by Hothorn *et al.* 2018) is performed based on the underlying infrastructure provided by the **mlt** package (Hothorn 2019a).

All models in the class (1) can be fitted using the general function call

```
R> cotram(<formula>, method = <link>, ...)
```

with `<formula>` being any R formula featuring counts as the response and the right hand side as series of terms determining a linear predictor. The specific models in the class can be fitted by choosing one of the link functions for `method = <link>`. The set of models specified by the different link functions and the interpretation of the explanatory variables in the linear predictor $\boldsymbol{x}^\top\boldsymbol{\beta}$ are outlined in more detail below.

The package further offers `predict()` and `plot()` functions to assess and illustrate the estimated linear predictor, conditional distribution and density function, quantiles and the estimated transformation function, both as step-functions and continuously (setting `smooth = TRUE`). Functionalities for model interpretation and evaluation, such as `summary()`, `coef()`, `confint()`, and `logLik()` are available in this framework.

2. Discrete Hazards Cox Count Transformation Model

The count transformation model with complementary log-log link function F_Z^{-1} (`method = "cloglog"`) offers a discrete version of the Cox proportional hazards model with fully parameterised transformation function h_Y and interpretation of the linear predictor as discrete hazard ratio. The model explains the effects of the exponentiated linear predictor $\exp(-\boldsymbol{x}^\top\boldsymbol{\beta})$ on observed counts as multiplicative changes in discrete hazards $\mathbb{P}(Y = y \mid Y \geq y, \boldsymbol{x})$, comparing the conditional cumulative hazard function $\log(1 - F_{Y|\boldsymbol{X}=\boldsymbol{x}})$ with the baseline cumulative hazard function $\log(1 - F_Y)$, with $\boldsymbol{x}^\top\boldsymbol{\beta} = 0$.

Using the deer-vehicle collisions data from Hothorn *et al.* (2015), we can fit the Cox count transformation model to the roe deer-vehicle collision counts per day, recorded from 2002 to 2011 in Bavaria, Germany, and obtain the estimated multiplicative temporal changes in "risk" as discrete hazards. The `tvar` variables are sin-cosine transformed times (see Hothorn *et al.* 2015).

```
R> mod_cloglog <- cotram(DVC ~ year + weekday + tvar1 + tvar2 + tvar3 +
+                        tvar4 + tvar5 + tvar6 + tvar7 + tvar8 + tvar9 +
+                        tvar10 + tvar11 + tvar12 + tvar13 + tvar14 +
+                        tvar15 + tvar16 + tvar17 + tvar18 + tvar19 + tvar20,
+                        data = df, method = "cloglog")
R> logLik(mod_cloglog)
```

```
'log Lik.' -16545.5 (df=42)
```

To assess how the risk varies across days and seasons, we can now compute the estimated discrete hazards ratio for each day of the year, based on the predictor values of the year 2011. The results, shown in Figure 1, illustrate the changes in the hazard ratios, relative to baseline on January 1st (note that we plot $\exp(\boldsymbol{x}(\text{day})^\top \boldsymbol{\beta} - \boldsymbol{x}(2011\text{–}01\text{–}01)^\top \boldsymbol{\beta})$, such that large values correspond to large number of collisions and thus higher risk).

```
R> nd <- model.frame(mod_cloglog)[which(df$year == "2011"), -1]
R> nd$day <- df[which(df$year == "2011"), "day"]
R> nd$weekday <- factor("Monday", levels = levels(nd$weekday))
```

```
R> fit_cloglog <- predict(mod_cloglog, type = "lp", newdata = nd) -
+    predict(mod_cloglog, type = "lp", newdata = nd)[1]
R> xyplot(exp(fit_cloglog) ~ day , data = cbind(nd, fit_cloglog),
+         ylab = "Hazard ratio", xlab = "Day of year", panel = panel)
```

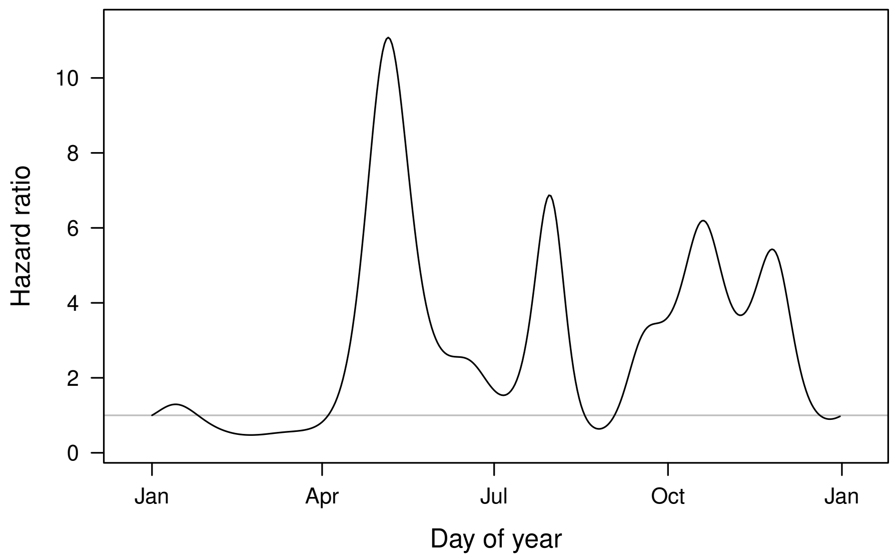

Figure 1: Deer-vehicle collisions. Temporal changes in risk for deer-vehicle collisions across the year as discrete hazard ratios estimated by model `mod_cloglog` with reference: January 1st. The curve indicates, that the hazard ratio is increased associated with animal activity due to search for new habitats and food resources in April and rut season in July and August. The peak in October does not seem to have a clear explanation in terms of increased roe deer activity.

3. Logistic Count Transformation Model

Odds ratios are often used in practice to compare two different configurations of the set of explanatory variables $\boldsymbol{x}$. Conveniently, for the class of count transformation models we can obtain the estimated effects on this scale by specifying a logit link. The exponentiated linear predictor $\exp(-\boldsymbol{x}^\top\boldsymbol{\beta})$ estimated by such a logistic count transformation model can be interpreted as odds ratio

$$\frac{\mathbb{P}(Y \le y \mid \boldsymbol{x})}{\mathbb{P}(Y > y \mid \boldsymbol{x})} = \frac{\mathbb{P}(Y \le y)}{\mathbb{P}(Y > y)} \exp(-\boldsymbol{x}^\top\boldsymbol{\beta}),$$

comparing the conditional odds of a configuration $\boldsymbol{x}$ with the baseline odds $F_Y/1-F_Y$ (with $\boldsymbol{x}^\top\boldsymbol{\beta} = 0$). The response-varying intercept $h_Y(y)$ cancels out in the odds ratio, resulting in an estimate, which can be interpreted simultaneously across all cut-offs y.

To explain the temporal risk of roe deer-vehicle collisions on the odds ratio scale, the only modification to the model formulation of Section 2 required, is the link specification in the function call as `method = "logit"`.

```
R> mod_logit <- cotram(DVC ~ year + weekday + tvar1 + tvar2 + tvar3 +
+                         tvar4 + tvar5 + tvar6 + tvar7 + tvar8 + tvar9 +
+                         tvar10 + tvar11 + tvar12 + tvar13 + tvar14 +
+                         tvar15 + tvar16 + tvar17 + tvar18 + tvar19 + tvar20,
+                         data = df, method = "logit")
R> logLik(mod_logit)

'log Lik.' -16319.29 (df=42)
```

Comparison of the log-likelihoods of the fitted model and the Cox count transformation model from Section 2 shows almost matching values, with a slight improvement in model fit, when replacing the cloglog with the logit link.

We now could further assess the effect of the factor `year` on the deer-vehicle collision counts by computing the odds ratios (small values correspond to moving the distribution to the right and thus to larger number of collisions) along with the likelihood-based confidence intervals.

```
R> years <- grep("year", names(coef(mod_logit)), value = TRUE)
R> coef <- exp(-coef(mod_logit)[years])
R> ci <- exp(-confint(mod_logit)[years,])
R> round(cbind(coef, ci), 3)

          coef 2.5 % 97.5 %
year2003 0.595 0.765  0.463
year2004 0.337 0.433  0.262
year2005 0.305 0.393  0.237
year2006 0.406 0.525  0.314
year2007 0.156 0.203  0.120
year2008 0.096 0.123  0.074
year2009 0.104 0.135  0.081
year2010 0.090 0.116  0.069
year2011 0.097 0.125  0.075
```

Plotting the estimated conditional distribution functions of model `mod_logit` in Figure 2, demonstrates the linear shift in $F_{Y|\boldsymbol{X}=\boldsymbol{x}}$ guided by the different levels of the factor `year`.

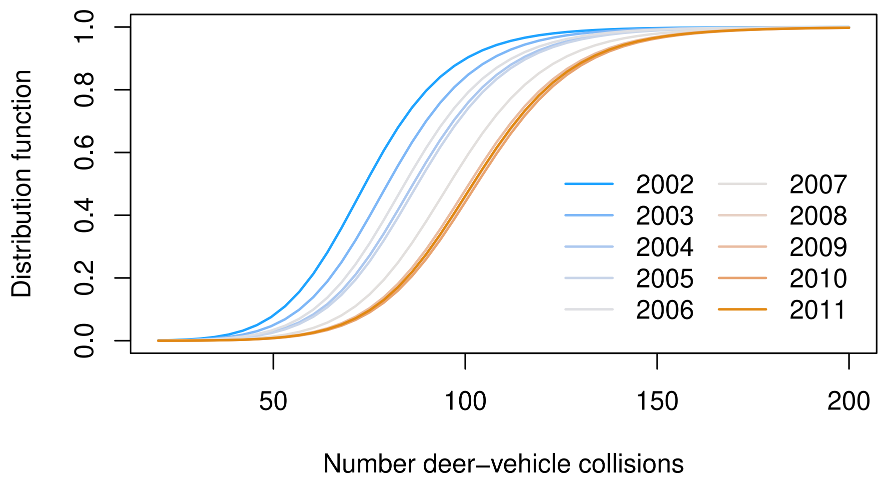

Figure 2: Deer-vehicle collisions. Illustration of the estimated conditional distribution functions of each year between 2002 and 2011.

4. Discrete Reverse Time Hazards Count Transformation Model

Specifying a count transformation model with log-log link F_Z^{-1} we get the model formulation

$$F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y \mid \boldsymbol{x}) = \mathbb{P}(Y \leq y \mid \boldsymbol{x}) = \exp\left(-\exp\left(h_Y(\lfloor y \rfloor) - \boldsymbol{x}^\top \boldsymbol{\beta}\right)\right)$$

with interpretation of the linear predictor $\exp(\boldsymbol{x}^\top \boldsymbol{\beta})$ as discrete reverse hazard ratio with multiplicative changes in $\log(F_Y)$. To fit the model, we again only need to adapt the model specification in terms of the link function by setting `method = "loglog"`.

```
R> mod_loglog <- cotram(DVC ~ year + weekday + tvar1 + tvar2 + tvar3 +
+                       tvar4 + tvar5 + tvar6 + tvar7 + tvar8 + tvar9 +
+                       tvar10 + tvar11 + tvar12 + tvar13 + tvar14 +
+                       tvar15 + tvar16 + tvar17 + tvar18 + tvar19 + tvar20,
+                       data = df, method = "loglog")
R> logLik(mod_loglog)

'log Lik.' -16438.23 (df=42)
```

For further assessment we could evaluate the discrete conditional density of a set of $\boldsymbol{x}$. Figure 3 illustrates the estimated density function in terms of the predictor values recorded on 2002–01–01 along with the actually observed deer-vehicle collision count.

```
R> nd <- model.frame(mod_loglog)[1,]
```

```
R> plot(mod_loglog, type = "density", newdata = nd, q = 0:150, col = col,
+       xlab = "Number of deer-vehicle collisions", ylab = "Density function")
R> abline(v = nd$DVC)
```

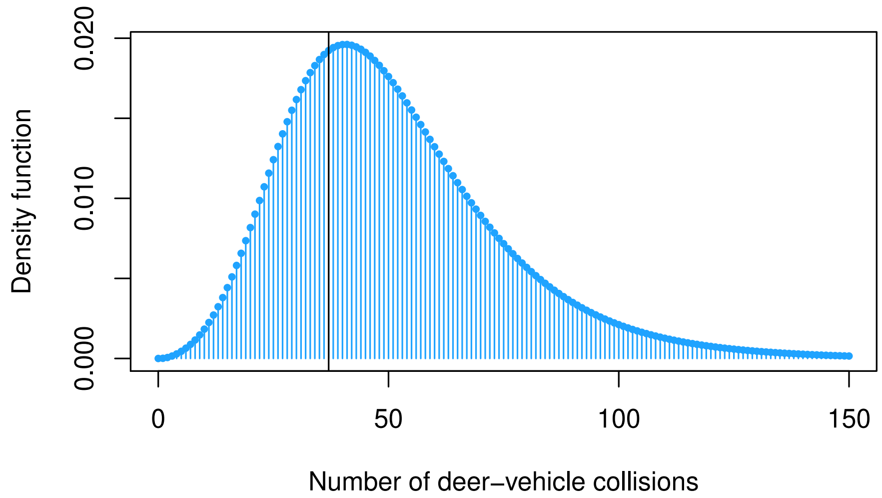

Figure 3: Deer-vehicle collisions. Estimated discrete density function for model `mod_loglog` with the actual observed count shown as vertical black line.

5. Probit Count Transformation Model

When applying a count transformation model with a probit link (`method = "probit"`) we can interpret the estimated effects as changes in the conditional mean of the transformed counts $\mathbb{E}(h_Y(y) \mid \boldsymbol{X} = \boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{\beta}$. This interpretation is the same, as obtained from fitting a normal linear regression model on a priori transformed counts, by *e.g.* a log or square-root transformation. However, for the probit count transformation model, as implemented in the **cotram** package, the transformation of the response y was not heuristically chosen, as in a least-squares approach, but estimated from data by optimising the exact count log-likelihood.

```
R> mod_probit <- cotram(DVC ~ year + weekday + tvar1 + tvar2 + tvar3 +
+                        tvar4 + tvar5 + tvar6 + tvar7 + tvar8 + tvar9 +
+                        tvar10 + tvar11 + tvar12 + tvar13 + tvar14 +
+                        tvar15 + tvar16 + tvar17 + tvar18 + tvar19 + tvar20,
+                        data = df, method = "probit")
R> logLik(mod_probit)

'log Lik.' -16310.32 (df=42)
```

A simple tool in this framework to check, whether, for example a log transformation, would have been appropriate, is to inspect the estimated transformation function $h_Y(y)$ as illustrated in Figure 4.

```
R> plot(mod_probit, type = "trafo", newdata = df[1,], smooth = TRUE,
+       xlab = "Number of deer-vehicle collisions",
+       ylab = expression(paste("Transformation function ", h[Y](y))),
+       col = col[10], lwd = 2)
```

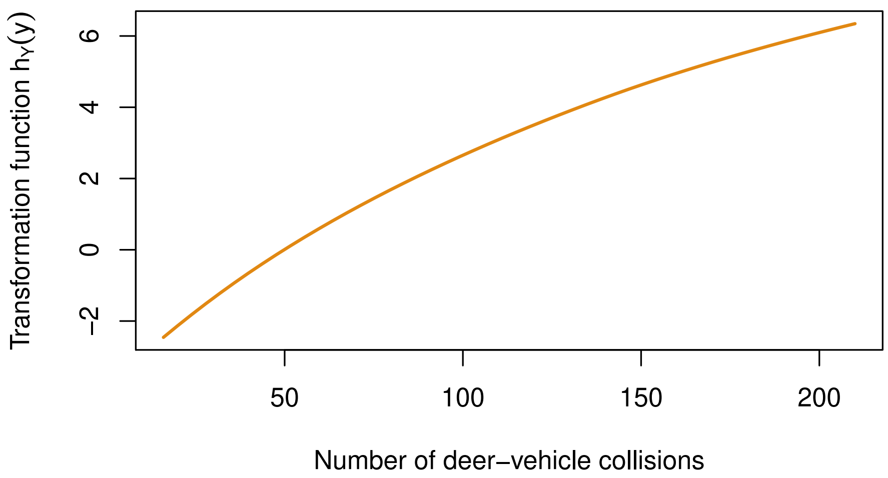

Figure 4: Deer-vehicle collisions. Baseline transformation h_Y estimated by the model `mod_probit` illustrated as smoothed line.

6. Summary

The implemented models and methods in the **cotram** package offer a unified framework for users to fit and evaluate transformation models for counts, by ensuring the correct handling of the discrete nature of the data. Simplifying the modelling procedure, the models are parameterised under general and empirically tested settings, eliminating the need for overly complicated model specifications.

References

Hothorn T (2018). "Most Likely Transformations: The **mlt** Package." *Journal of Statistical Software*. Accepted for publication 2018-03-05, URL https://cran.r-project.org/web/packages/mlt.docreg/vignettes/mlt.pdf.

Hothorn T (2019a). ***mlt**: Most Likely Transformations*. R package version 1.0-5, URL https://CRAN.R-project.org/package=mlt.

Hothorn T (2019b). ***tram**: Transformation Models*. R package version 0.2-6, URL https://CRAN.R-project.org/package=tram.

Hothorn T, Möst L, Bühlmann P (2018). "Most Likely Transformations." *Scandinavian Journal of Statistics*, **45**(1), 110–134. doi:10.1111/sjos.12291.

Hothorn T, Müller J, Held L, Möst L, Mysterud A (2015). "Temporal Patterns of Deer-vehicle Collisions Consistent with Deer Activity Pattern and Density Increase but not General Accident Risk." *Accident Analysis & Prevention*, **81**, 143–152. doi:10.1016/j.aap.2015.04.037.

Siegfried S, Hothorn T (2019). "Count Transformation Models." Submitted manuscript (see below).

Count Transformation Models

Sandra Siegfried[1] and Torsten Hothorn[1]

[1] Institut für Epidemiologie, Biostatistik und Prävention,

Universität Zürich, Hirschengraben 84, CH-8001 Zürich,

Switzerland

Number of words abstract: 350

Number of words text: 3738

Number of figures: 5

Number of tables: 1

Number of references: 26

Number of online appendices: 1

Abstract

1. The effect of explanatory environmental variables on a species' distribution is often assessed using a count regression model. Poisson generalised linear models or negative binomial models are common, but the traditional approach of modelling the mean after log or square-root transformation remains popular and in some cases is even advocated.

2. We propose a novel class of linear models for count data. Similar to the traditional approach, the new models apply a transformation to count responses; however, this transformation is estimated from the data and not defined a priori. In contrast to simple least-squares fitting and in line with Poisson or negative binomial models, the exact discrete likelihood is optimised for parameter estimation and inference. Interpretation of linear predictors is possible at various scales depending on the model formulation.

3. Count transformation models provide a new approach to regressing count data in a distribution-free yet fully parametric fashion, obviating the need to a priori commit to a specific parametric family of distributions or to a specific transformation. The model class is a generalisation of discrete Weibull models for counts and is thus able to handle over- and underdispersion. We demon-

strate empirically that the models are more flexible than Poisson or negative binomial models but still maintain interpretability of multiplicative effects. A re-analysis of deer-vehicle collisions and the results of artificial simulation experiments provide evidence of the practical applicability of the model class.

4. In ecology studies, uncertainties regarding whether and how to transform count data can be resolved in the framework of count transformation models, which were designed to simultaneously estimate an appropriate transformation and the linear effects of environmental variables by maximising the exact count log-likelihood. The application of data-driven transformations allows over- and underdispersion to be addressed in a model-based approach. Competing models in this class can be compared to Poisson or negative binomial models using the in- or out-of-sample log-likelihood. Extensions to non-linear additive or interaction effects, correlated observations, hurdle-type models and other, more complex situations are possible. A free software implementation is available in the **cotram** add-on package to the R system for statistical computing.

Keywords conditional distribution function, conditional quantile function, count regression, deer-vehicle collisions, transformation model;

1 Introduction

Information represented by counts is ubiquitous in ecology. Perhaps the most obvious instance of ecological count data is animal abundances, which are determined either directly, for example by birdwatchers, or indirectly, by the counting of surrogates, for example the number of deer-vehicle collisions as a proxy for roe deer abundance. This information is later converted into models of animal densities or species distributions using statistical models for count data. Distributions of count data are, of course, discrete and right-skewed, such that tailored statistical models are required for data analysis. Here we focus on models explaining the impact of explanatory environmental variables $\boldsymbol{x}$ on the distribution of a count response $Y \in \{0, 1, 2, \dots\}$. In the commonly used Poisson generalised linear model $Y \mid \boldsymbol{x} \sim \text{Po}(\exp(\alpha + \boldsymbol{x}^\top \boldsymbol{\beta}))$ with log-link, intercept α and linear predictor $\boldsymbol{x}^\top \boldsymbol{\beta}$, both the mean $\mathbb{E}(Y \mid \boldsymbol{x})$ and the variance $\mathbb{V}(Y \mid \boldsymbol{x})$ of the count response are given by $\exp(\alpha + \boldsymbol{x}^\top \boldsymbol{\beta})$. Overdispersion, *i.e.* the situation $\mathbb{E}(Y \mid \boldsymbol{x}) < \mathbb{V}(Y \mid \boldsymbol{x})$, is allowed in the more complex negative binomial model $Y \mid \boldsymbol{x} \sim \text{NB}(\exp(\alpha + \boldsymbol{x}^\top \boldsymbol{\beta}), \nu)$ with mean $\mathbb{E}(Y \mid \boldsymbol{x}) = \exp(\alpha + \boldsymbol{x}^\top \boldsymbol{\beta})$ and potentially larger variance $\mathbb{V}(Y \mid \boldsymbol{x}) = \mathbb{E}(Y \mid \boldsymbol{x}) + \mathbb{E}(Y \mid \boldsymbol{x})^2/\nu$. For independent observations, the model parameters are obtained by maximising the discrete log-likelihood function, in which an observation $(y, \boldsymbol{x})$ contributes the log-density $\log(\mathbb{P}(Y = y \mid \boldsymbol{x}))$ of either the

Poisson or the negative binomial distribution.

Before the emergence of these models tailored to the analysis of count data (generalised linear models were introduced by Nelder & Wedderburn 1972), researchers were restricted to analysing transformations of Y by normal linear regression models. Prominent textbooks at the time (Snedecor & Cochran 1967; Sokal & Rohlf 1967) suggested log transformations $\log(y+1)$ or square-root transformations $\sqrt{y+0.5}$ of observed counts y. The application of least-squares estimators to the log-transformed counts then leads to the mean $\mathbb{E}(\log(y+1) \mid \boldsymbol{x}) = \alpha + \boldsymbol{x}^\top\boldsymbol{\beta}$. Implicitly, it is assumed that the variance after transformation $\mathbb{V}(\log(y+1) \mid \boldsymbol{x}) = \sigma^2$ is constant and that errors are normally distributed. Although it is clear that the normal assumption $\log(Y+1) \mid \boldsymbol{x} \sim \mathrm{N}(\alpha+\boldsymbol{x}^\top\boldsymbol{\beta}, \sigma^2)$ is incorrect (the count data are still discrete after transformation) and, consequently, that the wrong likelihood is maximised by applying least-squares to $\log(y+1)$ for parameter estimation and inference, this approach is still broadly used both in practice and in theory (*e.g.* Ives 2015; Dean, Voss & Draguljić 2017; Gotelli & Ellison 2013; De Felipe, Sáez-Gómez & Camacho 2019; Mooney, Phillips, Tillberg, Sandrow, Nelson & Mooney 2016). Moreover, other deficits of this approach have been discussed in numerous papers (*e.g.* O'Hara & Kotze 2010; Warton, Lyons, Stoklosa & Ives 2016; St-Pierre, Shikon & Schneider 2018; Warton 2018).

As a compromise between the two extremes of using rather strict count distribution models (such as the Poisson or negative binomial) and the analysis of transformed counts by normal linear regression models, we suggest a novel class of transformation models for count data that combines the strengths of both approaches. Briefly stated, in the newly proposed method appropriate transformations of counts Y are estimated simultaneously with regression coefficients $\boldsymbol{\beta}$ from the data by maximising the correct discrete form of the likelihood in models that ensure the interpretability of a linear predictor $\boldsymbol{x}^\top\boldsymbol{\beta}$ on an appropriate scale. We describe the theoretical foundations of these novel count regression models in Section 2. Practical aspects of the methodology are demonstrated in Section 3 in a re-analysis of roe deer activity patterns based on deer-vehicle collision data, followed by an artificial simulation experiment contrasting the performance of Poisson, negative binomial and count transformation models under certain conditions.

2 Methods

The core idea of our count transformation model for describing the impact of explanatory environmental variables $\boldsymbol{x}$ on counts $Y \in \{0, 1, 2, \dots\}$ is the simultaneous estimation of a fully parameterised smooth transformation $h_Y(Y)$ of the discrete response and the regression coefficients in a linear predictor $\boldsymbol{x}^\top\boldsymbol{\beta}$. The aim of the approach is to model the discrete conditional distribution function $F_{Y|\boldsymbol{X}=\boldsymbol{x}}$ directly. Specifically, for any positive real number y we evaluate the conditional distribution function as

$$F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y \mid \boldsymbol{x}) = \mathbb{P}(Y \leq y \mid \boldsymbol{x}) = F_Z\left(h_Y\left(\lfloor y \rfloor\right) - \boldsymbol{x}^\top\boldsymbol{\beta}\right), \quad y \in \mathbb{R}^+ \tag{1}$$

with $h_Y : \mathbb{R}^+ \to \mathbb{R}$ being an unknown, montonically increasing continuous transformation function applied to the greatest integer $\lfloor y \rfloor$ less than or equal to y. Specific models in this class arise from the different a priori choices of the inverse link function $F_Z : \mathbb{R} \to [0, 1]$ and the parameterisation of h_Y. Hothorn, Möst & Bühlmann (2018) suggested the parameterisation of h_Y in terms of basis functions $\boldsymbol{a} : \mathbb{R} \to \mathbb{R}^P$ and the corresponding parameters $\boldsymbol{\vartheta}$ as

$$h_Y(y) = \boldsymbol{a}(y)^\top\boldsymbol{\vartheta}.$$

The only modification required for count data is to consider this transformation function as a step function with jumps at integers $0, 1, 2, \ldots$ only. This is achieved in model (1) by the floor function $\lfloor y \rfloor$. The very same approach was suggested by Padellini & Rue (2019) but to model quantile functions $F^{-1}_{Y|\boldsymbol{X}=\boldsymbol{x}}$ of count data instead of the distribution functions we consider here. Figure 1 shows a distribution function $F_Y(y) = F_Z(h_Y(\lfloor y \rfloor))$ and the corresponding transformation function h_Y, both as discrete step-functions (flooring the argument first) and continuously (without doing so). The two versions are identical for integer-valued arguments. Thus, the transformation function h_Y, and consequently the transformation model (1), are parameterised continuously but evaluated and interpreted discretely. A computationally attractive, low-dimensional representation of a smooth function in terms of a few basis functions $\boldsymbol{a}$ and corresponding parameters is therefore the core ingredient of our novel model class.

[Figure 1 about here.]

On a more technical level, the basis $\boldsymbol{a}$ is specified in terms of $\boldsymbol{a}_{\text{Bs},P-1}$, with P-dimensional basis functions of a Bernstein polynomial (Farouki 2012) of order $P-1$. Specifically, the basis $\boldsymbol{a}(y)$ can be chosen as: $\boldsymbol{a}_{\text{Bs},P-1}(y)$ or $\boldsymbol{a}_{\text{Bs},P-1}(y+1)$, or as a Bernstein polynomial on the log-scale: $\boldsymbol{a}_{\text{Bs},P-1}(\log(y))$ or $\boldsymbol{a}_{\text{Bs},P-1}(\log(y+1))$. The choice of $\boldsymbol{a}(y) = \boldsymbol{a}_{\text{Bs},P-1}(\log(y+1))$ is particularly well suited for modelling relatively small counts. For $P = 1$, the defined basis is equivalent to a linear function of either y, $\log(y)$ or $\log(y+1)$. Monotonicity of the transformation function h_Y can be obtained under the constraint $\vartheta_1 \leq \vartheta_2 \leq \cdots \leq \vartheta_P$ of the parameters $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_P)^\top \in \mathbb{R}^P$ (Hothorn et al. 2018).

The monotonically increasing continuous inverse link function $F_Z : \mathbb{R} \to [0, 1]$ governs the interpretation of the linear predictor $\boldsymbol{x}^\top\boldsymbol{\beta}$. The conditional distribution function $F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y \mid \boldsymbol{x})$ for different choices of the link function F_Z^{-1} and any configuration $\boldsymbol{x}$ are given in Table 1, with $F_Y(y) = F_Z(h_Y(\lfloor y \rfloor))$ denoting the distribution of the baseline configuration $\boldsymbol{x}^\top\boldsymbol{\beta} = 0$. Note that, with a sufficiently flexible parameterisation of the transformation function $h(y) = \boldsymbol{a}(y)^\top \boldsymbol{\vartheta}$, every distribution can be written in this way such that the model is distribution-free (Hothorn et al. 2018).

The parameters $\boldsymbol{\beta}$ describe a deviation from this baseline distribution in terms of the linear predictor $\boldsymbol{x}^\top \boldsymbol{\beta}$. For a probit link, the linear predictor is the conditional mean of the transformed counts $h_Y(Y)$. This interpretation, except for the fact that the intercept α is understood as being part of the transformation function h_Y, is the same as in the traditional approach of first transforming the counts and only then estimating the mean using least-squares. However, the transformation h_Y is not heuristically chosen or defined a priori but estimated from data through parameters $\boldsymbol{\vartheta}$, as explained below. For a logit link, $\exp(-\boldsymbol{x}^\top \boldsymbol{\beta})$ is the odds ratio comparing the conditional odds ${}^{F_{Y|\boldsymbol{X}=\boldsymbol{x}}}/_{1-F_{Y|\boldsymbol{X}=\boldsymbol{x}}}$ with the baseline odds ${}^{F_Y}/_{1-F_Y}$. The complementary log-log (cloglog) link leads to a discrete version of the Cox proportional hazards model, such that $\exp(-\boldsymbol{x}^\top \boldsymbol{\beta})$ is the hazard ratio comparing the conditional cumulative hazard function $\log(1 - F_{Y|\boldsymbol{X}=\boldsymbol{x}})$ with the baseline cumulative hazard function $\log(1-F_Y)$. The log-log link leads to the reverse time hazard ratio with multiplicative changes in $\log(F_Y)$. All models in Table 1 are parameterised to relate positive values of $\boldsymbol{x}^\top \boldsymbol{\beta}$ to larger means independent of the specified link F_Z^{-1}.

[Table 1 about here.]

There is a very close connection between generalised linear models for binary data and our transformation model (1). For any dichotomisation of the counts $\mathbb{1}(Y \leq y)$, the generalised linear model

$$F_Z^{-1}\left(\mathbb{E}(\mathbb{1}(Y \leq y) \mid \boldsymbol{x})\right) = F_Z^{-1}\left(\mathbb{P}(\mathbb{1}(Y \leq y) \mid \boldsymbol{x})\right) = \alpha(y) - \boldsymbol{x}^\top \boldsymbol{\beta}$$

features an intercept $\alpha(y)$ that depends on the cut-off y while the regression coefficients $\boldsymbol{\beta}$ are treated as constant across all possible cut-off values $y \in \{0, 1, 2, \dots\}$. Our transformation model (1) arises from the choice $\alpha(y) = h_Y(y)$, and the transformation function can thus be interpreted as a response-varying intercept in binomial generalised linear models with different link functions F_Z^{-1}.

In Section 3.1 of our empirical evaluation we consider a linear count transformation model for discrete hazards by specifying the cloglog link. The discrete Cox count transformation model

$$\begin{aligned} F_{Y \mid \boldsymbol{X}=\boldsymbol{x}}(y \mid \boldsymbol{x}) &= \mathbb{P}(Y \leq y \mid \boldsymbol{x}) \\ &= 1 - \exp\left(-\exp\left(\boldsymbol{a}_{\mathrm{Bs},P-1}\left(\log(\lfloor y+1 \rfloor)\right)^\top \boldsymbol{\vartheta} - \boldsymbol{x}^\top \boldsymbol{\beta}\right)\right) \end{aligned} \tag{2}$$

with P Bernstein basis functions $\boldsymbol{a}_{\mathrm{Bs},P-1}$ relates positive linear predictors to smaller hazards and thus larger means. The discrete hazard function $\mathbb{P}(Y = y \mid Y \geq y, \boldsymbol{x})$ is the probability that y counts will be observed given that at least y counts were already observed. The model is equivalent to

$$\mathbb{P}(Y = y \mid Y \geq y, \boldsymbol{x}) = \exp(-\boldsymbol{x}^\top \boldsymbol{\beta})\mathbb{P}(Y = y \mid Y \geq y)$$

and thus the hazard ratio $\exp(-\boldsymbol{x}^\top \boldsymbol{\beta})$ gives the multiplicative change in discrete hazards.

The Cox proportional hazards model with a simplified transformation function $h_Y(y) = \vartheta_1 + \vartheta_2 \log(y + 1)$ specifies a discrete form of a Weibull model (introduced by Nakagawa & Osaki 1975) that Peluso, Vinciotti & Yu (2019) recently discussed as an extension to other count regression models and that serves as a more flexible approach for both over- and underdispersed data. The discrete Weibull model is a special form of our Cox count transformation model (2), as the former features a linear basis function $\boldsymbol{a}$ with $P = 2$ parameters defined by a Bernstein polynomial of order one. Thus, model (2) can be understood as a generalisation moving away from the low-parametric discrete Weibull distribution while maintaining both the interpretability of the effects as log-hazard ratios and the ability to handle over- and underdispersion.

Simultaneous likelihood-based inference for $\boldsymbol{\vartheta}$ and $\boldsymbol{\beta}$ for fully parameterised transformation models was developed by Hothorn et al. (2018); here we refer only to the most important aspects. The exact log-likelihood of the model for independent observations $(y_i, \boldsymbol{x}_i), i = 1, \ldots, N$ is given by the sum of the

N contributions

$$\ell_i(\boldsymbol{\vartheta}, \boldsymbol{\beta}) = \log(\mathbb{P}(Y = y_i \mid \boldsymbol{x}_i)) = \begin{cases} \log\left[F_Z\left\{\boldsymbol{a}(0)^\top\boldsymbol{\vartheta} - \boldsymbol{x}_i^\top\boldsymbol{\beta}\right\}\right] & y_i = 0 \\ \log\left[F_Z\left\{\boldsymbol{a}(y_i)^\top\boldsymbol{\vartheta} - \boldsymbol{x}_i^\top\boldsymbol{\beta}\right\} - F_Z\left\{\boldsymbol{a}(y_i - 1)^\top\boldsymbol{\vartheta} - \boldsymbol{x}_i^\top\boldsymbol{\beta}\right\}\right] & y_i > 0. \end{cases}$$

The corresponding log-likelihood is then maximised simultaneously with respect to both $\boldsymbol{\vartheta}$ and $\boldsymbol{\beta}$ under suitable constraints:

$$(\hat{\boldsymbol{\vartheta}}_N, \hat{\boldsymbol{\beta}}_N) = \operatorname*{arg\,max}_{\boldsymbol{\vartheta},\boldsymbol{\beta}} \sum_{i=1}^{N} \ell_i(\boldsymbol{\vartheta}, \boldsymbol{\beta}) \quad \text{subject to } \vartheta_p \leq \vartheta_{p+1}, p \in 1, \dots, P - 1.$$

Score functions and Hessians are available from Hothorn et al. (2018).

3 Results

In our empirical evaluation of the proposed count transformation models, we demonstrate practical aspects of the model class in Section 3.1, by re-analysing data on deer-vehicle collisions, and examine their properties in the context of conventional count regression models, assuming either a conditional Poisson or a negative binomial distribution. In Section 3.2, we use simulated count data to evaluate the robustness of count transformation models under model misspecification.

3.1 Analysis of deer-vehicle collision data

In the following, we re-analyse a time series of 341'655 deer-vehicle collisions involving roe deer (*Capreolus capreolus*) that were documented between 2002–01–01 and 2011–12–31 in Bavaria, Germany. The roe deer-vehicle collisions, recorded in 30-minute time intervals in the whole of Bavaria, were originally analysed by Hothorn, Müller, Held, Möst & Mysterud (2015) with the aim of describing temporal patterns in roe deer activity. The raw data and a detailed description of their analysis are available in the original study. In our re-analysis, we explore the estimates and properties of count regression models explaining how the risk of roe deer-vehicle collisions varies over days (diurnal effects) as well as across weeks, seasons and the whole year. We applied a Poisson generalised linear model with a log link, a negative binomial model with a log link and a discrete Cox count transformation model (2) with $P = 7$ parameters $\boldsymbol{\vartheta}$ of a Bernstein polynomial. The latter two models allow for possible overdispersion. The temporal changes in the risk of roe deer-vehicle collisions were modelled as a function of the following explanatory variables: annual, weekly and diurnal effects, an interaction of the weekly and diurnal effects, and seasonal effects, encoded as interactions of diurnal effects with a smooth seasonal component $s(d)$ (based on Held & Paul 2012). The three models were fitted to the data of the first eight years (2002 to

2009) and evaluated based on the data from the remaining two years, 2010 and 2011.

For each model we computed the estimated multiplicative seasonal changes in risk depending on the time of day relative to baseline on January 1st, including 95% simultaneous confidence bands. We interpreted "risk" as a multiplicative change to baseline with respect to either the conditional mean ("expectation ratio"; Poisson and negative binomial models) or the conditional discrete hazard function ("hazard ratio") for the Cox count transformation model (2).

[Figure 2 about here.]

The results in Figure 2 show a rather strong agreement between the three models with respect to the estimated risk (expectation ratio or hazard ratio). However, the uncertainty, assessed by the 95% confidence bands, was underestimated in the Poisson model. The negative binomial and the Cox count transformation model (2) agree on the effects and the associated variability, with the possible exception of the risk at daylight (Day, am).

To assess the performance of the three count regression models, we computed the out-of-sample log-likelihoods of each model based on the data of the validation sample (year 2010 and 2011). The out-of-sample log-likelihood of the Cox count transformation model (2) with a value of -58'164.47 was the largest across the three count regression models. The Poisson model, with an out-of-sample log-likelihood of $-67'192.75$, was the most inconsistent with the data. Allowing for possible overdispersion by the negative binomial model increased the out-of-sample log-likelihood to $-58'234.72$, which was closer to but did not match the out-of-sample log-likelihood of model (2).

We further compared the three different models in terms of their conditional distribution functions for four selected time intervals of the year 2009. The discrete conditional distribution functions of the models, evaluated for all integers between 0 and 38, are given in Figure 3. The conditional medians obtained from all three models are rather close, but the variability assessed by the Poisson model is much smaller than that associated with the negative binomial and count transformation models, thus indicating overdispersion.

[Figure 3 about here.]

3.2 Artificial count-data-generating processes

We investigated the performance of the different regression models in a simulation experiment based on count data from various underlying data-generating processes (DGPs). Count responses Y were generated conditionally on a numeric predictor variable $x \in [0, 1]$ following a Poisson or negative-binomial distribution or one of the discrete distributions underlying the four count transformation models corresponding to the four link functions from Table 1. For the Poisson model, the mean and variance were assumed to be $\mathbb{E}(Y \mid \boldsymbol{x}) = \mathbb{V}(Y \mid \boldsymbol{x}) = \exp(1.2 + 0.8\boldsymbol{x})$. The negative binomial data were chosen to be moderately overdispersed, with $\mathbb{E}(Y \mid \boldsymbol{x}) = \exp(1.2 + 0.8\boldsymbol{x})$ and $\mathbb{V}(Y \mid \boldsymbol{x}) = \mathbb{E}(Y \mid \boldsymbol{x}) + \mathbb{E}(Y \mid \boldsymbol{x})^2/3$. The four data-generating processes arising from the count transformation models were specified by the different link functions in Table 1, a Bernstein polynomial $\boldsymbol{a}_{\mathrm{Bs},6}(\log(y + 1))$ and a regression coefficient $\beta_1 = 0.8$.

We repeated the simulation experiment for each count-data-generating process 100 times, with learning and validation sample sizes of $N = 250$ and $\tilde{N} = 750$ respectively. The centred out-of-sample log-likelihoods, contrasting the model fit, were computed by the differences between the out-of-sample log-likelihoods of the models and the out-of-sample log-likelihoods of the true generating processes.

[Figure 4 about here.]

The results as given in Figure 4 follow a clear pattern. When misspecified, the model fit of the Poisson model is inferior to that of all other models. As expected, the negative-binomial model well fits both the data arising from the Poisson distribution (limiting case of the negative-binomial distribution with $\nu \to \infty$) and the moderately overdispersed data. However, it lacks robustness for more complex data-generating processes, such as the underlying mechanisms specified by a count transformation model. The fit of the count transformation models is satisfactory across all DGPs, albeit with some differences within the model class.

4 Discussion

Motivated by the challenges posed by the statistical analysis of ecological count data, we present a novel class of count transformation models that provide a unified approach tailored to the analysis of count responses. The model class, as outlined in Section 2, offers a diverse set of count models and can be specified, estimated and evaluated in a simple but flexible maximum likelihood framework. The direct modelling of the conditional discrete distribution, while preserving the interpretability of the linear predictor $\boldsymbol{x}^\top\boldsymbol{\beta}$, is a key feature of our count transformation model. Furthermore, it eliminates the need to impose restrictive distributional assumptions, to choose transformations in a data-free manner or to rely on rough approximations of the exact likelihood. The models are flexible enough to handle different dispersion levels adaptively, without being restricted to either over- or underdispersion. Our results from the re-analysis of deer-vehicle collision data, presented in Section 3.1, demonstrate the favourable properties of count transformations in practice. They are especially compelling for the analysis of count responses arising from more complex data-generating processes, for which the Poisson and even the more flexible negative binomial distribution are of limited use (as illustrated in Section 3.2). Moreover, conditional quantiles can be easily extracted from the fitted model by numerical inversion of the smooth conditional distribution function $F_Z(h_Y(y) - \boldsymbol{x}^\top\boldsymbol{\beta})$. An additional advantage of count transformation models is that the model class allows researchers to flexibly choose the scale of the interpretation of the linear predictor $\boldsymbol{x}^\top\boldsymbol{\beta}$ by specifying a link function F_Z^{-1} from Table 1.

The model class can be easily tailored to the experimental design using strata-specific transformation functions $h_Y(\lfloor y \rfloor \mid \text{strata})$ or response-varying effects $\boldsymbol{\beta}(\lfloor y \rfloor)$. Correlated observations arising from clustered data require the inclusion of random effects with subsequent application of a Laplace approximation to the likelihood. Accounting for varying observation times or batch sizes is straightforward by the inclusion of an offset in the model specification. Random censoring is easy to incorporate in the likelihood (Hothorn et al. 2018), which can then appropriately handle uncertain recordings (for example, the observation "more than three roe-deer vehicle collisions in half an hour" corresponds to right-censoring at three). The same applies to truncation. By contrast, hurdle-like transformation models require modifications of the basis functions as well as interactions between the response and explanatory variables (see Section 4.5 in Hothorn et al. 2018).

Extensions to the proposed simple shift count transformation model can be made by boosting algorithms (Hothorn 2019b) that allow the estimation of conditional transformation models (Hothorn, Kneib & Bühlmann 2014) featuring complex, non-linear, additive or completely unstructured tree-based conditional parameter functions $\boldsymbol{\vartheta}(\boldsymbol{x})$. Similarly, count transformation models can be partitioned by transformation trees (Hothorn & Zeileis 2017), which in turn lead to transformation forests, as a statistical learning approach for computing predictive distributions.

The greatest challenge in applying count transformation models is their interpretability. The effects of the explanatory environmental variables are not directly interpretable as multiplicative changes in the conditional mean of the count response, as is the case in Poisson or negative binomial models with a log link. For the logit, cloglog and log-log link functions, the effects are still multiplicative, but at the scales of the discrete odds ratio, hazard ratio or reverse time hazard ratio, which might be difficult to communicate to practitioners. If the probit link is used, the effects are interpretable as changes in the conditional mean of the transformed counts. This interpretation is the same as that obtained from running a normal linear regression model on, for example, log-transformed counts, with the important difference that (i) the transformation was estimated from data by optimising (ii) the exact discrete likelihood. Nonetheless, it is possible to plot the estimated transformation function $\boldsymbol{a}(y)^\top\hat{\boldsymbol{\vartheta}}$ against $\log(y+1)$ ex post to assess the appropriateness of applying a log-transformation.

Computational details

All computations were performed using R version 3.6.1 (R Core Team 2019). A reference implementation of transformation models is available in the **mlt** R add-on package (Hothorn 2019a; 2018). A simple user interface to linear count transformation models is available in the **cotram** add-on package (Siegfried & Hothorn 2019).

The following example demonstrates the functionality of the **cotram** package in terms of a count transformation model with a cloglog link explaining how the number of tree pipits (*Anthus trivialis*) varies across different percentages of canopy overstorey cover (coverstorey).

```r
### package cotram available from CRAN.R-project.org
### install.packages(c("cotram", "coin"))
library("cotram")
### tree pipit data; doi: 10.1007/s10342-004-0035-5
data("treepipit", package = "coin")
### fit discrete Cox model to tree pipit counts
m <- cotram(counts ~ coverstorey,  ### log-hazard ratio of
                                   ### coverstorey
            data = treepipit,      ### data frame
            method = "cloglog",    ### link = cloglog
            order = 5,             ### order of Bernstein poly.
            prob = 1)              ### support is 0...5
logLik(m)                          ### log-likelihood

## 'log Lik.' -38.27244 (df=7)

exp(coef(m))                       ### hazard ratio

## coverstorey
##   0.9805453

exp(confint(m))                    ### 95% confidence interval

##                 2.5 %    97.5 %
## coverstorey 0.9697581 0.9914526

### more illustrations
# vignette("cotram", package = "cotram")
```

The data are shown in Figure 5 overlayed with the smoothed version of the estimated conditional distribution functions for varying values of coverstorey.

[Figure 5 about here.]

References

De Felipe, M.; Sáez-Gómez, P. & Camacho, C. (2019) Environmental factors influencing road use in a nocturnal insectivorous bird, *European Journal of Wildlife Research*, 65(3), 31, doi: 10.1007/s10344-019-1267-5.

Dean, A.; Voss, D. & Draguljić, D. (2017) *Design and Analysis of Experiments*, Springer Texts in Statistics, Springer International Publishing, 2nd edn., doi: 10.1007/978-3-319-52250-0.

Farouki, R.T. (2012) The Bernstein polynomial basis: A centennial retrospective, *Computer Aided Geometric Design*, 29(6), 379–419, doi: 10.1016/j.cagd.2012.03.001.

Gotelli, N.J. & Ellison, A.M. (2013) *A Primer of Ecological Statistics*, Sinauer Associates, 2nd edn.

Held, L. & Paul, M. (2012) Modeling seasonality in space-time infectious disease surveillance data, *Biometrical Journal*, 54(6), 824–843, doi: 10.1002/bimj.20120003.

Hothorn, T. (2018) Most likely transformations: The **mlt** package, *Journal of Statistical Software*, URL https://cran.r-project.org/web/packages/mlt.docreg/vignettes/mlt.pdf, accepted for publication 2018-03-05.

Hothorn, T. (2019a) ***mlt**: Most Likely Transformations*, URL https://CRAN.R-project.org/package=mlt, R package version 1.0-5.

Hothorn, T. (2019b) Transformation boosting machines, *Statistics and Computing*, doi: 10.1007/s11222-019-09870-4.

Hothorn, T.; Kneib, T. & Bühlmann, P. (2014) Conditional transformation models, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 76(1), 3–27, doi: 10.1111/rssb.12017.

Hothorn, T.; Möst, L. & Bühlmann, P. (2018) Most likely transformations, *Scandinavian Journal of Statistics*, 45(1), 110–134, doi: 10.1111/sjos.12291.

Hothorn, T.; Müller, J.; Held, L.; Möst, L. & Mysterud, A. (2015) Temporal patterns of deer-vehicle collisions consistent with deer activity pattern and density increase but not general accident risk, *Accident Analysis & Prevention*, 81, 143–152, doi: 10.1016/j.aap.2015.04.037.

Hothorn, T. & Zeileis, A. (2017) Transformation forests, Tech. rep., arXiv 1701.02110, URL https://arxiv.org/abs/1701.02110.

Ives, A.R. (2015) For testing the significance of regression coefficients, go ahead and log-transform count data, *Methods in Ecology and Evolution*, 6(7), 828–835, doi: 10.1111/2041-210X.12386.

Mooney, E.H.; Phillips, J.S.; Tillberg, C.V.; Sandrow, C.; Nelson, A.S. & Mooney, K.A. (2016) Abiotic mediation of a mutualism drives herbivore abundance, *Ecology Letters*, 19(1), 37–44, doi: 10.1111/ele.12540.

Nakagawa, T. & Osaki, S. (1975) The discrete Weibull distribution, *IEEE Transactions on Reliability*, 24(5), 300–301, doi: 10.1109/TR.1975.5214915.

Nelder, J.A. & Wedderburn, R.W. (1972) Generalized linear models, *Journal of the Royal Statistical Society: Series A (General)*, 135(3), 370–384, doi: 10.2307/2344614.

O'Hara, R.B. & Kotze, D.J. (2010) Do not log-transform count data, *Methods in Ecology and Evolution*, 1(2), 118–122, doi: 10.1111/j.2041-210X.2010.00021.x.

Padellini, T. & Rue, H. (2019) Model-aware quantile regression for discrete data, Tech. rep., arXiv 1804.03714, v2, URL https://arxiv.org/abs/1804.03714v2.

Peluso, A.; Vinciotti, V. & Yu, K. (2019) Discrete Weibull generalized additive model: an application to count fertility data, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(3), 565–583, doi: 10.1111/rssc.12311.

R Core Team (2019) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, URL https://www.R-project.org/.

Siegfried, S. & Hothorn, T. (2019) ***cotram**: Count Transformation Models*, URL http://CRAN.R-project.org/package=cotram, R package version 0.1-0.

Snedecor, G.W. & Cochran, W.G. (1967) *Statistical Methods*, The Iowa State University Press, 6th edn.

Sokal, R.R. & Rohlf, F.J. (1967) *Biometry*, W.H. Freeman and Company.

St-Pierre, A.P.; Shikon, V. & Schneider, D.C. (2018) Count data in biology - data transformation or model reformation?, *Ecology and Evolution*, 8(6), 3077–3085, doi: 10.1002/ece3.3807.

Warton, D.I. (2018) Why you cannot transform your way out of trouble for small counts, *Biometrics*, 74(1), 362–368, doi: 10.1111/biom.12728.

Warton, D.I.; Lyons, M.; Stoklosa, J. & Ives, A.R. (2016) Three points to consider when choosing a LM or GLM test for count data, *Methods in Ecology and Evolution*, 7(8), 882–890, doi: 10.1111/2041-210X.12552.

List of Tables

Link F_Z^{-1}	Interpretation of $\boldsymbol{x}^\top\boldsymbol{\beta}$
probit	$\mathbb{E}(h_Y(Y) \mid \boldsymbol{x}) = \boldsymbol{x}^\top\boldsymbol{\beta}$
logit	$\frac{F_{Y\mid\boldsymbol{X}=\boldsymbol{x}}(y\mid\boldsymbol{x})}{1-F_{Y\mid\boldsymbol{X}=\boldsymbol{x}}(y\mid\boldsymbol{x})} = \exp(-\boldsymbol{x}^\top\boldsymbol{\beta})\frac{F_Y(y)}{1-F_Y(y)}$
cloglog	$1 - F_{Y\mid\boldsymbol{X}=\boldsymbol{x}}(y \mid \boldsymbol{x}) = (1 - F_Y(y))^{\exp(-\boldsymbol{x}^\top\boldsymbol{\beta})}$
loglog	$F_{Y\mid\boldsymbol{X}=\boldsymbol{x}}(y \mid \boldsymbol{x}) = F_Y(y)^{\exp(\boldsymbol{x}^\top\boldsymbol{\beta})}$

Table 1: Transformation Model. Interpretation of linear predictors $\boldsymbol{x}^\top\boldsymbol{\beta}$ under different link functions F_Z^{-1}.

List of Figures

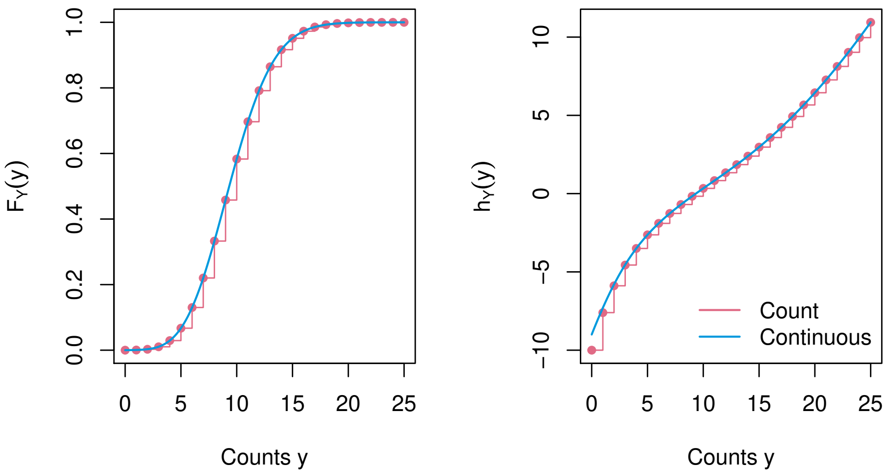

Figure 1: Transformation model. Illustration of a cumulative distribution function ($F_Y(y) = F_Z(h_Y(\lfloor y \rfloor))$, left) and of a transformation function (h_Y, right) of a count response (red) and a corresponding continuous variable (blue). Note that the two functions coincide for counts $0, 1, 2, \ldots$.

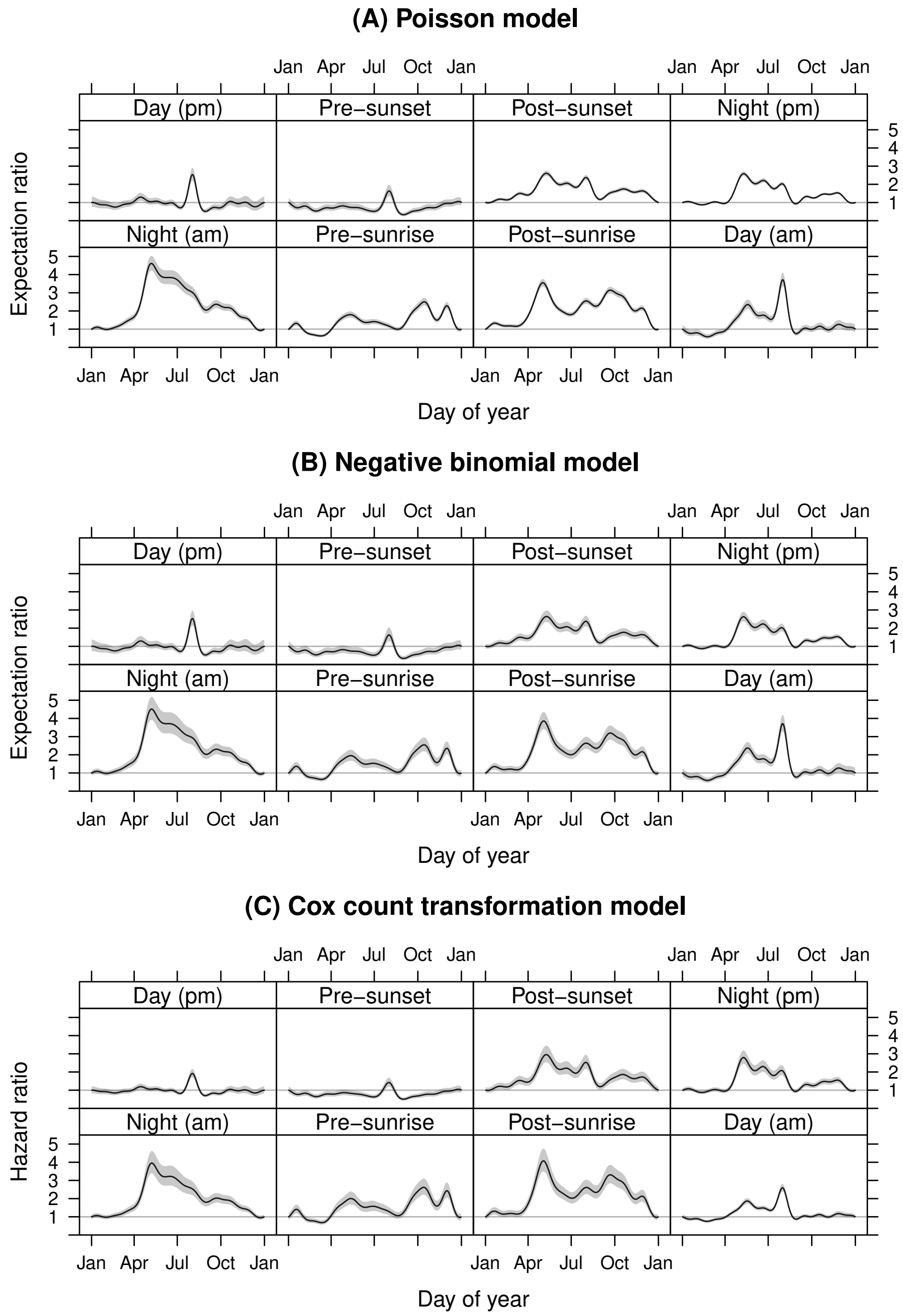

Figure 2: Deer-vehicle collisions. Multiplicative seasonal changes (reference: January 1 at the corresponding time of day) with simultaneous 95% confidence bands for the expected number of deer-vehicle collisions (modelled by the Poisson model with a log link (A) and the negative binomial model with a log link (B)), and for the discrete hazard ratios modelled by the Cox count transformation model (2) (C).

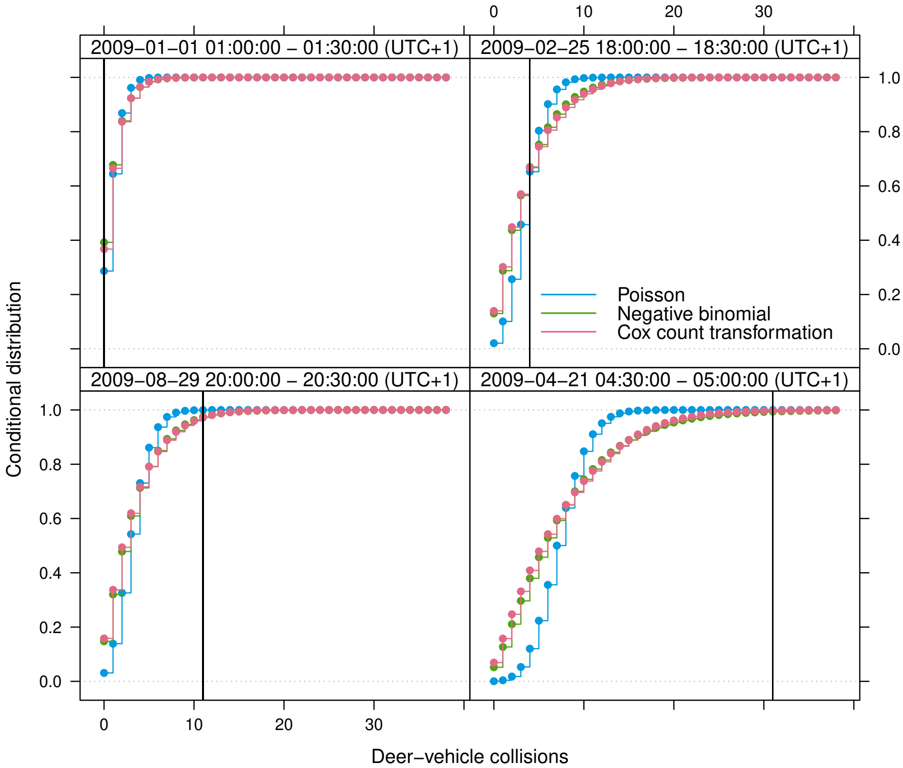

Figure 3: Deer-vehicle collisions. Distributions of the deer-vehicle collision counts conditional on the explanatory environmental parameters of four different time intervals of the year 2009 evaluated for the discrete Cox count transformation model (2) (red), the Poisson model (blue) and the negative binomial model (green). The actually observed deer-vehicle collision counts are shown as a vertical black line.

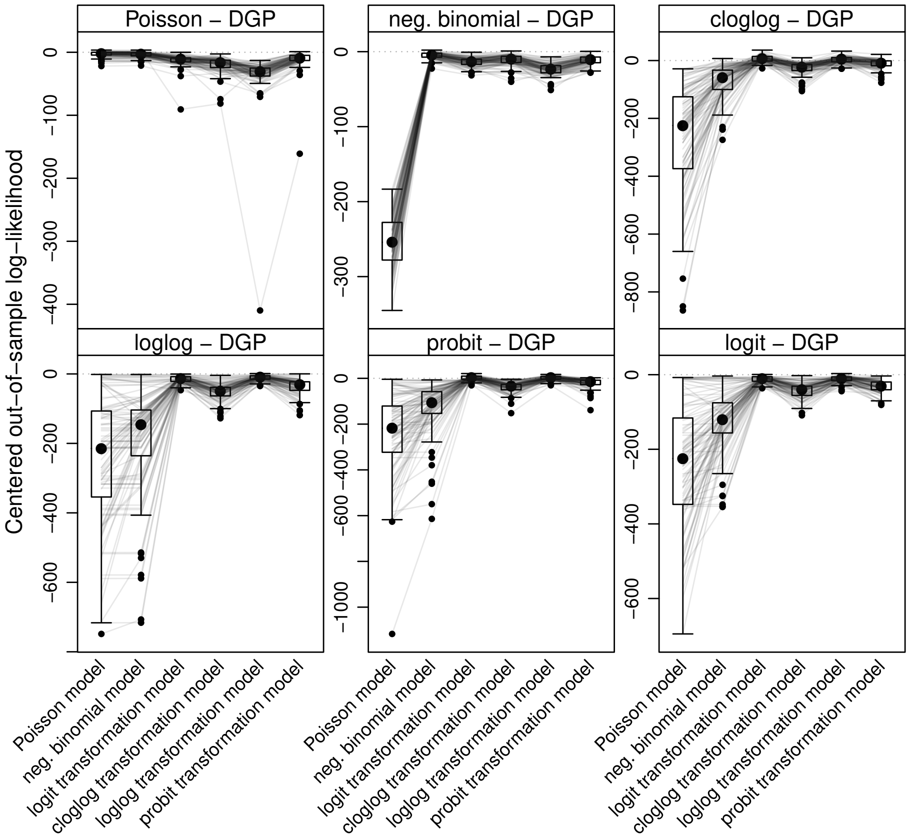

Figure 4: Artificial count-data-generating processes (DGPs). The performance of the count regression models (Poisson, negative binomial and count transformation models outlined in Table 1) assessed by the centered out-of-sample log-likelihood of the corresponding model. Larger values of the out-of-sample log-likelihood indicate a better performance of the corresponding count regression model.

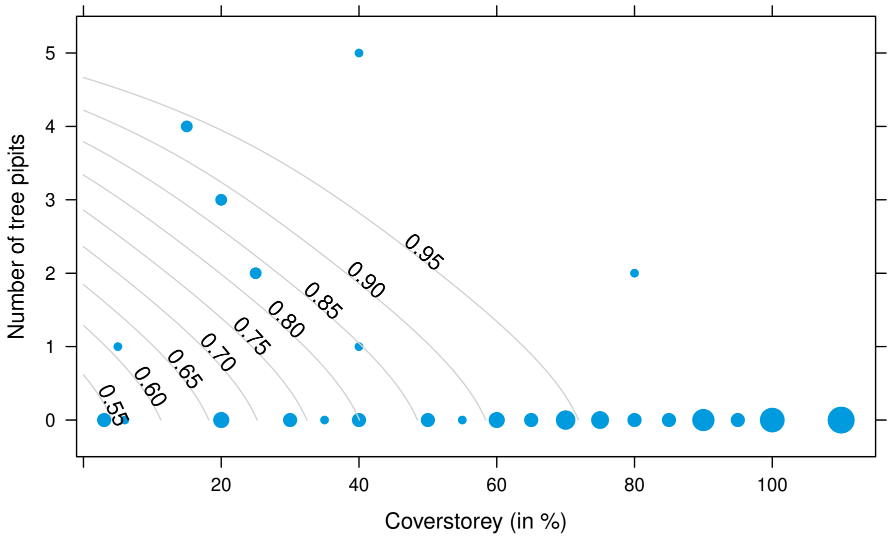

Figure 5: Tree pipit illustration. Number of tree pipits counted at 86 different plots with varying coverstorey. The sizes of the circles are proportional to the square-root of the sample size. Observations are overlayed with the smoothed conditional distribution functions. For a coverstorey of 20%, for example, the probability of not observing any tree pipit is slightly larger than 0.65, the probability of observing at most one tree pipit is somewhat larger than 0.70. For a coverstorey of 60%, the probability of observing at least one tree pipit is less than 0.1.

Affiliation:

Sandra Siegfried and Torsten Hothorn
Institut für Epidemiologie, Biostatistik und Prävention
Universität Zürich
Hirschengraben 84, CH-8001 Zürich, Switzerland
`Torsten.Hothorn@uzh.ch`