Regression Models for Count Data: beyond the Poisson model

Wagner Hugo Bonat
Walmes Marques Zeviani
Eduardo Elias Ribeiro Jr

Preface

The main goal of this material is to provide a technical support for the students attending the course "Regression models for count data: beyond the Poisson model", given as part of the XV Brazilian School of Regression models - March/2017 in Goiânia, Brazil.

The main goal of this course is to present a wider range of statistical models to deal with count data. In particular, we focus on parametric and second-moments specified models. We shall present the model specification along with strategies for model fitting and the associated `R` code. Furthermore, this book-course and supplementary material as `R` (R Core Team 2015) code and data sets are available for the students.

We intend to keep the course in a level suitable for bachelor students who already attended a course on generalized linear models (Nelder and Wedderburn 1972). However, since the course also covers updated topics, it can be of interest of postgraduate students and researches in general.

We designed the course for three hours of tuition. In the first part (two hours) of the course, we shall present the analysis of count data based on fully parametric models. After a brief introduction and motivation on count data, we present the Gamma-Count, Poisson-Tweedie and COM-Poisson distributions. We explore their properties through a consideration of dispersion, zero-inflated and heavy tail indexes, and illustrate their applications with four data analyses. The estimation of these models based on the likelihood paradigm is discussed along with `R` code and worked examples.

In the second part (one hour) of the course, we provide a brief introduction to the estimating function approach (Jørgensen and Knudsen 2004 ; Bonat and Jørgensen 2016) and discuss models based on second-moments assumptions in the style of Wedderburn (1974). In particular, we focus on the recently proposed Extended Poisson-Tweedie (Bonat et al. 2016) and the quasi-Poisson models. The estimating function approach adopted for estimation

and inference is presented along with `R` code and data examples. The use of the `R` package `mcglm` (Bonat 2016) is discussed for fitting the extended Poisson-Tweedie model.

Chapter 1

Introduction

The analysis of count data has received attention from the statistical community in the last four decades. Since the seminal paper published by Nelder and Wedderburn (Nelder and Wedderburn 1972), the class of generalized linear models (GLMs) have a proeminent role for regression modelling of normal and non-normal data including count data. The success enjoyed by the GLM framework comes from its ability to deal with a wide range of normal and non-normal data. GLMs are fitted by a simple and efficient Newton score algorithm relying only on second-moment assumptions for estimation and inference. Furthermore, the theoretical background for GLMs is well established in the class of dispersion models (B. Jørgensen 1987; B. Jørgensen 1997) as a generalization of the exponential family of distributions.

In spite of the flexibility of the GLM class, the Poisson distribution is the only choice for the analysis of count data in this framework. Thus, in practice there is probably an over-emphasis on the use of the Poisson distribution for count data. A well known limitation of the Poisson distribution is its mean and variance relationship, which implies that the variance equals the mean, referred to as equidispersion. In practice, however, count data can present other features, namely underdispersion (mean > variance) and overdispersion (mean < variance). There are many different possible causes for departures from the equidispersion. Furthermore, in practical data analysis a number of these could be involved.

One possible cause of under/overdispersion is departure from the Poisson process. It is well known that the Poisson counts can be interpreted as the number of events in a given time interval where the arrival's times are exponential distributed. When this assumption is violated the resulting counts

can be under or overdispersed (Zeviani et al. 2014). Another possibility and probably more frequent cause of overdispersion is unobserved heterogeneity of experimental units. It can be due, for example, to correlation between individual responses, cluster sampling, ommitted covariates and others.

In general, these departures from the Poisson distribution are manifested in the raw data as a zero-inflated or heavy-tail count distribution. It is important to discuss the consequences of failing to take into account the under or overdispersion when analysing count data. In the case of overdispersion, the standard errors associated with the regression coefficients calculated under the Poisson assumption are too optimistic and associated hypothesis tests will tend to give false positive results by incorrectly rejecting null hypotheses. The opposite situation will appear in case of underdispersed data. In both cases, the Poisson model provides unreliable standard errors for the regression coefficients and hence potentially misleading inferences. However, the regression coefficients are still consistently estimated.

The strategies for constructing alternative count distributions are related with the causes of the non-equidispersion. When departures from the Poisson process are plausible the class of duration time models (Winkelmann 2003) can be employed. This class of models changes the distribution of the time between events from the exponential to more general distributions, like gamma and inverse Gaussian. In this course, we shall discuss one example of this approach, namely, the Gamma-Count distribution (Zeviani et al. 2014). This distribution assumes that the time between events is gamma distributed, thus it can deal with under, equi and overdispersed count data.

On the other hand, if unobserved heterogeneity is present its in general implies extra variability and consequently overdispersed count data. In this case, a Poisson mixtures is commonly applied. This approach consists of include random effects on the observation level, and thus take into account the unobserved heterogeneity. Probably, the most popular example of this approach is the negative binomial model, that corresponds to a Poisson-gamma mixtures. In this course, we shall present the Poisson-Tweedie family of distributions, which in turn corresponds to Poisson-Tweedie mixtures (Bonat et al. 2016; Jørgensen and Kokonendji 2015). Finally, a third approach to deal with non-equidispersed count data consists of generalize the Poisson distribution by adding an extra parameter to model under and overdispersion. Such a generalization can be done using the class of weighted Poisson distributions (Del Castillo and Pérez-Casany 1998). One popular example of this approach is the Conway–Maxwell–Poisson distribution (COM-Poisson) (K. F. Sellers and Shmueli 2010). The COM-Poisson is a member of the exponential family, has the Poisson and geometric distributions as special cases and the Bernoulli distribution as a limiting case. It can deal with both

under and overdispersed count data. Thus, given the nice properties of the COM-Poisson distribution for handling count data, we choose to present this model as part of this short course.

In this short course we shall highlight and compare the ability of these distributions to deal with count data through a consideration of dispersion, zero-inflated and heavy tail indexes. Furthermore, we specify regression models and illustrate their application with four worked examples.

In Chapter 2 we present the properties and regression models based on the Poisson, Gamma-count, Poisson-Tweedie and COM-Poisson distributions. Moreover, we compare these distributions using the dispersion, zero-inflated and heavy-tail indexes. Estimation and inference for these models based on the likelihood paradigm are described in Chapter 3. In Chapter 4, we extend the Poisson and Poisson-Tweedie by specifying models based only on second-moments assumptions and present the related estimation and inference based on the estimating function approach. Chapter 5 presents four worked examples. Finally, in Chapter 6 we discuss the general methods and propose some topics for future works.

Chapter 2

Count distributions: properties and regression models

In this chapter, we present the probability mass function and discuss the main properties of the Poisson, Gamma-Count, Poisson-Tweedie and COM-Poisson distributions.

Poisson distribution

The Poisson distribution is a notorious discrete distribution. It has a dual interpretation as a natural exponential family and as an exponential dispersion model. The Poisson distribution denoted by $P(\mu)$ has probability mass function

$$
\begin{aligned}
p(y;\mu) &= \frac{\mu^y}{y!}\exp\{-\mu\} & (2.1)\\
&= \frac{1}{y!}\exp\{\phi y - \exp\{\phi\}\}, \quad y \in \mathbb{N}_0, (\#eq:Poisson) & (2.2)
\end{aligned}
$$

where $\phi = \log\{\mu\} \in \mathbb{R}$. Hence the Poisson is a natural exponential family with cumulant generator $\kappa(\phi) = \exp\{\phi\}$. We have $\mathrm{E}(Y) = \kappa'(\phi) = \exp\{\phi\} = $

μ and $\text{var}(Y) = \kappa''(\phi) = \exp\{\phi\} = \mu$. The probability mass function @ref(eq:Poisson) can be evaluated in R through the `dpois()` function.

In order to specify a regression model based on the Poisson distribution, we consider a cross-section dataset, (y_i, x_i), $i = 1, \ldots, n$, where y_i's are iid realizations of Y_i according to a Poisson distribution. The Poisson regression models is defined by

$$Y_i \sim P(\mu_i), \quad \text{with} \quad \mu_i = g^{-1}(\boldsymbol{x_i}^\top \boldsymbol{\beta}).$$

In this notation, $\boldsymbol{x_i}$ and $\boldsymbol{\beta}$ are $(q \times 1)$ vectors of known covariates and unknown regression parameters, respectively. Moreover, g is a standard link function, for which we adopt the logarithm link function, but potentially any other suitable link function could be adopted.

Gamma-Count distribution

The Poisson distribution as presented in @ref(eq:Poisson) follows directly from the natural exponential family and thus fits in the generalized linear models (GLMs) framework. Alternatively, the Poisson distribution can be derived by assuming independent and exponentially distributed times between events (Zeviani et al. 2014). This derivation allows for a flexible framework to specify more general models to deal with under and overdispersed count data.

As point out by Winkelmann (2003) the distributions of the arrival times determine the distribution of the number of events. Following Winkelmann (1995), let $\tau_k, k \in \mathbb{N}$ denote a sequence of waiting times between the $(k-1)$th and the kth events. Then, the arrival time of the yth event is given by $\nu_y = \sum_{k=1}^{y} \tau_k$, for $y = 1, 2, \ldots$. Furthermore, denote Y the total number of events in the open interval between 0 and T. For fixed T, Y is a count variable. Indeed, from the definitions of Y and ν_y we have that $Y < y$ iff $\nu_y \geq T$, which in turn implies $P(Y < y) = P(\nu_y \geq T) = 1 - F_y(T)$, where $F_y(T)$ denotes the cumulative distribution function of ν_y. Furthermore,

$$
\begin{aligned}
P(Y = y) &= P(Y < y + 1) - P(Y < y) & (2.3) \\
&= F_y(T) - F_{y+1}(T). (\#eq : DURATION) & (2.4)
\end{aligned}
$$

Equation @ref(eq:DURATION) provides the fundamental relation between the distribution of arrival times and the distribution of counts. Moreover, this type of specification allows to derive a rich class of models for count data

by choosing a distribution for the arrival times. In this material, we shall explore the Gamma-Count distribution which is obtained by specifying the arrival times distribution as gamma distributed.

Let τ_k be identically and independently gamma distributed, with density distribution (dropping the index k) given by

$$f(\tau; \alpha, \gamma) = \frac{\gamma^\alpha}{\Gamma(\alpha)} \tau^{\alpha-1} \exp\{-\gamma\tau\}, \quad \alpha, \gamma \in \mathbb{R}^+. \tag{2.5}$$

In this parameterization $E(\tau) = \alpha/\gamma$ and $var(\tau) = \alpha/\gamma^2$. Thus, by applying the convolution formula for gamma distributions, it is easy to show that the distribution of ν_y is given by

$$f_y(\nu; \alpha, \gamma) = \frac{\gamma^{y\alpha}}{\Gamma(y\alpha)} \nu^{y\alpha-1} \exp\{-\gamma\nu\}. \tag{2.6}$$

To derive the new count distribution, we have to evaluate the cumulative distribution function, which after the change of variable $u = \gamma\alpha$ can be written as

$$F_y(T) = \frac{1}{\Gamma(y\alpha)} \int_0^{\gamma T} u^{n\alpha-1} \exp\{-u\} du, (\#eq: INTEGRAL) \tag{2.7}$$

where the integral is the incomplete gamma function. We denote the right side of @ref{eq:INTEGRAL} as $G(\alpha y, \gamma T)$. Thus, the number of event occurrences during the time interval $(0, T)$ has the two-parameter distribution function

$$P(Y = y) = G(\alpha y, \gamma T) - G(\alpha(y+1), \gamma T), (\#eq: MASSFUNCTION) \tag{2.8}$$

for $y = 0, 1, \ldots$, where $\alpha, \gamma \in \mathbb{R}^+$. Winkelmann (1995) showed that for integer α the probability mass function defined in @ref{eq:MASSFUNCTION} is given by

$$P(Y = y) = \exp^{\{-\gamma T\}} \sum_{i=0}^{\alpha-1} \frac{(\gamma T)^{\alpha y+i}}{\alpha y + i}!. \tag{2.9}$$

For $\alpha = 1$, $f(\tau)$ is the exponential distribution and @ref{eq:MASSFUNCTION} clearly simplifies to the Poisson distribution. The following `R` function can

be used to evaluate the probability mass function of the Gamma-Count distribution.

```r
dgc <- function(y, gamma, alpha, log = FALSE) {
  p <- pgamma(q = 1,
              shape = y * alpha,
              rate = alpha * gamma) -
    pgamma(q = 1,
           shape = (y + 1) * alpha,
           rate = alpha * gamma)
  if(log == TRUE) {p <- log(p)}
  return(p)
}
```

Although, numerical evaluation of @ref{eq:MASSFUNCTION} can easily be done, the moments (mean and variance) cannot be obtained in closed form. Winkelmann (1995) showed for a random variable $Y \sim GC(\alpha, \gamma)$, where $GC(\alpha, \gamma)$ denotes a Gamma-Count distribution with parameters α and γ, $E(Y) = \sum_{i=1}^{\infty} G(\alpha i, \gamma T)$.

Furthermore, for increasing T it holds that

$$Y(T) \stackrel{a}{\sim} N\left(\frac{\gamma T}{\alpha}, \frac{\gamma T}{\alpha^2}\right),\qquad(2.10)$$

thus the limiting variance-mean ratio equals a constant $1/\alpha$. Consequently, the Gamma-Count distribution displays overdispersion for $0 < \alpha < 1$ and underdispersion for $\alpha > 1$. Figure @ref(fig:PlotGC) presents the probability mass function for some Gamma-Count distributions. We fixed the parameter $\gamma = 10$ and fit the parameter α in order to have dispersion index (DI = var$(Y)/$E(Y)) equaling to 0.5, 2, 5 and 20.
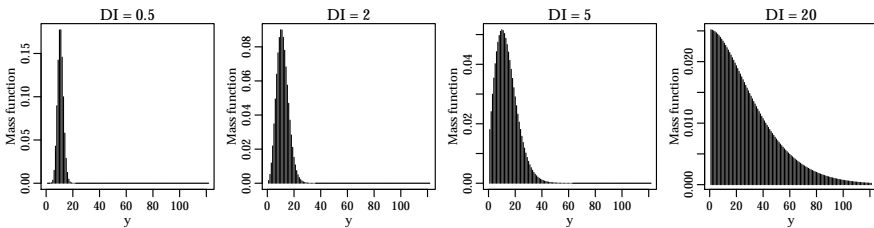

Figure 2.1: Gamma-Count probability mass function by values of the dispersion index (DI).

The Gamma-Count regression model assumes that the period at risk (T)

is identical for all observations, thus T may be set to unity without loss of generality. In the Gamma-count regression model, the parameters depend on a vector of individual covariates $\boldsymbol{x}_i$. Thus, the Gamma-Count regression model is defined by

$$\mathrm{E}(\tau_i|\boldsymbol{x}_i) = \frac{\alpha}{\gamma} = g^{-1}(-\boldsymbol{x_i}^\top\boldsymbol{\beta}). \qquad (2.11)$$

Consequently, the regression model is for the waiting times and not directly for the counts. Note that, $\mathrm{E}(N_i|\boldsymbol{x}_i) = \mathrm{E}(\tau_i|\boldsymbol{x}_i)-1$ iff $\alpha = 1$. Thus, $\hat{\boldsymbol{\beta}}$ should be interpreted accordingly. $-\beta$ measures the percentage change in the expected waiting time caused by a unit increase in x_i. The model parameters can be estimated using the maximum likelihood method as we shall discuss in Chapter 3.

Poisson-Tweedie distribution

The Poisson-Tweedie distribution (Bonat et al. 2016; Jørgensen and Kokonendji 2015; El-Shaarawi, Zhu, and Joe 2011) consists of include Tweedie distributed random effects on the observation level of Poisson random variables, and thus to take into account unobserved heterogeneity. The Poisson-Tweedie family is given by the following hierarchical specification

$$\begin{aligned} Y|Z &\sim \mathrm{Poisson}(Z) & (2.12) \\ Z &\sim \mathrm{Tw}_p(\mu,\phi), (\#eq:conditional) \end{aligned}$$

where $\mathrm{Tw}_p(\mu,\phi)$ denotes a Tweedie distribution (B. Jørgensen 1987 ; B. Jørgensen 1997) with probability function given by

$$f_Z(z;\mu,\phi,p) = a(z,\phi,p)\exp\{(z\psi - k_p(\psi))/\phi\}.(\#eq:tweedie) \qquad (2.13)$$

In this notation, $\mu = k'_p(\psi)$ is the expectation, $\phi > 0$ is the dispersion parameter, ψ is the canonical parameter and $k_p(\psi)$ is the cumulant function. Furthermore, $\mathrm{var}(Z) = \phi V(\mu)$ where $V(\mu) = k''_p(\psi)$ is the variance function. Tweedie densities are characterized by power variance functions of the form $V(\mu) = \mu^p$, where $p \in (-\infty, 0] \cup [1, \infty)$ is an index determining the distribution. The support of the distribution depends on the value of the power parameter. For $p \geq 2$, $1 < p < 2$ and $p = 0$ the support corresponds to the

positive, non-negative and real values, respectively. In these cases $\mu \in \Omega$, where Ω is the convex support (i.e. the interior of the closed convex hull of the corresponding distribution support). Finally, for $p < 0$ the support corresponds to the real values, however the expectation μ is positive. Here, we required $p \geq 1$, to make $\mathrm{Tw}_p(\mu, \phi)$ non-negative.

The function $a(z, \phi, p)$ cannot be written in a closed form apart of the special cases corresponding to the Gaussian ($p = 0$), Poisson ($\phi = 1$ and $p = 1$), non-central gamma ($p = 3/2$), gamma ($p = 2$) and inverse Gaussian ($p = 3$) distributions (B. Jørgensen 1997). The compound Poisson distribution is obtained when $1 < p < 2$. This distribution is suitable to deal with non-negative data with probability mass at zero and highly right-skewed (Andersen and Bonat 2016).

The Poisson-Tweedie is an overdispersed factorial dispersion model (Jørgensen and Kokonendji 2015) and its probability mass function for $p > 1$ is given by

$$f(y; \mu, \phi, p) = \int_0^\infty \frac{z^y \exp{-z}}{y!} a(z, \phi, p) \exp\{(z\psi - k_p(\psi))/\phi\}dz. (\#eq:pmfPTW)$$

$$(2.14)$$

The integral @ref(eq:pmfPTW) has no closed-form apart of the special case corresponding to the negative binomial distribution, obtained when $p = 2$, i.e. a Poisson gamma mixture. In the case of $p = 1$, the integral @ref(eq:pmfPTW) is replaced by a sum and we have the Neyman Type A distribution. Further special cases include the Hermite ($p = 0$), Poisson compound Poisson ($1 < p < 2$), factorial discrete positive stable ($p > 2$) and Poisson-inverse Gaussian ($p = 3$) distributions (Jørgensen and Kokonendji 2015 ; C. C. Kokonendji, Dossou-Gbété, and Demétrio 2004).

In spite of other approaches to compute the probability mass function of the Poisson-Tweedie distribution are available in the literature (Esnaola et al. 2013 ; Barabesi, Becatti, and Marcheselli 2016). In this material, we opted to computed it by numerical evaluation of the integral in @ref(eq:pmfPTW) using the Monte Carlo method as implemented by the following functions.

```
# Integrand Poisson X Tweedie distributions
integrand <- function(x, y, mu, phi, power) {
    int = dpois(y, lambda = x)*dtweedie(x, mu = mu,
                                        phi = phi, power = power)
    return(int)
}

# Computing the pmf using Monte Carlo
```

```r
dptw <- function(y, mu, phi, power, control_sample) {
    pts <- control_sample$pts
    norma <- control_sample$norma
    integral <- mean(integrand(pts, y = y, mu = mu, phi = phi,
                               power = power)/norma)
    return(integral)
}
dptw <- Vectorize(dptw, vectorize.args = "y")
```

When using the Monte Carlo method we need to specify a proposal distribution, from which samples will be taken to compute the integral as an expectation. In the Poisson-Tweedie case is sensible to use the Tweedie distribution as proposal. Thus, in our function we use the argument `control_sample` to provide these values. The advantage of this approach is that we need to simulate values once and we can reuse them for all evaluations of the probability mass function, as shown in the following code.

```r
require(tweedie)
set.seed(123)
pts <- rtweedie(n = 1000, mu = 10, phi = 1, power = 2)
norma <- dtweedie(pts, mu = 10, phi = 1, power = 2)
control_sample <- list("pts" = pts, "norma" = norma)
dptw(y = c(0, 5, 10, 15), mu = 10, phi = 1, power = 2,
     control_sample = control_sample)
```

```
## [1] 0.0937 0.0590 0.0354 0.0217
```

```r
dnbinom(x = c(0, 5, 10, 15), mu = 10, size = 1)
```

```
## [1] 0.0909 0.0564 0.0350 0.0218
```

It is also possible to use the Gauss-Laguerre method to approximate the integral in @ref(eq:pmfPTW). In the supplementary material `Script2.R`, we provide R functions using both Monte Carlo and Gauss-Laguerre methods to approximate the probability mass function of Poisson-Tweedie distribution.

Figure @ref(fig:ptwpmfplot) presents the empirical probability mass function of some Poisson-Tweedie distributions computed based on a sample of size 100000 (gray). Furthermore, we present an approximation (black) for the probability mass function obtained by Monte Carlo integration. We considered different values of the Tweedie power parameter $p = 1.1$, 2, and 3 combined with different values of the dispersion index. In all scenarios the expectation μ was fixed at 10.

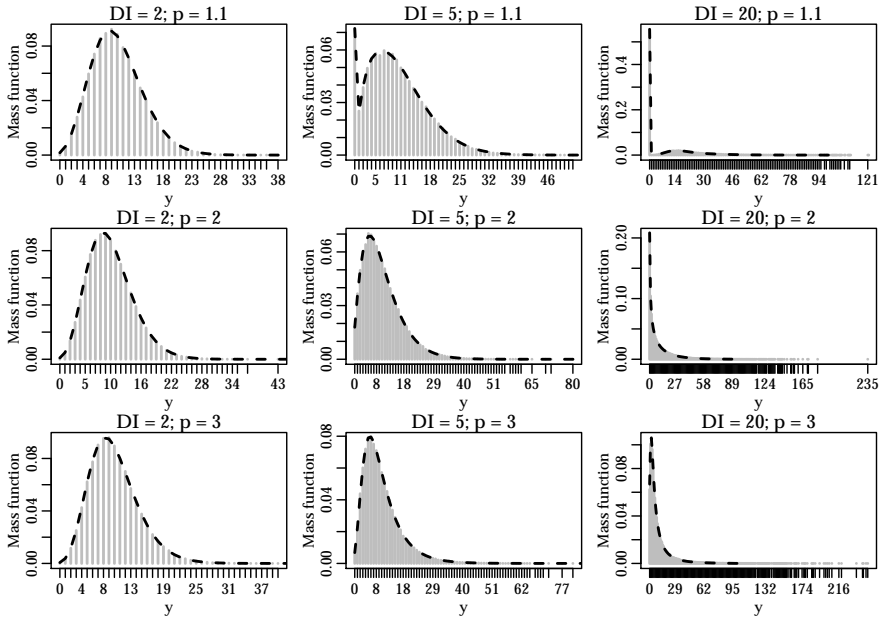For all scenarios considered the Monte Carlo method provides a quite accurate

Figure 2.2: Empirical (gray) and approximated (black) Poisson-Tweedie probability mass function by values of the dispersion index (DI) and Tweedie power parameter.

approximation to the empirical probability mass function. For these examples, we used 5000 random samples from the proposal distribution.

Finally, the Poisson-Tweedie regression model is defined by

$$Y_i \sim PTw_p(\mu_i, \phi), \quad \text{with} \quad \mu_i = g^{-1}(\boldsymbol{x_i}^\top \boldsymbol{\beta}),$$

where $\boldsymbol{x}_i$ and $\boldsymbol{\beta}$ are $(q \times 1)$ vectors of known covariates and unknown regression parameters. The estimation and inference of Poisson-Tweedie regression models based on the maximum likelihood method are challenged by the presence of an intractable integral in the probability mass function and non-trivial restrictions on the power parameter space. In Chapter 3, we discuss maximum likelihood estimation for Poisson-Tweedie regression. Furthermore, in Chapter 4 we extended the Poisson-Tweedie model by using an estimating function approach in the style of Wedderburn (1974).

COM-Poisson distribution

The COM-Poisson distribution belongs to the family of weighted Poisson distributions. A random variable Y is a weighted Poisson distribution if its probability mass function can be written in the form

$$f(y; \lambda, \nu) = \frac{\exp^{\{-\lambda\}} \lambda^y w_y}{W y!}, \quad y = 0, 1, \ldots,$$

where $W = \sum_{i=0}^{\infty} \exp^{\{-\lambda\}} \lambda^i w_s / i!$ is a normalizing constant (Kimberly F. Sellers, Borle, and Shmueli 2012). The COM-Poisson is obtained when $w_y = (y!)^{1-\nu}$ for $\nu \geq 0$. In general, the expectation and variance of the COM-Poisson distribution cannot be expressed in closed-form. However, they can be approximated by

$$\mathrm{E}(Y) \approx \lambda^{1/\nu} - \frac{\nu - 1}{2\nu} \quad \text{and} \quad \mathrm{var}(Y) \approx (1/\nu)\lambda^{1/\nu}.$$

These approximations are accurate when $\nu \leq 1$ or $\lambda > 10^\nu$. The infinite sum involved in computing the probability mass function of the COM-Poisson distribution can be approximated to any level of precision. It can be evaluated in R using the function `dcom()` from the `compoisson` package (Dunn 2012). Figure @ref(fig:comPoispmfplot) presents some COM-Poisson probability mass functions. We tried to find parameters λ and ν in order to have $\mathrm{E}(Y) = 10$ and dispersion index equals to DI $= 0.5, 2, 5$ and 20. However, we could not find any parameter combination to have DI $= 20$.
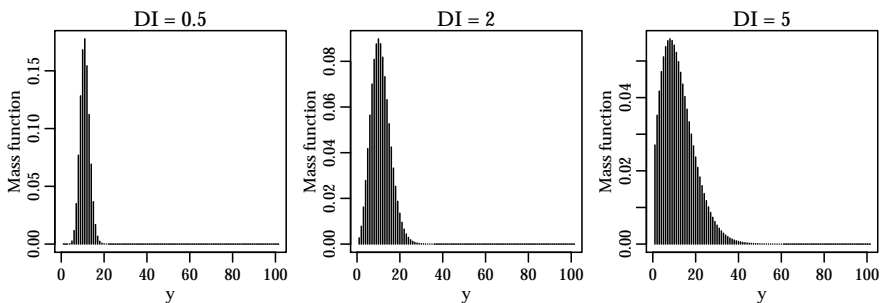

Figure 2.3: COM-Poisson probability mass function by values of the dispersion index (DI).

K. F. Sellers and Shmueli (2010) proposed a regression model based on the COM-Poisson distribution where the parameter λ is described by the values

of known covariates in a generalized linear models style. The COM-Poisson regression model is defined by

$$Y_i \sim CP(\lambda_i, \nu), \quad \text{with} \quad \lambda_i = g^{-1}(\boldsymbol{x_i}^\top \boldsymbol{\beta}).$$

In this notation, the parameter ν is considered the dispersion parameter such that $\nu > 1$ represents underdispersion and $\nu < 1$ overdispersion. The Poisson model is obtained for $\nu = 0$ and as usual we adopt the logarithm link function for g.

Comparing count distributions

Let Y be a count random variable and $\mathrm{E}(Y) = \mu$ and $\mathrm{var}(Y)$ denote its mean and variance, respectively. To explore and compare the flexibility of the models aforementioned, we introduce the dispersion (DI), zero-inflation (ZI) and heavy-tail (HT) indexes, which are respectively given by

$$\mathrm{DI} = \frac{\mathrm{var}(Y)}{\mathrm{E}(Y)}, \quad \mathrm{ZI} = 1 + \frac{\log \mathrm{P}(Y = 0)}{\mathrm{E}(Y)} \tag{2.15}$$

and

$$\mathrm{HT} = \frac{\mathrm{P}(Y = y + 1)}{\mathrm{P}(Y = y)} \quad \text{for} \quad y \to \infty. \tag{2.16}$$

These indexes are defined in relation to the Poisson distribution. Thus, the dispersion index indicates underdispersion for $\mathrm{DI} < 1$, equidispersion for $\mathrm{DI} = 1$ and overdispersion for $\mathrm{DI} > 1$. Similarly, the zero-inflated index is easily interpreted, since $\mathrm{ZI} < 0$ indicates zero-deflation, $\mathrm{ZI} = 0$ corresponds to no excess of zeroes and $\mathrm{ZI} > 0$ indicates zero-inflation. Finally, $\mathrm{HT} \to 1$ when $y \to \infty$ indicates a heavy tail distribution.

For the Poisson distribution the dispersion index equals $1 \ \forall \mu$. In the Poisson case, it is easy to show that $\mathrm{ZI} = 0$ and $\mathrm{HT} \to 0$ when $y \to \infty$. Thus, it is quite clear that the Poisson model can deal only with equidispersed data and has no flexibility to deal with zero-inflation and/or heavy tail count data. In fact, the presented indexes were proposed in relation to the Poisson distribution in order to highlight its limitations. Figure @(fig:indexes) presents the relationship between mean and variance, the dispersion and zero-inflation indexes as a function of the expected values μ for different scenarios and count distributions. Scenario 1 corresponds to the case of underdispersion. Thus, we fixed the dispersion index at $DI = 0.5$ when the mean equaling

10. Since the Poisson-Tweedie cannot deal with underdispersion, in this scenario we present only the Gamma-Count and COM-Poisson distributions. Similarly, scenarios $2 - -4$ are obtained by fixing the dispesion index at $DI = 2, 5$ and 10 when mean equaling 10. In the scenario 4 we could not find a parameter configuration in order to have a COM-Poisson distribution with dispersion index equals 20. Consequently, we present results only for the Gamma-Count and Poisson-Tweedie distributions. Furthermore, Figure @ref(fig:heavytail) presents the heavy tail index for some extreme values of the random variable Y.
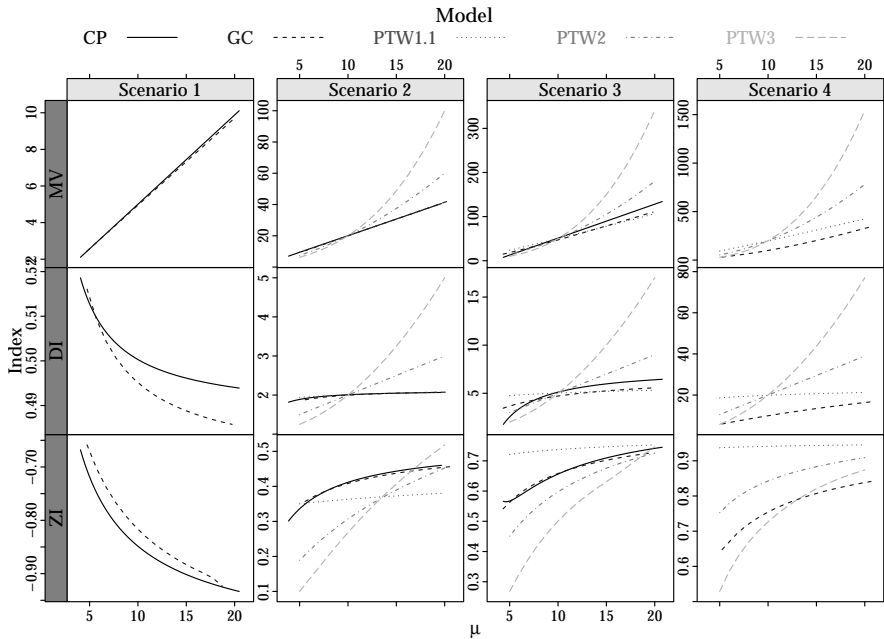

Figure 2.4: Mean and variance relationship (first line), dispersion (DI) and zero-inflation (ZI) indexes as a function of the expected values by simulation scenarios and count distributions.

The indexes presented in Figures @ref(fig:indexes) and @ref(fig:heavytail) show that for all considered scenarios the Gamma-Count and COM-Poisson distributions are quite similar. In general, for these distributions, the indexes slightly depend on the expected values and tend to stabilize for large values of μ. Consequently, the mean and variance relationship is proportional to the dispersion parameter value. In the overdispersion case, the Gamma-Count and COM-Poisson distributions can handle with a limited ammount of zero-inflation and are in general light tailed distributions, i.e. $HT \to 0$ for
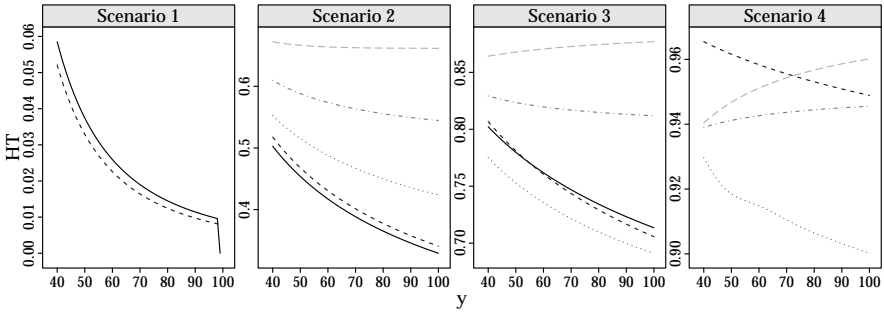
Figure 2.5: Heavy tail index for some extreme values of the random variable Y by simulation scenarios and count distributions.

$y \to \infty$.

Regarding the Poisson-Tweedie distributions the indexes show that for small values of the power parameter the Poisson-Tweedie distribution is suitable to deal with zero-inflated count data. In that case, the DI and ZI are almost not dependent on the values of the mean. Furthermore, the HT decreases as the mean increases. On the other hand, for large values of the power parameter the HT increases with increasing mean, showing that the model is specially suitable to deal with heavy tailed count data. In this case, the DI and ZI increase quickly as the mean increases giving an extremely overdispersed model for large values of the mean. In general, the DI and ZI are larger than one and zero, respectively, which, of course, show that the corresponding Poisson-Tweedie distributions cannot deal with underdispersed and zero-deflated count data.

In terms of regression models, the Poisson and Poisson-Tweedie models are easy and convenient to interpret because the expected value is directly modelled as a function of known covariates in a generalized linear models manner. On the other hand, the Gamma-Count specifies the regression model for the expectation of the times between events and, thus requires careful interpretation. The COM-Poisson regression model is hard to interpret and compare with the traditional Poisson regression model, since it specifies the regression model for the parameter λ that has no easy interpretation in relation to the expectation of the count response variable.

Finally, in terms of computational implementation the simplicity of the Poisson regression model is unquestionable. The probability mass function of the Gamma-Count distribution requires the evaluation of the difference between two cumulative gamma distributions. For large values of the random variable Y, such a difference can be time consuming and inaccurately computed.

Similarly, the COM-Poisson probability mass function involves the evaluation of a infinity sum, which can be computational expensive and inaccurate for large values of Y. Furthermore, for extreme values of λ and ν the infinite sum can numerically diverge making impossible to evaluate the probability mass function. Finally, the Poisson-Tweedie probability mass function involves an intractable integral, which makes the estimation and inference based on likelihood methods computationally intensive.

Chapter 3

The method of maximum likelihood

The estimation and inference for the models discussed in Chapter 2 can be done by the method of maximum likelihood (Silvey 1975). In this Chapter, we present the maximum likelihood method and its main properties along with some examples in R. The maximum likelihood method is applicable mainly in situations where the true distribution of the count random variable Y is known apart of the values of a finite number of unknown parameters. Let $p(y; \boldsymbol{\theta})$ denote the true probability mass function of the count random variable Y. We assume that the family $p(y; \boldsymbol{\theta})$ is labelled by a $(p \times 1)$ parameter vector $\boldsymbol{\theta}$ taking values in Θ a subset of $\mathbb{R}^n$. For a given observed value y of the a random variable Y, the likelihood functioncorresponding to the observation y is defined as $L(\boldsymbol{\theta}; y) = p(y; \boldsymbol{\theta})$. It is important to highlight that $p(y; \boldsymbol{\theta})$ is a probability mass function on the sample space. On the other hand, $L(\boldsymbol{\theta}; y) = p(y; \boldsymbol{\theta})$ is a function on the parameter space Θ. The likelihood function expresses the plausibilities of different parameters after we have observed y, in the absence of any other information we may have about these different values. In particular, for count random variables the likelihood function is the probability of the point y when $\boldsymbol{\theta}$ is the true parameter.

The method of maximum likelihood has a strong intuitive appeal and according to it, we estimate the true parameter $\boldsymbol{\theta}$ by any parameter which maximizes the likelihood function. In general, there is a unique maximizing parameter which is the most plausible and this is the maximum likelihood estimate (Silvey 1975). In other words, a maximum likelihood estimate $\hat{\boldsymbol{\theta}}(y)$ is any elemente of Θ such that $L(\hat{\boldsymbol{\theta}}(y); y) = \max_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}; y)$. At this stage we make

the distinction between the estimate $\hat{\boldsymbol{\theta}}(y)$ and the estimator $\hat{\boldsymbol{\theta}}$. However, we are not maintain this distinction and we shall use only $\hat{\boldsymbol{\theta}}$ leaving the context to make it clear whether we are thinking of $\hat{\boldsymbol{\theta}}$ as a function or as a particular value of a function.

Let Y_i be independent and identically distributed count random variables with probability mass function $p(y; \boldsymbol{\theta})$, whose observed values are denoted by y_i for $i = 1, \ldots, n$. In this case, the likelihood function can be written as the product of the individuals probability mass distributions, i.e.

$$L(\boldsymbol{\theta}; \boldsymbol{y}) = \prod_{i=1}^{n} L(\boldsymbol{\theta}; y_i) = \prod_{i=1}^{n} p(y_i; \boldsymbol{\theta}). (\#eq : LIK) \tag{3.1}$$

For convenience, in practical situations is advisable to work with the log-likelihood function obtained by taking the logarithm of Eq. @ref(eq:LIK). Thus, the maximum likelihood estimator (MLE) for the parameter vector $\boldsymbol{\theta}$ is obtained by maximizing the following log-likelihood function,

$$\updownarrow(\boldsymbol{\theta}) = \sum_{i=1}^{n} L(\boldsymbol{\theta}; y_i). (\#eq : LOGLIK) \tag{3.2}$$

Often, it is not possible to find a relatively simple expression in closed form for the maximum likelihood estimates. However, it is usually possible to assume that maximum likelihood estimates emerge as a solution of the likelihood equations or also called score function, i.e.

$$\mathcal{U}(\boldsymbol{\theta}) = \left(\frac{\partial \updownarrow(\boldsymbol{\theta})}{\partial \theta_1}^{\top}, \ldots, \frac{\partial \updownarrow(\boldsymbol{\theta})}{\partial \theta_p}^{\top} \right)^{\top} = \mathbf{0}. (\#eq : SCORE) \tag{3.3}$$

However, the system of non-linear equations in @ref(eq:SCORE) often have to be solved numerically. The entry (i, j) of the $p \times p$ Fisher information matrix $\mathcal{F}_{\boldsymbol{\theta}}$ for the vector of parameter $\boldsymbol{\theta}$ is given by

$$\mathcal{F}_{\boldsymbol{\theta}_{ij}} = -\mathrm{E}\left\{ \frac{\partial^2 \updownarrow(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right\}. \tag{3.4}$$

In order to solve the system of equations $\mathcal{U}(\boldsymbol{\theta}) = \mathbf{0}$, we employ the Newton scoring algorithm, defined by

$$\boldsymbol{\theta}^{(i+1)} \quad = \quad \boldsymbol{\theta}^{(i)} - \mathcal{F}_{\boldsymbol{\theta}}^{-1} \mathcal{U}(\boldsymbol{\theta}^{(i)}). \tag{3.5}$$

Finally, the well known distribution of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is $N(\boldsymbol{\theta}, \mathcal{F}_{\boldsymbol{\theta}}^{-1})$. Thus, the maximum likelihood estimator is asymptotically consistent, unbiased and efficient.

A critical point of the approach described so far, is that we should be able to compute the first and second derivatives of the log-likelihood function. However, for the Gamma-Count where the log-likelihood function is given by the difference between two integrals, we cannot obtain such derivatives analytically. Simiarly, for the COM-Poisson the log-likelihood function involves an infinite sum and consequently such derivatives cannot be obtained analitycally. Finally, in the Poisson-Tweedie distribution the log-likelihood function is defined by an intractable integral, which implies that we cannot obtain a closed-form for the score function and Fisher information. In these cases, an often approach is to maximize directly the log-likelihood function in Eq.

Thus, an alternative approach is to maximize directly the log-likelihood function in equation @ref(eq:LOGLIK) using a derivative-free algorithm as the Nelder-Mead method (Nelder and Mead 1965) or some other numerical method for maximizing the log-likelihood function, examples include the $BFGS$, conjugate gradient and simulated annealing. All of them are implemented in R through the `optim()` function. The package `bbmle` (Bolker and Team 2014) offers a suite of functions to work with numerical maximization of log-likelihood functions in R. As an example, consider the Gamma-Count distribution described in subsection 2.1. The log-likelihood function for the parameters γ and α in R is given by

```
ll_gc <- function(gamma, alpha, y) {
  ll <- sum(dgc(y = y, gamma = gamma, alpha = alpha, log = TRUE))
  return(-ll)
}
```

Now, for a given vector of observed count values, we can numerically maximize the log-likelihood function in @ref(llgc) using the function `mle2()` from the `bbmle`package. It is important to highlight that by default the `mle2()` function requires the negative of the log-likelihood function instead of the log-likelihood itself. Thus, our function returns the negative value of the log-likelihood function.

```
require(bbmle)
y <- rpois(100, lambda = 10)
fit_gc <- mle2(ll_gc, start = list("gamma" = 10, "alpha" = 1),
               data = list("y" = y))
```

The great advantage of the `bbmle` package for maximum likelihood estimation

in R, is that it already provides standard methods like `summary()`, `coef()`, `confint()`, `vcov()`, `profile()` and other for objects of `mle2` class.

```
summary(fit_gc)
```

```
## Maximum likelihood estimation
##
## Call:
## mle2(minuslogl = ll_gc, start = list(gamma = 10, alpha = 1),
##     data = list(y = y))
##
## Coefficients:
##       Estimate Std. Error z value    Pr(z)
## gamma    9.842      0.335    29.38 < 2e-16 ***
## alpha    0.929      0.139     6.68 2.5e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## -2 log L: 518
```

Similar functions can be done for the Poisson, Poisson-Tweedie and COM-Poisson distributions. In the supplementary material `Script5.R`, we provide some functions for maximum likelihood estimation of Poisson-Tweedie and COM-Poisson distributions.

Andersen, D. A, and W. H. Bonat. 2016. "Double Generalized Linear Compound Poisson Models to Insurance Claims Data." *ArXiv.*

Barabesi, L., C. Becatti, and M. Marcheselli. 2016. "The Tempered Discrete Linnik Distribution." *ArXiv E-Prints.*

Bolker, Ben, and R Development Core Team. 2014. *Bbmle: Tools for General Maximum Likelihood Estimation.* https://CRAN.R-project.org/package= bbmle.

Bonat, W. H. 2016. "Mcglm: Multivariate Covariance Generalized Linear Models." *Gitlab Repository.* http://git.leg.ufpr.br/wbonat/mcglm; GitLab.

Bonat, W. H., and B. Jørgensen. 2016. "Multivariate Covariance Generalized Linear Models." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 65: 649–75.

Bonat, W. H., B. Jørgensen, J. Kokonendji C. C. andHinde, and C. G. B. Démetrio. 2016. "Extended Poisson-Tweedie: Properties and Regression Model for Count Data." *Arxiv.*

Del Castillo, Joan, and Marta Pérez-Casany. 1998. "Weighted Poisson

Distributions for Overdispersion and Underdispersion Situations." *Annals of the Institute of Statistical Mathematics* 50 (3): 567–85.

Dunn, Jeffrey. 2012. *Compoisson: Conway-Maxwell-Poisson Distribution.* https://CRAN.R-project.org/package=compoisson.

El-Shaarawi, A. H., R. Zhu, and H. Joe. 2011. "Modelling Species Abundance Using the Poisson-Tweedie Family." *Environmetrics* 22 (2): 152–64.

Esnaola, Mikel, Pedro Puig, David Gonzalez, Robert Castelo, and Juan R. Gonzalez. 2013. "A Flexible Count Data Model to Fit the Wide Diversity of Expression Profiles Arising from Extensively Replicated Rna-Seq Experiments." *BMC Bioinformatics* 14 (1): 254–76.

Jørgensen, B., and S. J. Knudsen. 2004. "Parameter Orthogonality and Bias Adjustment for Estimating Functions." *Scandinavian Journal of Statistics* 31 (1): 93–114.

Jørgensen, B., and C. Kokonendji. 2015. "Discrete Dispersion Models and Their Tweedie Asymptotics." *AStA Advances in Statistical Analysis.*

Jørgensen, Bent. 1987. "Exponential Dispersion Models." *Journal of the Royal Statistical Society. Series B (Methodological)* 49 (2): 127–62.

———. 1997. *The Theory of Dispersion Models.* Chapman & Hall, London.

Kokonendji, C. C., S. Dossou-Gbété, and C. G. B. Demétrio. 2004. "Some Discrete Exponential Dispersion Models: Poisson-Tweedie and Hinde-Demétrio Classes." *Statistics and Operations Research Transactions* 28 (2): 201–14.

Nelder, J. A., and R. Mead. 1965. "A Simplex Method for Function Minimization." *The Computer Journal* 7: 308–13.

Nelder, J. A., and R. W. M. Wedderburn. 1972. "Generalized Linear Models." *Journal of the Royal Statistical Society. Series A* 135 (3): 370–84.

R Core Team. 2015. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing.

Sellers, K. F., and G. Shmueli. 2010. "A Flexible Regression Model for Count Data." *Ann. Appl. Stat.* 4 (2): 943–61.

Sellers, Kimberly F., Sharad Borle, and Galit Shmueli. 2012. "The Com-Poisson Model for Count Data: A Survey of Methods and Applications." *Applied Stochastic Models in Business and Industry* 28 (2): 104–16.

Silvey, S.D. 1975. *Statistical Inference.* Chapman and Hall/Crc Monographs on Statistics and Applied Probability. Taylor; Francis.

Wedderburn, R. W. M. 1974. "Quasi-Likelihood Functions, Generalized

Linear Models, and the Gauss—Newton Method." *Biometrika* 61 (3): 439–47.

Winkelmann, R. 1995. "Duration Dependence and Dispersion in Count-Data Models." *Journal of Business & Economic Statistics* 13 (4): 467–74.

———. 2003. *Econometric Analysis of Count Data.* Springer.

Zeviani, W. M., P. J. Ribeiro Jr, W. H. Bonat, S. E. Shimakura, and J. A. Muniz. 2014. "The Gamma-Count Distribution in the Analysis of Experimental Underdispersed Data." *Journal of Applied Statistics* 41 (12): 2616–26.