Prediction of Overdispersed Count Data Using Real-Time Cluster-Based Discretization of Explanatory Variables

Evženie Uglickich[1(✉)] and Ivan Nagy[1,2]

[1] Department of Signal Processing, Institute of Information Theory and Automation of the CAS, Pod vodárenskou věží 4, 18208 Prague, Czech Republic
{uglickich,nagy}@utia.cas.cz
[2] Faculty of Transportation Sciences, Czech Technical University, Na Florenci 25, 11000 Prague, Czech Republic
nagy@fd.cvut.cz

Abstract. The chapter focuses on the description of the relationship of the count variable and explanatory Gaussian variables. The cluster-based model is proposed, which is constructed on conditionally independent Gaussian clusters captured in real time using recursive algorithms of the Bayesian mixture estimation theory. The resulting model is expected to be used for predicting count data using real time Gaussian observations. The Poisson distribution of the count data is used as a basic model. However, in reality, count data often do not satisfy the Poisson assumption of equal mean and variance. For this case, five cluster-based Poisson-related models of overdispersed data have been studied. The experimental part of the chapter demonstrates a comparison of the prediction accuracy of the considered models with two theoretical counterparts for the case of weak and strong overdispersion with the help of simulations. The paper reports that the most accurate prediction in average has been provided by the cluster-based Generalized Poisson models.

Keywords: Cluster-based model · Count data · Overdispersion · Recursive Bayesian mixture estimation

1 Introduction

The chapter focuses on modeling and predicting count data variables generally described by the Poisson distribution. From a practical point of view, this task is required in application areas, dealing with random independent events observed with a constant intensity per time unit (e.g., social sciences, medicine, transportation, etc.) [1]. Specific examples of count variables considered per time unit include, e.g., a number of bankruptcies [2], aircraft shutdowns, specific diagnoses, server virus attacks [3], website users, customers [4,5], passengers [6], etc.

This work was supported by the project Arrowhead Tools, the project number ECSEL 826452 and MSMT 8A19009.

O. Gusikhin et al. (Eds.): ICINCO 2021, LNEE 1006, pp. 163–184, 2023.
https://doi.org/10.1007/978-3-031-26474-0_9

In application fields such as, e.g., transportation data analysis, the question of modeling count data depending on other observable variables arises. It can be met, for instance, in predicting passenger demand [6] or electric vehicle plugin intensity [7], etc. As the Poisson distribution does not assume general conditional form, the description of the relationship of the count Poisson-distributed variable with explanatory variables with the aim of constructing the data prediction model is a complicated task. For this task, the Poisson regression (as well as those using some of the Poisson-related distributions of the target variable) is one of the approaches most frequently met in literature e.g., [8–13]. In some sources, the application of linear regression techniques to Poisson-distributed count data due to high number of their possible realizations is also mentioned, see, for instance, [14].

Mixture models known as the universal approximation of nonlinear relation between variables [15] are used as well for the description of multimodal Poisson-distributed data. In this field, studies focusing on mixtures of Poisson distributions [16], mixtures of Poisson regressions [17,18] and Poisson-gamma models [14] have been found.

In this chapter, the joint model of the count Poisson and multimodal Gaussian data is discussed. The similar problem is solved in the papers [19–22] via Gaussian-Poisson mixtures estimated with the help of the iterative expectation-maximization (EM) algorithm [23]. The prediction with the Poisson model is also considered in the papers [24,25] and using the Bayesian methodology in [26]. The paper [27] proposed the approach, where the Poisson prediction probability function is constructed using the joint model of the target Poisson variable and Gaussian explanatory multivariate multimodal variable. The main features of the proposed algorithm [27] for the model estimation and data prediction are as follows: (i) the cluster-based discretization of Gaussian measurements, (ii) estimation of local Poisson models corresponding the discretization intervals, which include Gaussian data belonging to the detected clusters, and (iii) prediction of the target variable based on currently measured data discretized in real time. It should be noted that the cluster-based discretization was investigated, for example, in the papers [28–31]. In [27], it is based on the recursive mixture estimation [32,33] under the Bayesian methodology.

The presented chapter is the extended version of the paper [27]. The aim of the chapter is to present the solution [27] for specific count data, which may be better described by special distributions based on the Poisson model. For example, to fit count observations with a high number of zeros, the zero-inflated Poisson model [34] as well as compound Poisson distributions [35] are used. Data without zeros can be fit by the zero-truncated Poisson distribution with a minimum at 1, see, e.g., [36]. Moreover, in reality, count data often do not satisfy the Poisson assumption of the equality of mean and variance, which means that the overdispersion or underdispersion of the data is observed.

This chapter focuses on overdispersed data as a more desired issue from the practical point of view; however, some of the used models are suitable for both the overdispersed and underdispersed count data [37]. The Generalized Poisson models [37,38] and negative binomial regressions [39] are often used for the description of such data. In this chapter, the prediction of the count variable using the cluster-based discretization of continuous data [27] is considered for the mentioned Poisson-related distributions

not requiring equidispersion, namely, the zero-truncated Poisson distribution (ZTP) and two Generalized Poisson models (GPM). The continuous Rayleigh distribution [40] has been used as well due to the shape of its probability density function similar to the Poisson one and high number of possible realizations of the count variable. The prediction with the Poisson, ZTP, GPM and Rayleigh distributions is experimentally compared with results obtained with the traditional Poisson and negative binomial regressions.

The layout of the chapter is organized as follows: Sect. 1.1 formulates the problem in general for the Poisson distribution. Section 2 provides the general solution with the cluster-based Poisson model. Its specification for overdispersed data described by the Poisson-related distributions can be found in Sect. 3. Section 4 discusses results of the experimental comparison of the mentioned models. Conclusions can be found in Sect. 5.

1.1 Problem Formulation

Let y be the count variable described by the Poisson distribution

$$f\left(y = y_t\right) = e^{-\lambda}\frac{\lambda^{y_t}}{y_t!} \tag{1}$$

with the parameter λ, and let $y_t \in \{0, 1, \ldots\}$ be a realization of y observed on a multimodal system at discrete time instants $t = 1, 2, \ldots, T$.

Let $x = [x_1, x_2, \ldots, x_{N_x}]'$ be the multivariate Gaussian variable, and $x_t = [x_{1;t}, x_{2;t}, \ldots, x_{N_x;t}]'$ contains realizations of the vector x.

Observing the considered multimodal system, for the time instants $t > T$ the realizations x_t are still being measured, but y_t cannot be observed for the time $t > T$. Thus, the problem is verbally formulated as follows:

Predict the values of the dependent Poisson variable based on their relationship with realizations of explanatory Gaussian variables measured at real time $t > T$.

To solve the problem, the multimodality of realizations x_t is going to be used to describe the relationship of y and x under assumption of conditional independence of the individual variables $x_1, \ldots, x_{N_x}$ in the vector x. This will be done by the cluster-based discretization of the individual realizations from x_t and estimation of the local Poisson distributions using the data y_t measured at the same time instants as x_t in the detected clusters. The labels of the clusters will represent the discretized values of the Gaussian variables as it is shown in Fig. 1, which provides an illustrative scheme of the presented approach of the local model estimation. In this figure, for example, the blue-colored Poisson distribution covers count data, which have been measured at the same time with Gaussian data belonging to the cluster described by the blue-colored Gaussian distribution. The number 1 is the label of this blue cluster, which means that regarding this Gaussian distribution, the Poisson distribution numbered 1 exists.

This is believed to allow obtaining the conditional model of y depending on x in clusters in the form of the cluster-based Poisson model and use it for the prediction in real time. The solution is presented below.

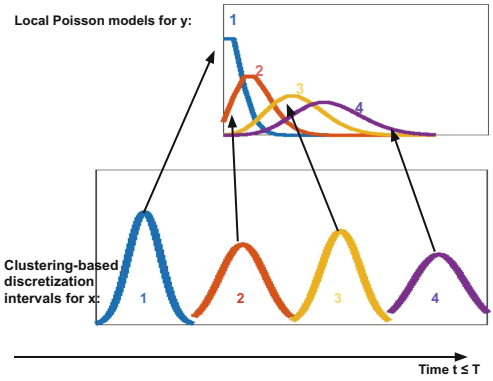

Fig. 1. The scheme of the estimation of the local Poisson distributions.

2 General Solution with Cluster-Based Poisson Model

Generally, the relationship between the dependent variable y and explanatory multivariate variable $x = [x_1, \ldots, x_{N_x}]'$ is described by the joint distribution $f(y = y_t, x)$, where the denotation $f(y = y_t)$ relates to the discrete random variable y described by the Poisson distribution (1) with the realization y_t. For the continuous Gaussian multivariate variable x, the denotation of realizations is omitted. This joint distribution is decomposed with the help of the chain rule [41] in the following way:

$$f(y = y_t, x) = f(y = y_t | x_1, \ldots, x_{N_x}) f(x_1, \ldots, x_{N_x}). \tag{2}$$

The conditional distribution in the right side of (2) is the main focus of the work. Assuming the conditional independence of the individual variables $x_1, \ldots, x_{N_x}$, according to the Naive Bayes approach [49] it is derived as follows:

$$f(y = y_t | x_1, \ldots, x_{N_x}) \propto \frac{\prod_{l=1}^{N_x} f(y = y_t | x_l)}{f(y = y_t)^{N_x - 1}}. \tag{3}$$

Proof. According to the Bayes rule, see, e.g., [42], it holds

$$f(y = y_t | x_1, \ldots, x_{N_x}) = \frac{f(x_1, \ldots, x_{N_x} | y = y_t) f(y = y_t)}{f(x_1, \ldots, x_{N_x})}. \tag{4}$$

Assuming the conditional independence of the individual $x_1, \ldots, x_{N_x}$, the right side of the above relation results in

$$\frac{\prod_{l=1}^{N_x} f(x_l | y = y_t) f(y = y_t)}{f(x_1, \ldots, x_{N_x})}. \tag{5}$$

Next, applying the Bayes rule to the individual conditional distributions in (5) gives

$$f(x_l | y = y_t) = \frac{f(y = y_t | x_l) f(x_l)}{f(y = y_t)}, \quad \forall l \in \{1, \ldots, N_x\}. \tag{6}$$

Substituting (6) into (5) provides

$$\begin{aligned}
&\prod_{l=1}^{N_x}\left[\frac{f(y=y_t|x_l)f(x_l)}{f(y=y_t)}\right]\frac{f(y=y_t)}{f(x_1,\ldots,x_{N_x})}\\
&=\frac{\prod_{l=1}^{N_x}f(y=y_t|x_l)\prod_{l=1}^{N_x}f(x_l)}{f(y=y_t)^{N_x}}\frac{f(y=y_t)}{f(x_1,\ldots,x_{N_x})}\\
&=\frac{\prod_{l=1}^{N_x}f(y=y_t|x_l)}{f(y=y_t)^{N_x-1}}\frac{\prod_{l=1}^{N_x}f(x_l)}{f(x_1,\ldots,x_{N_x})}\propto\frac{\prod_{l=1}^{N_x}f(y=y_t|x_l)}{f(y=y_t)^{N_x-1}},
\end{aligned}\tag{7}$$

where in view of the modeled variable y in the left side of (3), $\frac{\prod_{l=1}^{N_x}f(x_l)}{f(x_1,\ldots,x_{N_x})}$ is a constant value with the substituted realizations of each x_l.

In the numerator of the relation (3), the scalar models $f(y = y_t|x_l)$, $l \in \{1,\ldots,N_x\}$ express the dependence of y on the individual variables x_l. Since (i) x_l are continuous and (ii) the Poisson distribution does not have a general conditional form, a discrete-valued condition brings easier solution in the form of the Poisson distribution defined for each discrete value in the condition. That is why it is proposed to approximate each of the models $f(y = y_t|x_l)$ in (3) by the Poisson distribution existing for each discretized value of each variable x_l as follows:

$$f(y=y_t|x_l)\approx f(y=y_t|\tilde{x}_l=i)\equiv f_i(y=y_t),\tag{8}$$

where $\tilde{x}_l$ is the discretized variable of x_l and it has realizations $i \in \{1,2,\ldots,N_{c;l}\}$. The recursive cluster-based discretization is used in the paper, which means that the value i of $\tilde{x}_l$ is a label of the cluster, to which the observations of the corresponding x_l belong, and $N_{c;l}$ is the number of its clusters.

In this way, to discretize the data here means to find the clusters. For this aim, the scalar marginal distributions $f(x_l)$ are used. Each of them is approximated by the mixture of $N_{c;l}$ Gaussian components $\mathcal{N}_i(x_{l;t};\theta_{i;l},r_{i;l})$ with the collection of the unknown expectations $\theta_{i;l}$. Their variances $r_{i;l}$ are set known and fixed in order to locate tops of the data hills. The cluster-based discretization is presented below.

2.1 Cluster-Based Discretization of Explanatory Variables

The realizations $x_t = [x_{1;t},\ x_{2;t},\ \ldots\ x_{N_x;t}]'$ observed at time $t = 1,\ldots,T$ are used for this part of the proposed approach. Here, the aim is to discretize these realizations to the clusters and use labels of the clusters as the values of $\tilde{x}_l$ for (8). Specifically, this task covers the estimation of individual expectations $\theta_{i;l}$ identifying each i-th Gaussian component of each l-th variable and their labels $\tilde{x}_l$.

The cluster-based discretization of realizations of each x_l is based on the recursive Bayesian mixture estimation methodology [32,33,43]. Its significant advantages are: (i) the recursive running of the clustering algorithms based on actually measured data, which is suitable in view of the prediction task formulated in Sect. 1.1, and (ii) the simple and efficient mixture initialization of the univariate components based on prior knowledge. For the discretization, it worth setting relatively bigger number of components, for instance, 10 or 15, in order not to lose the important information in the data.

According to [32,43], the joint distribution of unknown expectations $\theta_{i;l}$ and discretized values $\tilde{x}_l$ is constructed for the i-th Gaussian component of the l-th variable using the Bayes and chain rules as follows: $\forall i \in \{1, 2, \ldots, N_{c;l}\}$, $\forall l \in \{1, \ldots, N_x\}$

$$\begin{aligned} f(\tilde{x}_{l;t} = i, \theta_{i;l} | x_l(t)) &\propto f(x_{l;t}, \tilde{x}_{l;t} = i, \theta_{i;l} | x_l(t-1)) \\ = f(x_{l;t} | \tilde{x}_{l;t} = i, \theta_{i;l}, x_l(t-1)) f(\tilde{x}_{l;t} = i | \theta_{i;l}, x_l(t-1)) f(\theta_{i;l} | x_l(t-1)) \\ = \underbrace{f_i(x_{l;t} | \theta_{i;l}, x_l(t-1))}_{\mathcal{N}_i(x_{l;t}; \theta_{i;l}, r_{i;l})} \underbrace{f_i(\theta_{i;l} | x_l(t-1))}_{\mathcal{N}_i(\theta_{i;l})} \underbrace{f_i(\tilde{x}_{l;t} = i | x_l(t-1))}_{\mathcal{U}_i(\tilde{x}_l)} \\ \propto \underbrace{f_i(x_{l;t} | \theta_{i;l}, x_l(t-1))}_{\mathcal{N}_i(x_{l;t}; \theta_{i;l}, r_{i;l})} \underbrace{f_i(\theta_{i;l} | x_l(t-1))}_{\mathcal{N}_i(\theta_{i;l})} \end{aligned} \tag{9}$$

where

- $\tilde{x}_{l;t} = i$ denotes the label of the i-th cluster described by the i-th component and detected at time t;
- $x_l(t) = \{x_{l;0}, x_{l;1}, \ldots, x_{l;t}\}$ is a collection of all data of x_l up to the time t with the prior knowledge $x_{l;0}$;
- $\mathcal{N}_i(x_{l;t}; \theta_{i;l}, r_{i;l})$ is the i-th Gaussian component of x_l;
- $\mathcal{N}_i(\theta_{i;l})$ is the prior Gaussian probability density function (pdf);
- $\mathcal{U}_i(\tilde{x}_l)$ is the prior uniform distribution of the discretized variable $\tilde{x}_l$ of x_l, which is a constant in view of the left side of the expression;
- and $\tilde{x}_l$ and $\theta_{i;l}$ are assumed to be mutually independent.

To obtain the estimate of the actual component (i.e., to derive the posterior distribution of $\tilde{x}_l$ based on the current data), the decomposed joint distribution from the right side of (9) is marginalized over the expectations $\theta_{i;l}$

$$f_i(\tilde{x}_{l;t} = i | x_l(t)) \propto \int_{\theta^*} \underbrace{f_i(x_{l;t} | \theta_{i;l}, x_l(t-1))}_{\mathcal{N}_i(x_{l;t}; \theta_{i;l}, r_{i;l})} \underbrace{f_i(\theta_{i;l} | x_l(t-1))}_{\mathcal{N}_i(\theta_{i;l})} \, d\theta_{i;l}, \tag{10}$$

where θ^* means the entire definition space of the expectation. Here, the approximation with the help of the Dirac delta function $\delta(\theta_{i;l}, \hat{\theta}_{i;l;t-1})$, which allows to substitute the prior point estimates $\hat{\theta}_{i;l;t-1}$ of the expectations into the Gaussian components, provides a significant simplification of the solution, see, e.g., [33]. Due to this approximation, the integral in the right side of (10) can be denoted by

$$q_{i;l} = \int_{\theta^*} \underbrace{f_i(x_{l;t} | \theta_{i;l}, x_l(t-1))}_{\mathcal{N}_i(x_{l;t}; \theta_{i;l}, r_{i;l})} \underbrace{f_i(\theta_{i;l} | x_l(t-1))}_{\mathcal{N}_i(\theta_{i;l})} \, d\theta_{i;l} \approx \underbrace{f_i(x_{l;t} | \hat{\theta}_{i;l;t-1}, x_l(t-1))}_{\mathcal{N}_i(x_{l;t}; \hat{\theta}_{i;l;t-1}, r_{i;l})}, \tag{11}$$

which is the value of the i-th Gaussian component pdf with the substituted prior point estimate $\hat{\theta}_{i;l;t-1}$ and current realization $x_{l;t}$. After the normalization

$$m_{i;l} = \frac{q_{i;l}}{\sum_{k=1}^{N_{c;l}} q_{k;l}}, \tag{12}$$

it provides the normalized proximity [33] of the current realization $x_{l;t}$ to the i-th component of the l-th variable, $\forall i \in \{1, 2, \ldots, N_{c;l}\}$, $\forall l \in \{1, \ldots, N_x\}$. Then the searched posterior distribution (10) of the discretized variable $\tilde{x}_l$ is

$$f_i(\tilde{x}_{l;t} = i | x_l(t)) \propto m_{i;l}, \tag{13}$$

which represents the probability of the membership of the current realization $x_{l;t}$ to each of the $N_{c;l}$ components of the variable x_l at time t. The point estimate of the discretized variable $\tilde{x}_l$ at time t, which labels the Poisson distributions is the argument of the maxima of (13)

$$\tilde{x}_{l;t} = \arg\max_i f_i(\tilde{x}_{l;t} = i | x_l(t)). \tag{14}$$

In this way, the continuous realizations of each variable x_l at time t are now represented by the discrete labels $\tilde{x}_l$ of its clusters described by the components.

The cluster-based discretization (10)–(14) is performed in the recursive way similarly to [32,43] to re-compute the point estimates of the expectations $\theta_{i;l}$ and discretized variables $\tilde{x}_l$ at time t. For this end, statistics of the prior Gaussian pdfs $\{V_{i;l;t-1}, \kappa_{i;l;t-1}\}$ conjugate to the i-th individual components describing the clusters of the variables x_l are updated with the new realizations weighted by the proximities as follows:

$$V_{i;l;t} = V_{i;l;t-1} + m_{i;l} x_{l;t}, \quad \kappa_{i;l;t} = \kappa_{i;l;t-1} + m_{i;l}. \tag{15}$$

They are used for re-computing the point estimates of the expectations at time t

$$\hat{\theta}_{i;l;t} = \frac{V_{i;l;t}}{\kappa_{i;l;t}}, \quad \forall i \in \{1, 2, \ldots, N_{c;l}\}, \forall l \in \{1, \ldots, N_x\} \tag{16}$$

to be used in (11). The recursive discretization (10)–(14) runs until the time $t = T$, i.e., for a data set with realizations of x_l and y observed at the same time. As a result, the point estimates of $\tilde{x}_l$ (14), which discretize the Gaussian realizations, are now used to label the Poisson distributions (8). This is explained in the next section.

2.2 Estimation of the Local Poisson Distributions on Clusters

The cluster-based Poisson model is constructed using observations of the individual variables x_l discretized to their clusters and the realizations y_t measured at the same time instants as the clustered data up to the time $t = T$. The searched Poisson distributions (8) labeled by the point estimates i of the discretized variable $\tilde{x}_l$ from (14)

$$\underbrace{f_i(y = y_t)}_{\mathcal{P}oi(y_t;\lambda_{i;l})} \tag{17}$$

are given by the point estimates of their parameters $\lambda_{i;l}$, which are the averages of the realizations y_t measured simultaneously with the data $x_{l;t}$ belonging to each i-th cluster, $\forall\, i \in \{1, 2, \ldots, N_{c;l}\}$, $\forall l \in \{1, \ldots, N_x\}$.

2.3 Prediction with the Cluster-Based Poisson Model

According to the problem formulation in Sect. 1.1, the prediction of the realizations y_t for time $t > T$ should be based on the new data $x_{l;t}$. The cluster-based Poisson model is used for the prediction in real time $t > T$ as follows:

1. The current realizations of each x_l are measured;
2. The actual proximities $m_{i;l}$ to their components are computed;
3. The Poisson distributions (17) estimated in Sect. 2.2 are used to obtain their weighted averages $\forall l \in \{1, \ldots, N_x\}$

$$f(y = y_t|x_l) = \sum_{i=1}^{N_{c;l}} m_{i;l} \underbrace{f_i(y = y_t)}_{\mathcal{P}oi(y_t;\lambda_{i;l})}, \tag{18}$$

   which is the resulting cluster-based Poisson model for the variable x_l.
4. To obtain the predictive model, the individual models (18) with each x_l in the condition are substituted into (3), where the denominator contains values of the Poisson distribution for all previously available realizations y_t.
5. The result of (3) is either the predictive distribution $f(y = y_t|x_1, \ldots, x_{N_x})$ or the point prediction of the realization y_t in real time $t > T$ as the argument of the maxima

$$\hat{y}_t = \arg\max_j f(y = y_t|x_1, \ldots, x_{N_x}), \quad j \in \{0, 1, \ldots, N_y\}, \tag{19}$$

   where N_y is the maximum observed value of the count variable. This is the main result according to the problem formulation in Sect. 1.1.

3 Cluster-Based Models of Overdispersed Count Data

This section is devoted to the specification of the solution presented in Sect. 2 to distributions describing count observations, where the Poisson assumption of the equality of mean and variance is violated, i.e., overdispersion or underdispersion is present. In real applications, the overdispersion is frequently met, that is why this section focuses on this case. The distributions considered in the section are suitable for modeling overdispersed count data because, even though they are based on the Poisson distribution, they allow the variance to be a function of mean as a rule through an additional dispersion parameter.

In this chapter, the following distributions are used to describe the count variable y defined in Sect. 1.1, which is assumed to be overdispersed.

3.1 Cluster-Based Zero-Truncated Poisson Model

The zero-truncated Poisson distribution (ZTP) is taken in the chapter using the following denotation

$$f(y = y_t|y_t > 0) = \frac{\lambda^{y_t}}{(e^{\lambda} - 1)y_t!} \tag{20}$$

with the parameter λ and the mean

$$E[y] = \frac{\lambda}{1 - e^{-\lambda}} \tag{21}$$

and the variance as a function of the mean

$$D[y] = \frac{\lambda + \lambda^2}{1 - e^{-\lambda}} - \frac{\lambda^2}{(1 - e^{-\lambda})^2} = E[y](1 + \lambda - E[y]). \tag{22}$$

The ZTP distribution (20) excludes zero realizations $y_t = 0$ of the variable y. For this reason, the data recoding via their shifting to a minimum at 1 is necessary to use this model instead of the Poisson distribution (1) in the presented approach. For the prediction accuracy evaluation, the re-shift at zero must be applied.

The ZTP model (20) is estimated locally on the clusters instead of the Poisson distributions in (17) (see Sect. 2.2) and substituted into (18) in the prediction part of the algorithm (see Sect. 2.3). According to [44], the maximum likelihood estimation of the parameter λ is obtained numerically via solving the equation (21) using the sample mean instead of $E[y]$. This model has been tested for the approach in Sect. 4; however, for bigger values of the parameter λ, the mean is approximately equal to this parameter, and consequently, $D[y] \approx E[y]$ in (22) is obtained. The proposed approach with the ZTP model (20) is therefore restricted for small values of λ.

3.2 Cluster-Based Consul's Generalized Poisson Model

The Consul's Generalized Poisson model (GP1) [38] is used in this work in the form of the following probability function

$$f(y = y_t) = \frac{e^{-(\lambda + a y_t)} (\lambda + a y_t)^{y_t - 1}}{y_t!}, \tag{23}$$

where λ is the parameter of GP1 and a is the dispersion parameter. The mean and variance of the distribution are respectively

$$E[y] = \frac{\lambda}{1 - a}, \quad D[y] = \frac{\lambda}{(1 - a)^3}. \tag{24}$$

The dispersion parameter a is estimated according to the following formula [38,45,46]

$$a = \frac{\sum_{t=T+1}^{\tilde{T}} \left(\frac{|y_t - \hat{y}_t|}{\sqrt{\hat{y}_t}} - 1 \right)}{(\tilde{T} - T) - N_x - 1}, \tag{25}$$

where $\hat{y}_t$ are the predictions obtained either with the cluster-based Poisson model or the traditional Poisson regression for the testing data set of y_t and x_t measured for time $t = T + 1, \ldots, \tilde{T}$, and N_x is the number of the explanatory variables. Further, the point estimate of a is substituted along with the sample mean instead of $E[y]$ in (24) to obtain the estimate of the parameter λ.

Similarly, within the presented approach, the parameter λ of the GP1 model is estimated locally on the clusters instead of the Poisson distributions in (17) from Sect. 2.2 and substituted into (18) in the prediction part of the algorithm (see Sect. 2.3).

3.3 Cluster-Based Famoye's Generalized Poisson Model

Famoye's Generalized Poisson model (GP2) [38] is the next distribution tested within the bounds of the presented approach. It has the form

$$f(y = y_t) = \frac{\lambda}{1 + a\lambda} \frac{(\lambda + a y_t)^{y_t - 1}}{y_t!} e^{\frac{-\lambda(1+a y_t)}{1+a\lambda}}, \tag{26}$$

where λ is the parameter of the GP2 model. This distribution has the following mean and variance:

$$E[y] = \lambda, \quad D[y] = \lambda(1 + a\lambda)^2, \tag{27}$$

while the dispersion parameter a is estimated according to [38,45,46] as follows:

$$a = \frac{\sum_{t=T+1}^{\tilde{T}} \left(\frac{|y_t - \hat{y}_t|}{\sqrt{\hat{y}_t}} - 1 \right) \frac{1}{\hat{y}_t}}{(\tilde{T} - T) - N_x - 1}, \tag{28}$$

with $\hat{y}_t$ and the rest of denotations defined similarly as for (25). Here, the estimation of the parameter λ is straightforward due to the use of the sample mean in (27) locally on the clusters similarly to (17). In the prediction part, it is again substituted into (18) according to Sect. 2.3.

3.4 Cluster-Based Rayleigh Model

The continuous Rayleigh distribution suitable for non-negative data has the shape of the probability density function relatively close to the Poisson distribution. With a high number of possible realizations of the count data it can serve as an approximation of the modeled variable. That is why it is used in this work to test the proposed approach. The Rayleigh distribution has the following form

$$f(y = y_t) = \frac{y_t}{\sigma^2} e^{-y_t^2/(2\sigma^2)}, \tag{29}$$

where σ is the parameter of the Rayleigh distribution. Its mean and variance are approximately connected with σ through the following relations:

$$E[y] \approx 1.253\sigma, \quad D[y] \approx 0.429\sigma^2. \tag{30}$$

The parameter estimation is performed according to [47] using

$$\sigma \approx \sqrt{\frac{1}{2T} \sum_{t=1}^{T} y_t^2}, \tag{31}$$

which is calculated locally on the clusters similarly to (17) and then substituted into (18) instead of the Poisson distribution according to Sect. 2.3.

3.5 Theoretical Counterparts

The distributions introduced in the above sections are used as the cluster-based models instead of the Poisson distribution according to Sect. 2. Results of predicting based on these models are compared with (i) the Poisson regression (which can be used if the equality of mean and variance is not violated) and (ii) the negative binomial regression suitable for overdispersed data.

Poisson Regression. The Poisson regression assumes that the relation between the Poisson parameter $\lambda = E[y] = D[y]$ and realizations of the explanatory variables in the vector x_t has the following form

$$\ln(\lambda) = \theta' x_t = b_0 + b_1 x_{1;t} + b_2 x_{2;t} + \ldots + b_{N_x} x_{N_x;t}, \tag{32}$$

where the vector $\theta = [b_0 \;\; b_1 \;\; b_2 \;\; \ldots \;\; b_{N_x}]'$ contains regression coefficients. Here, they are estimated with the help of the linearization of the Poisson regression and subsequent application of the least square estimator in the following form:

$$\underbrace{\begin{bmatrix} \ln(y_1) \\ \ln(y_2) \\ \ldots \\ \ln(y_T) \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} 1 & x_{1;1} & \ldots & x_{N_x;1} \\ 1 & x_{1;2} & \ldots & x_{N_x;2} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1;T} & \ldots & x_{N_x;T} \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ \ldots \\ b_{N_x} \end{bmatrix}}_{\theta}. \tag{33}$$

The vector of the regression coefficients is estimated using

$$\theta = (X'X)^{-1} X'Y, \tag{34}$$

see, e.g., [41]. The prediction is then obtained using the observations of the explanatory variables for $t > T$ as follows:

$$\hat{y}_t = e^{\theta' x_t}. \tag{35}$$

Negative Binomial Regression. The negative binomial regression is a generalization of the Poisson regression, which does not require the assumption of equidispersion to be satisfied. It assumes that the count variable y follows the negative binomial distribution (NB)

$$f(y = y_t) = \binom{y_t + r - 1}{r - 1} (1 - p)^{y_t} p^r, \tag{36}$$

where p and r are parameters of the distribution. The NB distribution can be also defined with the alternative parametrization through the mean μ and variance σ^2

$$f(y = y_t) = \binom{y_t + \frac{\mu^2}{\sigma^2 - \mu} - 1}{y_t} \left(\frac{\sigma^2 - \mu}{\sigma^2}\right)^{y_t} \left(\frac{\mu}{\sigma^2}\right)^{\mu^2/(\sigma^2 - \mu)}. \tag{37}$$

The NB regression with the target variable y is given by the relation

$$\mu = \exp\{\ln t + b_0 + b_1 x_{1;t} + b_2 x_{2;t} + \ldots + b_{N_x} x_{N_x;t}\}, \tag{38}$$

or

$$\ln(\mu) = \theta' \begin{bmatrix} t \\ x_t \end{bmatrix} = t + b_0 + b_1 x_{1;t} + b_2 x_{2;t} + \ldots + b_{N_x} x_{N_x;t}, \tag{39}$$

where, similarly to the Poisson regression, the vector $\theta = [1 \quad b_0 \quad b_1 \quad b_2 \quad \ldots \quad b_{N_x}]'$ contains regression coefficients. After the linearization, the regression coefficients are estimated similarly to (33–34) with the corresponding arrangements. The prediction is done according to (35) using the extended vector of the explanatory variables instead of x_t.

Two types of the NB regression is distinguished [48]: the first one denoted by NB1 has the variance

$$D[y] = E[y](1 + a_{nb1}), \tag{40}$$

and the second one denoted by NB2 has the variance

$$D[y] = E[y] + a_{nb2} * E[y]^2, \tag{41}$$

while the mean is the same for both of them. According to [48], the dispersion parameter a_{nb1} for the NB1 model with the variance (40) is estimated using auxiliary least squares (similarly to (34)) with the equation

$$\frac{(y_t - \hat{y}_t)^2 - \hat{y}_t}{\hat{y}_t} = a_{nb1} + 0, \tag{42}$$

where $\hat{y}_t$ is the prediction obtained with the Poisson regression in (35). The column of the values of the left side of this equation for each time instant t is denoted by Y; the unit vector of the corresponding dimension is denoted by X and the intercept is equal to zero.

The dispersion parameter a_{nb2} for the NB2 regression with the variance (41) is estimated in a similar way, see, e.g., [48], with the help of auxiliary least squares solved for the equation

$$\frac{(y_t - \hat{y}_t)^2 - \hat{y}_t}{\hat{y}_t} = a_{nb2}\hat{y}_t + 0, \tag{43}$$

where the columns of the observations and predictions are denoted as follows: $Y = \frac{(y_t - \hat{y}_t)^2 - \hat{y}_t}{\hat{y}_t}$, $B_1 = a_{nb2}$, $X = \hat{y}_t$, $B_0 = 0$, and it holds $Y = B_1 x + B_0$.

In this work, the NB1 regression with the variance (40) is used.

Remark: The cluster-based NB model has been also tested with the bounds of the presented approach, assuming the usage of the NB distribution instead of the Poisson model according to Sect. 2. However, due to the computational complexity caused by complex numbers because of the negative number in factorial this attempt has been stopped so far.

The following section demonstrates results of experiments with the introduced models.

4 Experiments

The presented experiments have been conducted in a free and open source programming environment Scilab (www.scilab.org) aimed at engineering and scientific computations. The aim of the experiments was to verify the presented algorithm with the simulated data and compare the accuracy of predicting the count variable y using all of the given models.

4.1 Simulations

3000 data items have been simulated within an individual data set. Each of the data sets contains realizations of the count variable y and four-dimensional multimodal Gaussian variable $x = [x_1, x_2, x_3, x_4]'$, i.e., $N_x = 4$. For the cluster-based discretization of the Gaussian single variables, a different number of components have been chosen during the simulation for each of them. Four variables had $10, 12, 3$ and 15 components respectively in order to test both the high and low number of components.

Realizations of the count variable y have been simulated so that to have different values of the Poisson parameter λ for all of the clusters of each Gaussian variable. The equality of the mean and variance of the resulting count simulations were not supposed due to their multimodal nature, otherwise a single Poisson distribution would be enough to describe the data, which is irrelevant within the bounds of the considered problem. However, the interesting issue is the variances of the overdispersed count data used for the estimation of parameters of the considered distributions corresponding to the clusters of the Gaussian variables.

Approximately 100 data sets have been simulated for the case of weaker and stronger overdispersion of count data. Examples of the data mean and variances of both the cases are given in Table 1, while examples of histograms are shown in Fig. 2.

Table 1. Examples of the data means and variances of the data sets.

Data overdispersion	Mean	Variance
Weak	4.9023333	6.8630822
Strong	34.140667	275.99821

The following section demonstrates results of the prediction of the count variable according to the presented algorithm from Sect. 2 and its comparison with the Poisson and NB regressions.

4.2 Results and Discussion

Prediction of Weakly Overdispersed Count Data. In this section, the presented algorithm has been applied to approximately 100 data sets of 3000 values of x and y, where the values of y have been generated with the weak overdispersion. For each data set, 2800 data items have been utilized for the cluster-based discretization and local model estimation according to Sects. 2.1 and 2.2, i.e., $T = 2800$. The rest of 200 values have been used for the prediction algorithm according to Sect. 2.3 with the given models.

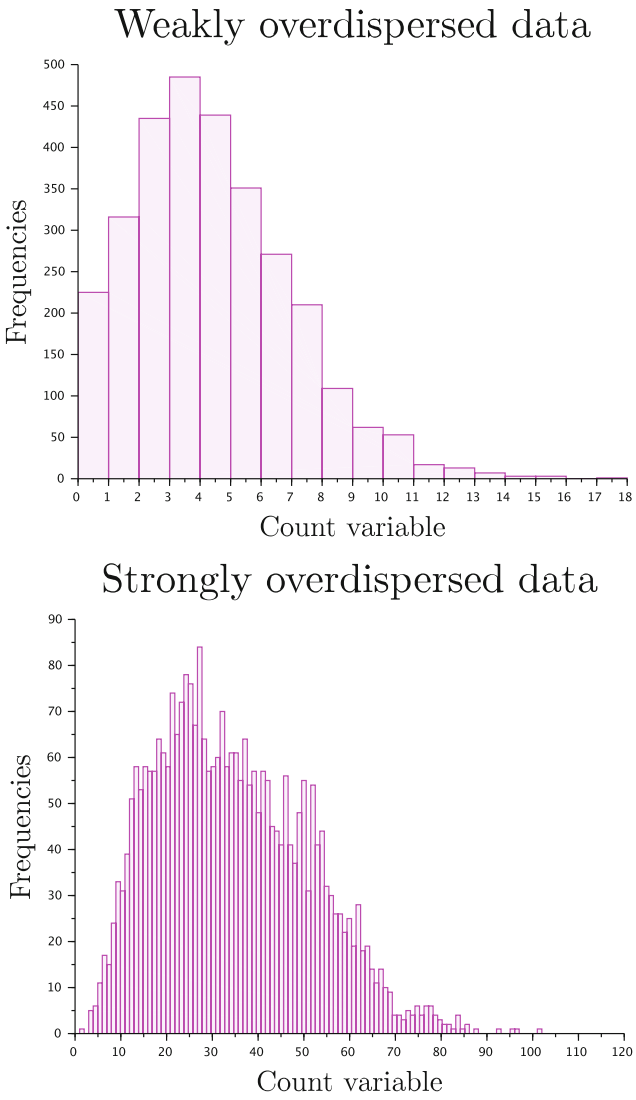

Fig. 2. Histogram examples of data with a weak (top) and strong (bottom) overdispersion.

The results of the experiments have been compared from the prediction accuracy point of view with the help of the following criteria:

$$\text{root-mean-square error RMSE} = \sqrt{\frac{\sum_{t=T+1}^{\tilde{T}} (y_t - \hat{y}_t)^2}{\tilde{T} - T}}, \tag{44}$$

where $\hat{y}_t$ stands for the predicted value obtained with each model, and $\tilde{T} = 3000$;

$$\text{relative-prediction error RPE} = \frac{D[y_t - \hat{y}_t]}{D[y_t]}, \tag{45}$$

where D denotes variance, and $t = T + 1, \ldots, \tilde{T}$;

$$\text{Akaike information criterion AIC} = -2\ln(\mathcal{L}) + 2N_x \tag{46}$$

and

$$\text{Bayesian information criterion BIC} = -2\ln(\mathcal{L}) + N_x \ln(\tilde{T} - T), \tag{47}$$

where $\mathcal{L}$ denotes the likelihood.

Table 2 demonstrates the comparison of the prediction accuracy with the weakly overdispersed data among all of the considered models, including the cluster-based Poisson model and Poisson regression. Due to the smaller values of the count variable, the cluster-based ZTP model has been used as well. The average values of RMSE, RPE, AIC and BIC calculated on the used sets of simulations are presented.

Table 2. The average prediction accuracy for weakly overdispersed data.

Model	RMSE	RPE	AIC	BIC
Cluster-based Poisson model	2.7724659	0.6870497	974.28897	987.50218
Cluster-based ZTP model	2.7760525	0.6887404	1001.3196	1014.5328
Cluster-based GP1 model	2.7965864	0.6901744	1802.6274	1815.8406
Cluster-based GP2 model	2.9883688	0.701766	1082.8114	1096.0246
Cluster-based Rayleigh model	5.5489679	0.7176861	1259.2657	1272.4789
Poisson regression	3.2819422	0.8568735	1131.8207	1148.3372
NB1 regression	3.1889111	0.8374878	1067.8391	1084.3556

It can be seen in Table 2 that all of the compared models provide the relatively high prediction error in view of the small mean and variance of the weakly dispersed data with the average range 20 and the minimum at 0. However, to evaluate the comparison among the obtained results, it can be noted that regarding RMSE and RPE, the cluster-based models (except the Rayleigh model) show the higher prediction accuracy than the traditional Poisson and NB1 regressions. The lowest RMSE, RPE, AIC and BIC have been obtained naturally with the cluster-based Poisson model. It is explained by the weak overdispersion, which means that the data distributions obtained on clusters were close to the Poisson one.

If one omits the cluster-based Poisson model and Poisson regression in view of the overdispersion, the cluster-based ZTP model has the lowest RMSE, RPE, AIC as well as BIC and shows improvements against the NB1 regression. However, it is sensitive to bigger values of the count data (see Sect. 3.1).

From this point of view, the cluster-based GP2 model shows the most balanced improvements of the prediction accuracy in the comparison with the rest of the models

for the case of the weakly overdispersed data: it has the lowest RMSE and RPE. Its AIC and BIC are insignificantly higher than the NB1 regression results.

The comparison of variances of the count data captured on clusters of individual Gaussian variables during the discretization shows that they are close to the variances of the GP2 distributions on these clusters. However, the corresponding test of hypothesis has not been performed because of the small number of components. For the illustration, Table 3 demonstrates the means, variances of data and GP2 variances on clusters of the variables x_1, x_2 and x_3 obtained with one of the data sets. The variable x_4 was omitted to save space. The histograms of the count data collected on the clusters of the variable x_3 are shown in Fig. 3 as an example of the distributions on the clusters.

Table 3. The mean, variance of data and the GP2 variances on clusters of the variables.

Mean	Variance of data	GP2 variance on clusters of x_1
3.2553606	3.4327105	3.1201919
3.7605893	4.5366727	3.5805055
3.9872204	4.9470364	3.7849268
4.3664596	5.196118	4.1241545
4.6079734	5.7258029	4.3383356
4.45	5.9447368	4.1984032
5.0644068	6.9176064	4.7391929
4.2	3.2	3.9756952
5.2034483	7.1107147	4.8602884
5.8913043	7.4323671	5.4524133
Mean	Variance of data	GP2 variance on clusters of x_2
2.952	3.3686452	2.8407398
2.7964602	2.9492731	2.6965651
3.35	3.2589744	3.2069019
3.7756654	4.6097872	3.5941437
3.8919861	5.0407397	3.6991806
4.1469194	5.1565107	3.9282105
4.2212121	5.285263	3.9946516
4.2061856	5.9054538	3.9812241
4.4304348	6.019157	4.1810295
4.5490909	6.3725813	4.2862496
4.689243	5.6870438	4.4100846
4.9888889	6.6206939	4.6732235
Mean	Variance of data	GP2 variance on clusters of x_3
3.5916667	4.4086896	3.4273074
4.3343109	5.7217864	4.0955355
4.6511628	6.2274282	4.3764856

Prediction of Strongly Overdispersed Count Data. In this section, results of predicting of strongly overdispersed count data are discussed. In this part of the experiments, the cluster-based Poisson and ZTP models as well as Poisson regression were not used due to the violated Poisson assumption and higher values of the counts (see Sect. 3.1). The experiments have been conducted under the same conditions as specified in the previous section.

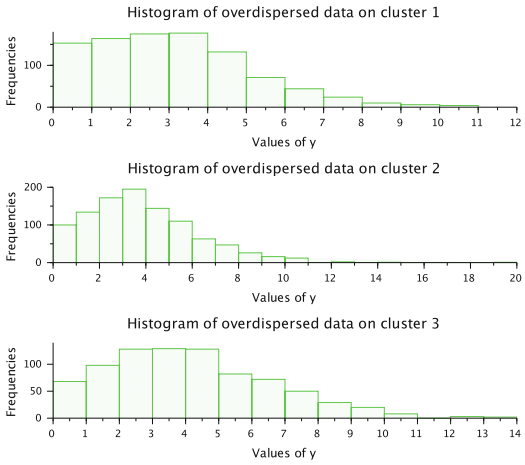

Fig. 3. The histograms of the count data collected on the clusters of the variable x_3.

Table 4. The average prediction accuracy for strongly overdispersed data.

Model	RMSE	RPE	AIC	BIC
Cluster-based GP1 model	5.3527321	0.2450608	2586.4615	2599.6747
Cluster-based GP2 model	10.156487	0.4225181	2988.1786	3001.3918
Cluster-based Rayleigh model	8.5742296	0.5272263	6235.3781	6248.5913
NB1 regression	8.6107072	0.6270219	1784.7504	1801.267

Table 4 provides the comparison of the average values of RMSE, RPE, AIC and BIC of the models calculated on the sets of simulations. Here, it can be seen, that the lowest RMSE and RPE have been obtained with the cluster-based GP1 model. However, its AIC and BIC are in the second place after the NB1 regression.

For the illustration, a fragment of the prediction is shown in Fig. 4, where the GP1 predictions follow the simulations. Figure 5 compares the histograms of the data from one of the testing sets with the GP1 and NB1 predictions. Both the simulations and GP1 predictions have the values approximately from 5 to 50, while the NB1 histogram provides the values from 10 to 44.

For the strongly overdispersed data, there is not the similar concordance in the variances of data and distributions on clusters as observed for weak overdispersion. It is explained by a high measure of the uncertainty in such data.

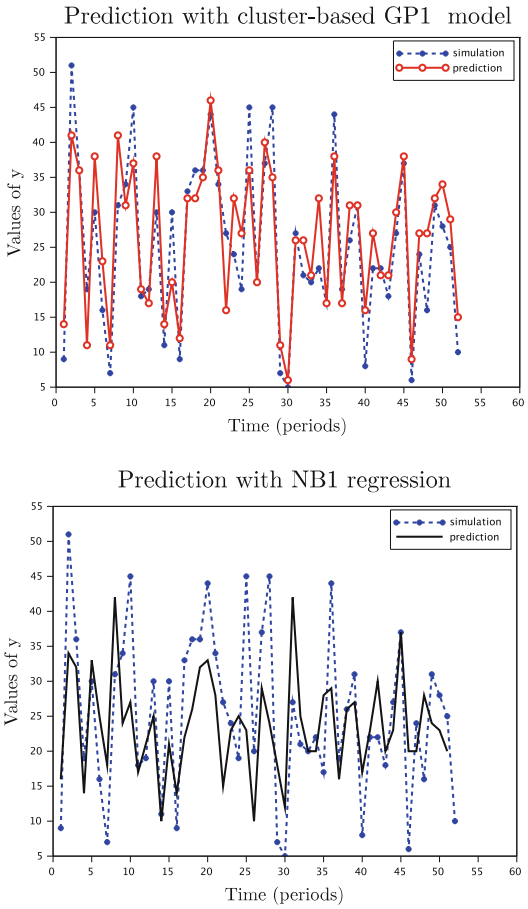

Fig. 4. The prediction with the cluster-based GP1 (top) and NB1 (bottom) models.

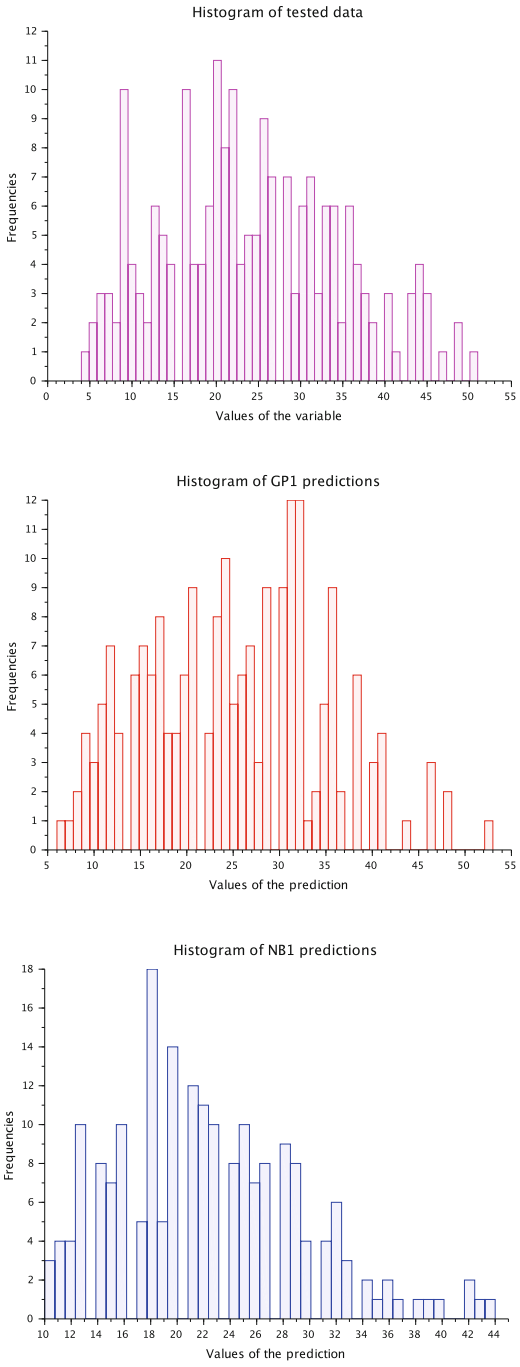

Fig. 5. The histograms of data from the testing set (top), the cluster-based GP1 (middle) and NB1 (bottom) predictions.

5 Conclusion

The presented chapter describes the prediction of the count variable with overdispersed realizations with the help of the five Poisson-related cluster-based models and compares them experimentally with the Poisson and negative binomial regressions. The cluster-based models are constructed using the relationship of the count variable with multimodal multivariate Gaussian observations, using their cluster-based discretization. For the discretization, recursive algorithms of the Bayesian mixture estimation theory are applied. Experiments show that the cluster-based models excepting one of them demonstrate improvements in the prediction in the comparison with theoretical counterparts. In practice, the cluster-based Generalized Poisson models seem to be a balanced choice for predicting the considered type of data.

One of the main contributions of the proposed approach consists in the use of real-time continuous data to predict the target count variable described either by the Poisson or Poisson-related distributions. This advantage can be beneficial in specific applications, especially in view of its extension to other count data distributions, which do not have general conditional form.

Acknowledgements. This work was supported by the project Arrowhead Tools, the project number ECSEL 826452 and MSMT 8A19009. The authors would also like to thank Olga Cinková for her brilliant help and fruitful discussions during preparing this manuscript.

References

1. Guenni, L.B.: Poisson Distribution and Its Application in Statistics. In: Lovric, M. (ed.) International Encyclopedia of Statistical Science, pp. 1071–1072. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-04898-2_448
2. Jaggia, S., Kelly, A.: Business Statistics: Communicating with Numbers, 3rd edn. McGraw-Hill Education, New York (2018)
3. Doane, D., Seward, L.: Applied Statistics in Business and Economics, 3rd edn. Mcgraw-Hill, New York (2010)
4. Donnelly, R., Jr.: Business Statistics, 3rd edn. Pearson, London (2019)
5. Anderson, D.R., Sweeney, D.J., Williams, T.A., Camm, J.D., Cochran, J.J.: Essentials of Modern Business Statistics with Microsoft Office Excel (Book Only), 7th edn. Cengage Learning, Boston (2017)
6. Petrouš, M., Suzdaleva, E., Nagy, I.: Modeling of passenger demand using mixture of Poisson components. In: Gusikhin, O., Madani K., Zaytoon J. (eds.) Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2019), pp. 617–624 (2019)
7. Sørensen, Å.L., Lindberg, K.B., Sartori, I., Andresen, I.: Analysis of residential EV energy flexibility potential based on real-world charging reports and smart meter data. Energy Build. **241**, 110923 (2021)
8. Heeringa, S.G., West, B.T., Berglung, P.A.: Applied Survey Data Analysis. Chapman & Hall/CRC, Boca Raton (2010)
9. Falissard, B.: Analysis of Questionnaire Data with R. Chapman & Hall/CRC, Boca Raton (2012)

10. Armstrong, B.G., Gasparrini, A., Tobias, A.: Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis. BMC Med. Res. Methodol. **14**, 122–128 (2014)
11. Agresti, A.: An Introduction to Categorical Data Analysis, 3rd edn. Wiley, Hoboken; New Jersey (2018)
12. Long, J.S., Freese, J.: Regression Models for Categorical Dependent Variables Using Stata, 3rd edn. Stata Press, College Station (2014)
13. Diallo, A.O., Diop, A., Dupuy, J.-F.: Analysis of multinomial counts with joint zero-inflation, with an application to health economics. J. Stat. Plan. Inference **194**, 85–105 (2018)
14. Agresti, A.: Categorical Data Analysis, 3rd edn. Wiley, Hoboken; New Jersey (2012)
15. Haykin, S.: Neural Networks: A Comprehensive Foundation. Macmillan, New York (1994)
16. Congdon, P.: Bayesian Models for Categorical Data. Wiley, Hoboken (2005)
17. Lim, H.K., Li, W.K., Yu, P.L.H.: Zero-inflated Poisson regression mixture model. Comput. Stat. Data Anal. **71**, 151–158 (2014)
18. Počuča, N., Jevtić, P., McNicholas, P.D., Miljkovic, T.: Modeling frequency and severity of claims with the zero-inflated generalized cluster-weighted models. Insur. Math. Econ. **94**, 79–93 (2020)
19. Perrakis, K., Karlis, D., Cools, M., Janssens, D.: Bayesian inference for transportation origin-destination matrices: the Poisson-inverse Gaussian and other Poisson mixtures. J. R. Stat. Soc. A. Stat. Soc. **178**, 271–296 (2015)
20. Yu, J., Gwak, J., Jeon, M.: Gaussian-Poisson mixture model for anomaly detection of crowd behaviour. In: Proceedings of 2016 International Conference on Control, Automation and Information Sciences (ICCAIS), pp. 106–111 (2016). https://doi.org/10.1109/ICCAIS.2016.7822444
21. Zha, L., Lord, D., Zou, Y.: The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data. J. Transp. Saf. Secur. **8**, 18–35 (2016)
22. Silva, A., Rothstein, S.J., McNicholas, P.D., Subedi, S.: A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data. BMC Bioinform. **20**, 394 (2019)
23. Gupta, M.R., Chen, Y.: Theory and Use of the EM Method. (Foundations and Trends(r) in Signal Processing). Now Publishers Inc., Norwell (2011)
24. Li, Y., Sha, Y., Zhao, R.: Poisson prediction of the loss of teachers in high schools. In: Proceedings of 2010 International Conference on Multimedia Technology, Ningbo, China, pp. 1–3 (2010). https://doi.org/10.1109/ICMULT.2010.5629866
25. Bejleri, V., Nandram, B.: Bayesian and frequentist prediction limits for the Poisson distribution. Commun. Stat. Theory Methods **47**(17), 4254–4271 (2018)
26. Petrouš, M., Uglickich, E.: Modeling of mixed data for Poisson prediction. In: Proceedings of IEEE 14th International Symposium on Applied Computational Intelligence and Informatics (SACI 2020), Timisoara, RO, pp. 77–82 (2020). https://doi.org/10.1109/SACI49304.2020.9118836
27. Uglickich, E., Nagy, I., Petrouš, M.: Prediction of multimodal poisson variable using discretization of gaussian data. In: Gusikhin, O., Nijmeijer, H., Madani, K. (eds.) Proceedings of the 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2021), pp. 600–608 (2021). https://doi.org/10.5220/0010575006000608
28. Gupta, A., Mehrotra, K., Mohan, C.K.: A clustering-based discretization for supervised learning. Stat. Probab. Lett. **80**(9–10), 816–824 (2010)
29. Kianmehr, K., Alshalalfa, M., Alhajj, R.: Fuzzy clustering-based discretization for gene expression classification. Knowl. Inf. Syst. **24**, 441–465 (2010)
30. Dash, R., Paramguru, R., Dash, R.: Comparative analysis of supervised and unsupervised discretization techniques. Int. J. Adv. Sci. Technol. **2**(3), 29–37 (2011)

31. Sriwanna, K., Boongoen, T., Iam-On, N.: Graph clustering-based discretization approach to microarray data. Knowl. Inf. Syst. **60**, 879–906 (2019)
32. Kárný, M., et al.: Optimized Bayesian Dynamic Advising: Theory and Algorithms. Springer, London (2006). https://doi.org/10.1007/1-84628-254-3
33. Nagy, I., Suzdaleva, E.: Algorithms and Programs of Dynamic Mixture Estimation. Unified Approach to Different Types of Components. SpringerBriefs in Statistics, Springer, Heidelberg (2017). https://doi.org/10.1007/978-3-319-64671-8
34. Lambert, D.: Zero-Inflated Poisson regression, with an application to defects in manufacturing. Technometrics **34**(1), 1–14 (1992). https://doi.org/10.2307/1269547
35. Zhang, H., Liu, Y., Li, B.: Notes on discrete compound Poisson model with applications to risk theory. Insur. Math. Econ. **59**, 325–336 (2014)
36. Best, D.J., Rayner, J.C.W., Thas, O.: Goodness of fit for the zero-truncated Poisson distribution. J. Stat. Comput. Simul. **77**(7), 585–591 (2007)
37. Yadav, B., et al.: Can Generalized Poisson model replace any other count data models? An evaluation. Clin. Epidemiology Glob. Health **11**, 100774 (2021)
38. Consul, P.C., Famoye, F.: Generalized Poisson regression model. Commun. Stat. Theor. Methods **21**, 89–109 (1992)
39. Hilbe, J.M.: Negative Binomial Regression. Cambridge University Press, Cambridge (2011)
40. Roy, D.: Discrete Rayleigh distribution. IEEE Trans. Reliab. **53**(2), 255–260 (2004)
41. Peterka, V.: Bayesian system identification. In: Eykhoff, P. (ed.) Trends and Progress in System Identification, pp. 239–304. Pergamon Press, Oxford (1981)
42. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B.: Bayesian Data Analysis (Chapman & Hall/CRC Texts in Statistical Science), 3rd edn. Chapman and Hall/CRC, Boca Raton (2013)
43. Kárný, M., Kadlec, J., Sutanto, E.L.: Quasi-Bayes estimation applied to normal mixture. In: Rojíček, J., Valečková, M., Kárný, M., Warwick, K. (eds.) Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing, CZ, Prague, pp. 77–82 (1998)
44. Cohen, A.C.: Estimating the parameter in a conditional Poisson distribution. Biometrics **16**(2), 203–211 (1960)
45. Consul, P.C.: Generalized Poisson Distributions: Properties and Applications. Marcel Dekker, New York (1989)
46. Date, S.: Generalized Poisson Regression for real world datasets. https://towardsdatascience.com. Accessed 4 Apr 2020
47. Siddiqui, M. M.: Statistical inference for Rayleigh distributions. J. Res. Natl. Bureau Stand. Sect. D Radio Sci. **68D**(9), 1007 (1964)
48. Date, S.: Negative Binomial Regression: A Step by Step Guide. https://towardsdatascience.com. Accessed 6 Oct 2019
49. Lewis, D.D.: Naive (Bayes) at forty: the independence assumption in information retrieval. In: Nédellec, C., Rouveirol, C. (eds.) ECML 1998. LNCS, vol. 1398, pp. 4–15. Springer, Heidelberg (1998). https://doi.org/10.1007/BFb0026666