The 12th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
7-9 September, 2023, Dortmund, Germany

Count Predictive Model with Mixed Categorical and Count Explanatory Variables

Evženie Uglickich[1] , Ivan Nagy[1,2] , Tetiana Reznychenko[2]
[1] Department of Signal Processing
Institute of Information Theory and Automation of the Czech Academy of Sciences
Pod vodárenskou věží 4, 18208 Prague, Czech Republic, uglickich@utia.cas.cz, https://orcid.org/0000-0003-1764-5924
[2] Faculty of Transportation Sciences, Czech Technical University
Na Florenci 25, 11000 Prague, Czech Republic
nagy@fd.cvut.cz, https://orcid.org/0000-0002-7847-1932, tetiana.reznychenko@cvut.cz

***Abstract* — The paper considers the problem of online prediction of a count variable based on real-time explanatory data of mixed count and categorical nature. The presented solution is based on (i) recursive Bayesian estimation of a mixture model of Poisson-distributed explanatory counts, using the categorical explanatory variable as a measurable pointer of the mixture, (ii) construction of a mixture of local Poisson regressions on the clustered data, and (iii) use of the pre-estimated mixtures for online prediction of the target count using actual measured explanatory data. The latter is one of the main contributions of the proposed approach. In addition, the dynamic model of the categorical explanatory variable preserves the functionality of the algorithm in case of its measurement failure. The experiments with simulations and real data report lower prediction errors compared to theoretical counterparts.**

***Keywords* — count data; Poisson mixtures; Poisson regression; recursive Bayesian mixture estimation**

I. Introduction

Count data is a type of discrete data whose values are non-negative integers generated by counting specific events. Unlike ordinal categorical data, the number of possible realizations of a count variable may be relatively high and may not be known. Because of these specific features, statistical methods for analyzing count data differ from those generally used for discrete data [1], [2]. In practice, predicting/estimating a target count variable depending on explanatory data is a highly desirable task in application fields, where random independent events are observed with a constant intensity per time unit. A significant part of the applications belong to transportation sciences, where count variables are numerously represented (e.g., counts of vehicles, pedestrians, cyclists, passengers, etc.). For example, the number of vehicles on a key section of urban road should be predicted based on data from explanatory sections to avoid measurement errors and improve driving conditions in the city. In safety research, counts of independent events per time unit (accidents, aircraft shutdowns, server virus attacks, crimes, terroristic attacks, etc.) should be analyzed. Other applications include social sciences and finance (customer counts, website visits, bankruptcies), medicine (patient counts, specific diagnoses), etc.

A count random variable is generally described by the Poisson distribution [3] or, in specific cases, by certain Poisson-related distributions. The zero-inflated Poisson model [4] and compound Poisson distributions [5] are used for count data with a high number of zeros, while the zero-truncated Poisson distribution with a minimum at 1 can be applied to data without zeros [6].

For predicting a count variable based on explanatory data, one of the basic approaches is Poisson regression [7]–[12]. However, in reality, counts do not often satisfy the Poisson assumption of the equality of mean and variance, which means that the overdispersion or underdispersion of the data is observed. In this case, the trivial use of Poisson regression does not allow to obtain an accurate model of the target variable, which leads to a higher prediction error. To cope with over- or underdispersed data, the following approaches are used.

Negative binomial regression (NB) [13] assumes that the target count variable is described by the NB distribution instead of the Poisson one. It uses additional dispersion parameters estimated by auxiliary least squares to describe the variance as a function of mean. This model is widespread and relatively successful. However, studies [14], [15] report sensitivity to small and large counts.

Mixture models are known as the universal approximation of nonlinear relation between variables [16]. The target counts can be described by a mixture of Poisson

This work was supported by the StorAIge project and the corresponding Czech institutional support project No. 8A21009. This project has received funding from the ECSEL Joint Undertaking (JU) under grant agreement No 101007321. The JU receives support from the European Union's Horizon 2020 research and innovation program and France, Belgium, Czech Republic, Germany, Italy, Sweden, Switzerland, Turkey.

distributions [17], which does not require meeting the equidispersion. However, it does not solve the prediction from explanatory data. Studies dealing with mixtures of Poisson regressions [18], [19], Poisson-gamma models [20] and negative binomial mixtures [17] were found. Their estimation is solved mainly via the iterative expectation-maximization (EM) algorithm [21] assuming offline computations during the data analysis.

Models of mixed distributions is a similar approach, describing the target and explanatory counts by the mixture of the Poisson and Gaussian distributions [22]–[24], etc. Again, to the best of our knowledge, they are estimated offline using the EM algorithm, even though the attempts of the online estimation can be found in [25] based on the recursive Bayesian mixture estimation [26]–[28].

Generalized Poisson Models (GPM), namely, the Consul's GPM and Famoye's GPM [29], are suitable for modeling both the overdispersed and underdispersed count data [30]. They also assume the variance to be a function of mean via dispersion parameters. However, they do not possess a closed form for recursive computations.

The presented paper approaches this problem using paradigms of the recursive Bayesian mixture estimation theory [26]–[28]. It proposes the algorithm of online prediction of a count variable based on real-time explanatory data of mixed count and categorical nature. The presented solution is based on (i) recursive Bayesian estimation of a mixture model of Poisson-distributed explanatory counts, using the categorical explanatory variable as a measurable pointer of the mixture, (ii) construction of a mixture of local Poisson regressions on the clustered data, and (iii) use of the pre-estimated mixtures for online prediction of the target count using actually measured explanatory data. The latter is one of the main contributions of the proposed approach. In addition, the dynamic model of the categorical explanatory variable preserves the functionality of the algorithm in case of its measurement failure. The experiments with simulations and real data report a higher accuracy of the online prediction compared to the Poisson and NB regressions in the offline mode.

The layout of the paper is organized as follows: Section II-A formulates a problem and introduces the models. Section III presents the methodology. Section IV provides results of experiments. Section V provides conclusions.

II. Mixture Model of Count Data

A. Problem Formulation

Let us consider a system, which generates values of the *target* count variable y_t, the vector of *explanatory* independent count variables $x_t = [x_{1;t}, \ldots, x_{N_x;t}]$, and the *explanatory* dynamic categorical variable $z_t \in \{1, 2, \ldots, N_z\}$ at each time instant $t = 1, 2, \ldots, T$. The measurements of the variables provide data sets $\{y_t\}_{t=1}^{T}$, $\{x_t\}_{t=1}^{T}$ and $\{z_t\}_{t=1}^{T}$.

The main task is to (i) construct a model describing the relationship of the target count y_t and the explanatory multivariate count x_t and categorical variable z_t, (ii) predict the values of y_t online based on the explanatory data measured at each time t.

B. Models

In order to describe the relationship of the target count y_t and all explanatory variables x_t, z_t, the joint probability function (pf) of these variables is decomposed into the product of the following conditional pfs

$$f(y_t, x_t, z_t, z_{t-1}) = f(y_t|x_t, z_t) f(x_t|z_t) f(z_t|z_{t-1}), \quad (1)$$

assuming that y_t and x_t are independent of z_{t-1}.

In this paper, the last of these pfs is specified as the dynamic parametrized model in the form of the conditional categorical distribution

$$f(z_t|\beta, z_{t-1}) = \begin{array}{|c|c|c|c|} \hline z_{t-1} \backslash z_t & 1 & \ldots & N_z \\ \hline 1 & \beta_{1|1} & \ldots & \beta_{1|N_z} \\ \hline \ldots & \ldots & \ldots & \ldots \\ \hline N_z & \beta_{N_z|1} & \ldots & \beta_{N_z|N_z} \\ \hline \end{array}, \quad (2)$$

where parameters $\beta \equiv \{\beta_{i|j}\}_{i,j=1}^{N_z}$ are probabilities of $z_t = i$ under condition that $z_{t-1} = j$.

The pf $f(x_t|z_t)$ from (1) describes the explanatory multivariate count variable x_t depending on z_t. The Poisson distribution is a widely used model for describing count data. However, it is known that it does not have a general conditional form that directly covers the dependence of the variables. In this paper, the dependence of x_t on z_t in the pf $f(x_t|z_t)$ from (1) is considered as a mixture of the Poisson components existing for each independent count $x_{l;t}$ $\forall l \in \{1, 2, \ldots, N_x\}$ and each realization of z_t, i.e.,

$$f(x_{l;t}|\lambda_l, z_t = i) = e^{-\lambda_{l,i}} \frac{\lambda_{l,i}^{x_{l;t}}}{x_{l;t}!}, \quad i = \{1, \ldots, N_z\}, \quad (3)$$

where the parameter $\lambda_l \equiv \{\lambda_{l,i}\}_{l,i=1}^{N_x, N_z}$ is equal to $\lambda_{l,i}$ for $x_{l;t}$ under condition that $z_t = i$. This makes the measurable categorical variable z_t a known pointer [26] of the Poisson mixture, which significantly simplifies a derivation of the mixture estimation algorithm.

The pf $f(y_t|x_t, z_t)$ from (1) is the main focus of the study. It describes the relationship between the target y_t and the explanatory counts $x_{l;t}$, each of which changes its behavior for each value of z_t. This leads to using the mixture of Poisson regressions that exist in the data space of each $x_{l;t}$. Thus, the pf $f(y_t|x_t, z_t)$ is specified as

$$f(y_t|x_t, \theta, z_t = i) = \frac{e^{y_t \theta_i' x_t}}{y_t!} e^{-e^{\theta_i' x_t}}, \quad (4)$$

which means that the Poisson regression with the expectation of y_t is

$$\mathrm{E}\left[y_t|x_t, \theta, z_t = i\right] = e^{\theta_i' x_t}, \ \forall i \in \{1, \ldots, N_z\} \quad (5)$$

with $\theta = \{\theta_i\}_{i=1}^{N_z}$, and θ_i is a set of regression coefficients. In this paper, the Poisson regression (5) will be used for the point prediction of y_t instead of the distribution (4).

Using the introduced models, the solution to the problem formulated in Section II-A is presented below.

III. Methodology

A. Model Estimation of Explanatory Data

The first phase of the solution focuses on the analysis of the explanatory data x_t, z_t. The main objective here is to estimate the parameters of the Poisson components (3) that exist for each value of z_t and to find the data $x_{l;t}$ belonging to these components. As for the estimation of the parameters of (2), it is needed for the possibility of predicting z_t in case of missing data.

The posterior pf of the unknown parameters of (2) and (3) is derived using the Bayes and chain rules [31] according to the recursive Bayesian mixture estimation methodology [26]–[28], i.e.,

$$\begin{aligned} & f\left(\lambda_l, \beta | \{x_{l;t}\}_{t=0}^T, \{z_t\}_{t=0}^T\right) \\ & \propto f\left(\lambda_l, \beta, x_{l;t}, z_t = i | \{x_{l;t}\}_{t=0}^{T-1}, \{z_t\}_{t=0}^{T-1}\right) \\ & = f\left(x_{l;t} | \lambda_l, z_t = i\right) f\left(\lambda_l | \{x_{l;t}\}_{t=0}^{T-1}, z_t = i\right) \\ & \times f\left(z_t = i | \beta, \{z_t\}_{t=0}^{T-1}\right) f\left(\beta | \{z_t\}_{t=0}^{T-1}\right). \end{aligned} \tag{6}$$

The right side of (6) is the product of (i) the Poisson pf (3), (ii) the conjugate prior Gamma pf for the estimation of λ_l [32], (iii) the categorical model (2) and (iv) the conjugate prior Dirichlet distribution $f(\beta|\{z_t\}_{t=0}^{T-1})$ used for the recursive estimation of β according to [27]. According to the adopted theory [26]–[28], the relation (6) is marginalized over the unknown parameters. This marginalization leads to a recursive update of the statistics of the involved models with the subsequent computation of the point estimates of the parameters.

The statistics of the categorical model (2) is actualized according to [27] $\forall i, j \in \{1, \ldots, N_z\}$ as follows:

$$\nu_{i|j;t} = \nu_{i|j;t-1} + \delta(z_t | z_{t-1}, i|j), \tag{7}$$

where $\nu_t \equiv \{\nu_{i|j;t}\}_{t=1}^T$ is the statistics of the dimension identical to (2), and ν_0 is the random initial statistics. The Kronecker delta function $\delta(z_t|z_{t-1}, i|j)$ is equal to 1, if $z_t = i$ and $z_{t-1} = j$, and it is 0 otherwise. The updated statistics ν_t is normalized to obtain the point estimate of the parameter β [26], [27], i.e.,

$$\hat{\beta}_{i|j;t} = \frac{\nu_{i|j;t}}{\sum_{k=1}^{N_z} \nu_{k|j;t}}. \tag{8}$$

As for the Poisson pfs (3), the recursive update of their statistics has been discussed in detail in [33]. With z_t as the known pointer, this update is straightforward

$$S_{l,i;t} = S_{l,i;t-1} + \delta(z_t, i) x_{l;t}, \ \kappa_{l,i;t} = \kappa_{l,i;t-1} + \delta(z_t, i) \tag{9}$$

starting with the initial statistics $S_{l,i;0}$ and $\kappa_{l,i;0}$ of the i-th Poisson pf (3) of the count $x_{l;t}$ labeled by the value $z_t = i$. The Kronecker delta function in (9) is defined similarly. The point estimate of the parameter $\lambda_{l,i}$, of each component (3) of each $x_{l;t}$ is computed as

$$\hat{\lambda}_{l,i;t} = \frac{S_{l,i;t}}{\kappa_{l,i;t}}, \ \forall l \in \{1, \ldots, N_x\}, \forall i \in \{1, \ldots, N_z\}, \tag{10}$$

which gives the recursive version of estimating the Poisson expectation.

In this way, this part of the solution is applied to the data sets $\{x_t\}_{t=0}^T, \{z_t\}_{t=0}^T$ measured up to the time $t = T$. Its results are the identified components of each explanatory count $x_{l;t}$.

Remark: With z_t as the known pointer, the point estimate (10) is optional or can be done offline. However, it serves as preparation for the case of an unmeasured pointer.

B. Local Poisson Regressions on Data in Components

The second phase of the solution runs completely offline. The data of each $x_{l;t}$ belonging to its components detected above are substituted into the vector x_t in (5) to obtain the expectation θ_i of the Poisson regression for the i-th value of z_t. This is done via the maximum likelihood (ML) estimation of the parameters θ_i of N_z *local* Poisson regressions using numerical methods [7].

C. Target Count Online Prediction

The third phase is the online prediction of the target count y_t for the time $t > T$, where its values are no longer measured, but x_t, z_t are permanently available. Here, the value of z_t actually measured at time t indicates which of the pre-estimated local Poisson regressions (5) is to be used for the prediction. Then the current observations of all counts $x_t = [x_{1;t}, \ldots, x_{N_x;t}]$ at time t are fed into the given local model (5). The point prediction of the target count y_t is computed via (5) using the point estimate of θ_i corresponding to the value of z_t.

Remark 1: If the measurement of z_t fails, its point prediction obtained via the maximum probability in the j-th row corresponding to the past $z_{t-1} = j$ of the estimated model (2) is used instead of a missing value of z_t, i.e.,

$$\hat{z}_t = \arg\max[\hat{\beta}_{1|j;t}, \ldots, \hat{\beta}_{N_z|j;t}], \ \ j = z_{t-1}. \tag{11}$$

Remark 2: In case of missing values of x_t, the point estimates of $\lambda_{l,i}$ are used instead of them.

The algorithm is summarized below.

D. Algorithm

{Initialization $(t = 0)$}
Set the number of components $N_z = \max(\{z_t\}_{t=0}^T)$.
Set the initial statistics ν_t.
for all $l \in \{1, \ldots, N_x\}$ **do**
  for all $i \in \{1, \ldots, N_z\}$ **do**
    Set the initial statistics $S_{l,i;0}$, $\kappa_{l,i;0}$.
  end for
end for
{Model Estimation of Explanatory Data)}

for $t = 1, 2, \ldots, T$ **do**
  Measure the current values of z_t and x_t.
  for all $l \in \{1, \ldots, N_x\}$ **do**
    for all $i \in \{1, \ldots, N_z\}$ **do**
      Update ν_t using (7), $S_{l,i;t}$ and $\kappa_{l,i;t}$ via (9).
      Compute $\hat{\beta}_t$ using (8) and $\hat{\lambda}_{l,i;t}$ via (10).
    end for
  end for
end for
$\{$Local Poisson Regressions on Data in Components)$\}$
for all $z_t = i \in \{1, \ldots, N_z\}$ **do**
  Filter the values of x_t and substitute them into (5).
  Get the ML estimate of θ_i numerically.
end for
$\{$Target Count Online Prediction)$\}$
for $t = T+1, T+2, \ldots,$ **do**
  Measure the current value of z_t, x_t.
  if z_t is NaN **then**
    Use its prediction $\hat{z}_t$ via (11) instead.
  end if
  For $i = z_t$ use (5) to get the point prediction of y_t.
end for

The algorithm was implemented in the free and open source programming environments Scilab (www.scilab.org) and Python 3 (www.python.org).

IV. Experiments

A. Experiments with Simulations

The functionality of the algorithm was first validated with simulations. To test the target count prediction, data sets with a capacity of 3.000 values were simulated during each run of the algorithm for $N_z = 4$, $N_x = 2$ using the dynamic categorical model, Poisson distributions with random parameters, and Poisson regressions. Fig. 1 (top) shows histograms of explanatory and target counts from one of the data sets. Note that while the multimodal nature of the explanatory data is clearly visible in the figure, the shape of the histogram of the target counts can be much smoother. This explains why the first phase of the solution (see Section III-A) is needed instead of searching for clusters in the data space of y_t.

80% and 20% of each data set from each algorithm run were used for training and testing. Fig. 1 (bottom) shows clusters of explanatory count data for each z_t value.

To evaluate the contribution of the proposed algorithm, Poisson regression and NB regression were applied to the training and testing data sets with all explanatory variables including z_t, i.e., with $[x_{1;t}, \ldots, x_{N_x;t}, z_t, \ 1]$. The prediction accuracy was evaluated using the root mean square error (RMSE), mean absolute error (MAE), mean squared logarithmic error (MSLE) and negative log likelihood (NLL). The comparison of the average accuracy metrics is shown in Table I, which reports improvements in prediction accuracy for all metrics compared. A fragment of the visual comparison of all predictions and the absolute deviations between simulation and predictions are shown in Fig. 2. 100 and 50 values are shown for better visibility. Note that for most values, the proposed algorithm has the smallest deviations from the simulated values.

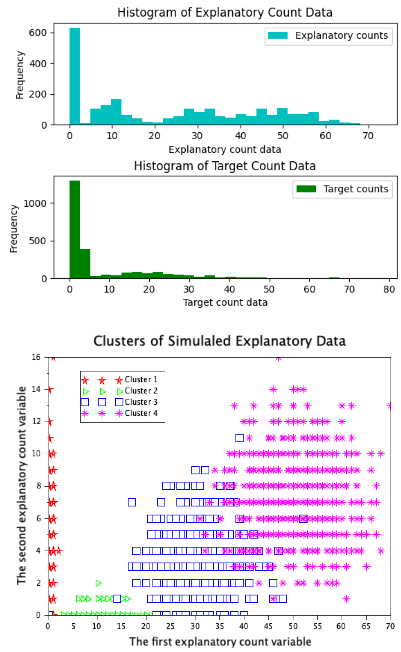

Figure 1. Simulated count data histograms (top) and explanatory count data clusters (bottom)

Table I. Comparison of Average Accuracy Metrics

	RMSE	MAE	MSLE	NLL
Proposed algorithm	2.601	1.656	0.242	1120.78
Poisson regression	3.249	2.165	0.402	1338.864
NB regression	4.445	2.656	0.422	1429.255

In the testing data, 3% of the values of the variable z_t were randomly replaced by NaN to simulate missing explanatory categorical data. In practice, such a situation can cause the algorithm to fail, since the values of z_t indicate the model to be used for prediction. Testing of the algorithm with random one-time failures and single longer failures gave accuracy in predicting missing data in the range of 70% to 97%. The accuracy of predicting target counts with missing data remained at the same level for all models compared.

B. Experiments with Real Traffic Count Data

The potential application is expected mainly in real-time traffic flow prediction for intelligent transportation systems. To validate the algorithm with real data, the data set [34] of traffic counts collected at New York City bridge crossings and roadways from 2014 to 2020 was used to test the target count prediction. The capacity of the data set was 33.409 values. Hourly traffic counts have been

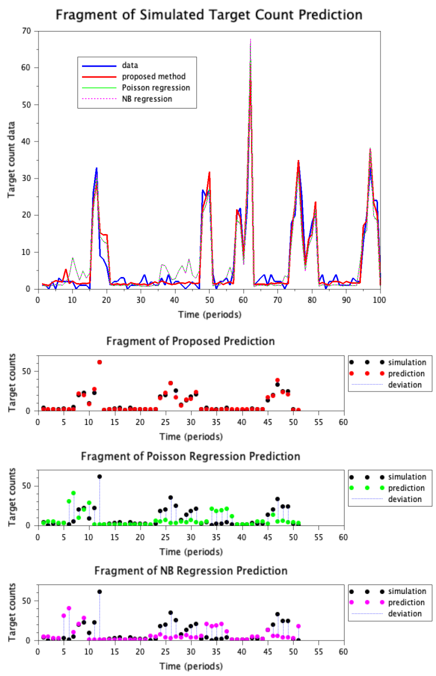

Figure 2. Visual comparison of predictions (top) and absolute deviations between simulation and prediction (bottom).

recalculated to minutes. The goal of the experiments was to predict the target traffic counts at the roadways based on the two explanatory count variables from the roadways and the categorical traffic state. The latter was unavailable in the data set and was obtained by discretizing the third explanatory count variable into 4 quartile-based values representing traffic states: "free flow", "stable flow", "approaching unstable flow", and "unstable flow". Traffic state from Google's live traffic maps can also be used.

Two types of experiments were performed: with original time series data and with randomly shuffled data. The prediction accuracy of the target traffic counts obtained with all compared methods is shown in Table II for the case of time series data and in Table III for randomly shuffled data. It can be seen that all accuracy metrics of all methods increased in the case of shuffled data. However, in both types of experiments the proposed prediction algorithm demonstrates higher accuracy than the compared methods.

Table II. Time Series Traffic Counts Prediction Accuracy

	RMSE	MAE	MSLE	NLL
Proposed algorithm	3.109	1.901	0.1454	15408.572
Poisson regression	6.837	2.928	0.271	18840.114
NB regression	6.512	2.815	0.262	18725.097

Fig. 3 shows a fragment of the prediction of the time series (top) and randomly shuffled traffic counts (bottom). Note that although the deviations of the time

Table III. Average Shuffled Traffic Counts Prediction Accuracy

	RMSE	MAE	MSLE	NLL
Proposed algorithm	7.082	3.025	0.179	22830.282
Poisson regression	10.516	4.177	0.295	27321.901
NB regression	10.777	4.129	0.287	27191.462

series predictions in the top plot appear visually higher, all methods follow the expectations of the counts, and looking at the y-axis, it can be seen that the deviations are not significant. The shuffled data deviations in the bottom plot are much higher (see the y-axis values), but the predictions appear visually more accurate. This explains the difference in accuracy between Tables II and III. This is also confirmed by comparing the absolute deviations between simulation and prediction in Fig. 4 (see values on the y-axis).

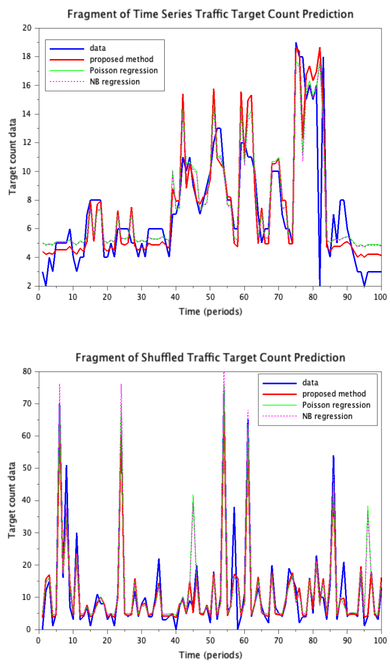

Figure 3. Fragments of traffic count prediction comparison.

V. Conclusion

The study's goal of validating online count prediction using simulations and real data was successfully achieved: lower prediction errors are reported compared to theoretical counterparts. In addition, simulated missing explanatory data did not affect the prediction accuracy due to the dynamic categorical model, which allows the use of predictions instead of missing values.

Note that the variable z_t used as the pointer is an auxiliary explanatory variable. If it cannot be measured, the problem becomes recursive Bayesian mixture estimation with Poisson components and regressions, where the data

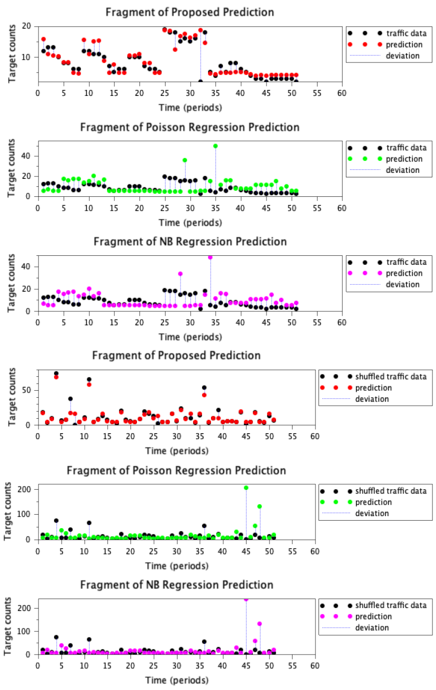

Figure 4. Fragments of absolute deviations comparison.

classification is based on estimating the pointer. This case will be discussed in a separate publication.

Acknowledgment

The authors thank to Olga Cinková for her brilliant help and fruitful discussions during preparing the paper.

References

[1] A. C. Cameron and P. K. Trivedi, *Regression Analysis of Count Data Book* (Second ed.). Cambridge University Press, 2013.
[2] R. Winkelmann, *Econometric Analysis of Count Data* (5th ed.). Springer, 2008.
[3] L. B. Guenni, "Poisson Distribution and Its Application in Statistics", In Lovric M. (eds) *International Encyclopedia of Statistical Science*, Springer, Berlin, Heidelberg, 2011.
[4] D. Lambert, "Zero-Inflated Poisson regression, with an application to defects in manufacturing", *Technometrics*, 1992, 34(1) 1–14.
[5] H. Zhang, Y. Liu, B. Li, "Notes on discrete compound Poisson model with applications to risk theory", *Insurance: Mathematics and Economics*, 2014, 59, p.325-336.
[6] D. J. Best, J. C. W. Rayner, O. Thas, "Goodness of fit for the zero-truncated Poisson distribution", *Journal of Statistical Computation and Simulation*, 2007, 77(7) 585–591.
[7] S.G. Heeringa, B.T. West, P.A. Berglung, *Applied Survey Data Analysis*, Chapman & Hall/CRC, 2010.
[8] B. Falissard, *Analysis of Questionnaire Data with R*. Chapman & Hall/CRC, Boca Raton, 2012.
[9] B.G. Armstrong, A. Gasparrini, A. Tobias, "Conditional Poisson models: a flexible alternative to conditional logistic case cross-over analysis", *BMC Medical Research Methodology*, 2014, 14(122).
[10] A. Agresti, *An Introduction to Categorical Data Analysis*. 3rd Ed. Wiley, 2018.
[11] J. S. Long and J. Freese, *Regression Models for Categorical Dependent Variables Using Stata*. 3rd Ed. Stata Press, 2014.
[12] A. O. Diallo, A. Diop, J.-F. Dupuy, "Analysis of multinomial counts with joint zero-inflation, with an application to health economics", *Journal of Statistical Planning and Inference*, 2018, 194, 85-105.
[13] J. M. Hilbe, *Negative Binomial Regression*, Cambridge University Press, 2011.
[14] R. Berk and J. M. MacDonald, "Overdispersion and Poisson regression", *Journal of Quantitative Criminology*, 2008, 24(3), 269-284.
[15] J.M. ver Hoef and P.L. Boveng, "Quasi-Poisson versus negative binomial regression: how should we model overdispersed count data", *Ecology*, 2007, 88, 2766-2772.
[16] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan, New York, 1994.
[17] P. Congdon, *Bayesian Models for Categorical Data*, John Wiley & Sons, 2005.
[18] H. K. Lim, W. K. Li, P. L.H. Yu, "Zero-inflated Poisson regression mixture model", *Computational Statistics & Data Analysis*, 2014, 71, p.151–158.
[19] N. Počuča, P. Jevtić, P.D. McNicholas, T. Miljkovic, "Modeling frequency and severity of claims with the zero-inflated generalized cluster-weighted models", *Insurance: Mathematics and Economics*, 2020, 94, p.79–93.
[20] A. Agresti, *Categorical Data Analysis. 3rd Ed.*, John Wiley & Sons, 2012.
[21] M. R. Gupta and Y. Chen, *Theory and Use of the EM Method. (Foundations and Trends(r) in Signal Processing)*, Now Publishers Inc., 2011.
[22] L. Zha, D. Lord, Y. Zou, "The Poisson inverse Gaussian (PIG) generalized linear regression model for analyzing motor vehicle crash data", *Journal of Transportation Safety & Security*, 2016, 8, p.18–35.
[23] A. Silva, S.J. Rothstein, P.D. McNicholas, S. Subedi, "A multivariate Poisson-log normal mixture model for clustering transcriptome sequencing data", *BMC Bioinformatic*, 2019, 20(394).
[24] V. Bejleri and B. Nandram, "Bayesian and frequentist prediction limits for the Poisson distribution", *Communications in Statistics – Theory and Methods*, 2018, 47(17), p.4254–4271.
[25] E. Uglickich, I. Nagy, M. Petrouš, "Prediction of multimodal Poisson variable using discretization of Gaussian data", In Proceedings of the 18th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2021), 2021, pp. 600–608.
[26] M. Kárný, J. Kadlec, and E.L. Sutanto, "Quasi-Bayes estimation applied to normal mixture," In J. Rojíček, M. Valečková, M. Kárný, and K. Warwick, K. Eds. Preprints of the 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing, Prague, CZ, 1998, pp. 77–82.
[27] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař, *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, London, 2006.
[28] I. Nagy and E. Suzdaleva, *Algorithms and Programs of Dynamic Mixture Estimation. Unified Approach to Different Types of Components*, SpringerBriefs in Statistics. Springer International Publishing, Heidelberg, 2017.
[29] P.C. Consul and F. Famoye, "Generalized Poisson regression model", *Commun Stat Theor Methods*, 1992, **21** 89–109.
[30] B. Yadav, L. Jeyaseelan, V. Jeyaseelan, J. Durairaj, S. George, K.G. Selvaraj, S. I. Bangdiwala, "Can Generalized Poisson model replace any other count data models? An evaluation", *Clinical Epidemiology and Global Health*, 2021, 11, 100774.
[31] V. Peterka, "Bayesian system identification," In Eykhoff, P. (Ed.), Trends and Progress in System Identification. Oxford, Pergamon Press, 1981, pp. 239–304.
[32] V. HyvPönen and T. Tolonen, "Bayesian Inference 2019," https://vioshyvo.github.io/Bayesian_inference/. Last accessed March 2019.
[33] E. Uglickich and I. Nagy, "Recursive mixture estimation with univariate multimodal Poisson variable," Research Report 2394. ÚTIA AV ČR. Prague, 2022, http://library.utia.cas.cz/separaty/2022/ZS/uglickich-0557467.pdf
[34] Department of Transportation (DOT), NYC OpenData. "Traffic Volume Counts", https://data.cityofnewyork.us/Transportation/Traffic-Volume-Counts/btm5-ppia, last access May 26, 2022.