Rayleigh model fitting to nonnegative discrete data

Matej Petrouš
Faculty of Transportation Sciences
Czech Technical University
Na Florenci 25, 11000 Prague, Czech Republic
Department of Signal Processing
Czech Academy of Sciences
Institute of Information Theory and Automation
Pod vodárenskou věží 4, 18208 Prague, Czech Republic
petromat@fd.cvut.cz

Evženie Uglickich
Department of Signal Processing
The Czech Academy of Sciences
Institute of Information Theory and Automation
Pod vodárenskou věží 4, 18208 Prague, Czech Republic
https://orcid.org/0000-0003-1764-5924

***Abstract*—The paper deals with modeling ordinal discrete random variables with a high number of nonnegative realizations. The prediction of the Rayleigh distribution learned on clusters of the explanatory variables is proposed. The proposed solution consists of the clustering and estimation phases based on the knowledge both of the target and explanatory variables, and the prediction phase using only the information from the explanatory variables. The main contributions of the approach are: (i) using the discretized knowledge of clusters of the explanatory variables and (ii) describing nonnegative discrete data by the multimodal Rayleigh distribution. Experiments with a data set from a tram network are provided.**

***Index Terms*—Poisson distribution, multimodal data, Rayleigh distribution, recursive estimation, passenger demand**

I. Introduction

The paper deals with modeling ordinal discrete random variables with a high number of possible nonnegative realizations. Models of discrete data can be met in various application fields, for example, marketing [1], medicine [2], queuing theory [3], passenger demand modeling [4], [5], accident analysis [6], etc. The extensive list of application fields shows that novel solutions for analysis of discrete data are highly desired. Working with multidimensional discrete variables causes their models being the tables of an enormous dimension [7], which may be problematic from the computational point of view. In the case of the specific type of data with a high number of realizations discussed in the paper, the problem of dimensionality is even more complicated. Therefore, using a continuous distribution to fit such large-scale discrete data opens a way to simplify the calculations [8]. This paper focuses on using the Rayleigh distribution.

A series of papers solving similar issues have been found. The study [9] discussed the problem of distribution fitting for a railway delay data set with a high number of possible discrete realizations. Discussed distributions included normal, exponential, log-normal, etc. Methods of distribution fitting based on the maximum likelihood approach and statistical hypothesis testing were described with a brief summary of their advantages and drawbacks.

This work has been supported by the project SILENSE, project number ECSEL 737487 and MSMT 8A17006.

The paper [10] proposed distribution fitting to discrete data using the algebraic algorithm of reconstructing the ellipsoids with discrete points with the help of the convex hull and covariance matrix methods. It was shown that the convex hull method was more sensitive to outliers and deviations from the normal distribution than the covariance matrix method. Moreover, it was observed that the convex hull method may produce inconsistent estimates for non-Gaussian data.

The study [11] is closer to the area discussed in the presented paper; it proposed the new two-parameter discrete version of the continuous generalized Rayleigh distribution. The parameter estimation has been solved using the method of proportions, method of moments, and maximum likelihood estimation, which have been compared using Monte Carlo simulations. In the work [12], the Rayleigh distribution, claimed as one of the most suitable distributions for histogram of underwater images, was applied to stretching the histogram to improve image quality.

In the presented paper, the Rayleigh distribution is considered for fitting the target discrete variable, which is modeled by means of explanatory variables. Their relationship could be described with the help of a regression. To describe nonnegative discrete variables with a high number of realizations, the Poisson distribution is often assumed. Hence, in the case of the Poisson distributed variables, the Poisson regression can be used [13], [14]. Having a multimodal system, which generates the Poisson data, a mixture of Poisson regressions seems to be a suitable tool [15]. However, in the applications where the estimation should be performed online, such as, e.g., passenger demand modeling, the mixture estimation with the Poisson regressions often fails because of the hardly initialized components. This is known to be a complicated task also for other distributions aimed at nonnegative data, such as, e.g., exponential [16], [17]. One of the possibilities is to use the approximation of the Poisson probability density function (pdf) by the Gaussian pdf with keeping the assumption of equal mean and variance, see, e.g., [18]. However, this also brings a series of limitations regarding the prediction accuracy.

The high number of possible values of the discrete random variable allows us to attempt to describe it using a contin-

978-1-7281-1059-2/20/$31.00 ©2020 IEEE

uous distribution. This paper proposes to use the Rayleigh distribution learned on clusters of explanatory variables for estimating and predicting nonnegative data. This choice is given primarily by the similar shape both of the Poisson and Rayleigh pdfs designed for nonnegative data. Moreover, the Rayleigh distributions are closer to the Gaussian pdfs, mixtures of which are known to be a universal approximation of nonlinear system behavior [19]. In addition, the assumption of the equal mean and variance necessary in the case of the Poisson distribution is difficult to be kept in practice [14]. The Rayleigh distribution is not restricted by this assumption, although the values of mean and variance are supposed to be close to each other. Along with the pdf shape, this gives a chance to fit the Rayleigh model to the considered type of data. The proposed solution enables to avoid the Poisson (or logistic) regression with the help of clustering the explanatory variables. The paper demonstrates the theoretical background of the solution and provides illustrative experiments.

The layout of the paper is organized as follows. Section II-A introduces the models and specifies the task to be solved. Section II-B presents the algorithm of clustering the explanatory variables, estimating the predictive models, and predicting the target variable. Experiments with real data along with a discussion can be found in Section III. Section IV provides conclusions.

II. Rayleigh Model Fitting to Discrete Data

A. Models

Let us have a target discrete variable $y_t \in \{1, 2, \ldots, n_y\}$ and the explanatory continuous multidimensional variable x_t, which can be measured on a multimodal system, where the subscript t denotes discrete time instants. In general, the relationship of the variables can be described by the following joint pdf

$$f(y_t = j, x_t|\mathcal{P}) = f(y_t = j|x_t, \mathcal{P})f(x_t|\mathcal{P}), \qquad (1)$$

where $j \in \{1, 2, \ldots, n_y\}$ and $\mathcal{P}$ generally denotes a set of unknown mutually independent parameters both of the pdfs in the right hand side of (1). The multimodality of the observed system supposes that the behavior of the variables also depends on the discrete pointer variable $c_t \in \{1, 2, \ldots, n_c\}$ [20], which expresses switching the system modes, i.e.,

$$f(y_t = j, x_t, c_t = i|\mathcal{P})$$
$$= f(y_t = j|x_t, c_t = i, \mathcal{P})f(x_t|c_t = i, \mathcal{P})f(c_t = i|\mathcal{P}), \qquad (2)$$

where $i \in \{1, 2, \ldots, n_c\}$ and the set of parameters $\mathcal{P}$ should be completed by the parameters of the pointer model. In reality, the pointer c_t is often unmeasured and should be estimated, which leads to the task of clustering the data.

As it was mentioned in Section I, modeling the discrete variable y_t depending on the continuous variable x_t is significantly complicated in the case of the online estimation of the multimodal data model leading to the use of the Poisson (or logistic) regression. In order to avoid it, the clusters of the variable x_t instead of the data x_t are assumed to be used in the pdf $f(y_t = j|x_t, c_t = i, \mathcal{P})$ via the estimated values of c_t only. This means that the variable y_t depends only on the cluster detected at time t. Under this assumption, the joint pdf (1) finally takes the form

$$f(y_t = j|c_t = i, \sigma)f(x_t|c_t = i, \Theta)f(c_t = i|\alpha), \qquad (3)$$

where $\mathcal{P} = \{\sigma, \Theta, \alpha\}$, which are mutually independent parameters of the involved pdfs.

In this paper, the models in the pdf (3) are specified as follows. As the variable y_t can have the high number of realizations on the interval from 0 to n_y, we propose to describe it by the Rayleigh distribution

$$f(y_t|c_t = i, \sigma) = \frac{y_t}{\sigma_i^2}\exp\{-\frac{y_t^2}{2\sigma_i^2}\} \qquad (4)$$

where the parameter $\sigma \equiv \sigma_i$ under condition that $c_t = i$.

The explanatory variable x_t is described by the multivariate normal distribution existing for each i

$$f(x_t|c_t = i, \Theta)$$
$$= (2\pi)^{-n_x/2}|r_i|^{-1/2}\exp\left\{-\frac{1}{2}[x_t - \theta_i']'r_i^{-1}[x_t - \theta_i']\right\}, \qquad (5)$$

where n_x is the dimension of the vector x_t, the parameter $\Theta \equiv \Theta_i$ for $c_t = i$ and $\Theta_i = \{\theta_i, r_i\}$ are the corresponding expectations and covariance matrices.

The pointer model from the joint pdf (3) is the transition table

$f(c_t = i|\alpha) \equiv$ (6)

$c_t = 1$	$c_t = 2$	$\ldots$	$c_t = n_c$
α_1	α_2	$\ldots$	α_{n_c}

where $\alpha = \{\alpha_i\}_{i=1}^{n_c}$ and α_i are nonnegative probabilities of the value i of the pointer c_t.

The specific case considered in the paper assumes that the values of x_t are generated permanently in real time, while y_t can be measured only for a period of time $t \leq n_t$. Thus, for the introduced models, the task can be formulated as follows: (i) classify the data x_t using the estimation of the pointer c_t and (ii) predict the variable y_t for the time $t > n_t$.

B. Clustering and Prediction Algorithm

1) Clustering and Estimation: For the task of clustering the data x_t, it is necessary to estimate the pointer c_t along with the parameters Θ and α. The knowledge of the detected cluster labeled by the pointer value is then used in the model (4). The posterior pdf for the pointer estimation is derived according to the recursive Bayesian mixture estimation theory [20], [21] via the marginalization of the joint pdf of all unknown variables

$$f(c_t = i|x(t)) = \int_{\Theta^*}\int_{\alpha^*} f(\Theta, \alpha, c_t = i|x(t))d\alpha d\Theta, \qquad (7)$$

where Θ^* and α^* stand for the entire definite space of the used variables and $x(t)$ denotes the collection of all the data x_t measured up to time t including prior knowledge. Using the

Bayes and chain rules [22], the joint pdf inside the integrals in (7) is decomposed into the models and prior pdfs

$$\underbrace{f(x_t|c_t=i,\Theta)}_{(5)}\underbrace{f(\Theta|x(t-1))}_{GiW\ prior\ pdf}\underbrace{f(c_t=i|\alpha)}_{(6)}\underbrace{f(\alpha|x(t-1))}_{Dir\ prior\ pdf}, \quad (8)$$

where GiW denotes the conjugate Gauss-inverse-Wishart prior pdf for the normal distribution (5), see, e.g., [20], [22], while Dir stands for the Dirichlet prior pdf conjugate to the pointer model (6) according to [21]. The statistics of the GiW pdf are updated recursively by each data item measured at time t using the initial information matrices $(V_0)_i$ and the counters $(\kappa_t)_i$ for each i-th model along with the initialized number of clusters in the following way

$$(V_t)_i = (V_{t-1})_i + w_{i;t}\begin{bmatrix} x_t \\ 1 \end{bmatrix}[x_t' \ \ 1], \quad (9)$$

$$(\kappa_t)_i = (\kappa_{t-1})_i + w_{i;t}, \quad (10)$$

where $w_{i;t}$ is the weight of the i-th model (5) expressing the probability that the current data item x_t belongs to the i-th cluster, see [20]–[22].

Similarly, the Dirichlet statistics update has the form [21]

$$(\nu_t)_i = (\nu_{t-1})_i + w_{i;t} \quad (11)$$

with the initial statistics $(\nu_0)_i$ defined for each i. The updated statistics are then used for re-computing the point estimates of the parameters. More information is available in the mentioned sources.

The weights are obtained in the following way. The weighting vector $w_t = [w_{1;t}, \ldots, w_{n_c;t}]'$ is computed [20], [21], [23] as

$$w_t = m. * \hat{\alpha}_{t-1} \quad (12)$$

and normalized. Here, $.*$ means multiplying by entries, $\hat{\alpha}_{t-1}$ is the previous (or initial) point estimate of the parameter α, and m is the vector of proximities of the current data item x_t to the i-th model (5), which is calculated by substituting x_t and the previous (or initial) point estimates of the parameters θ_i and r_i into the pdf (5).

According to the assumptions (3), the Rayleigh model (4) is estimated for each of the clusters detected at time t, i.e., for $c_t = i$. The Rayleigh statistics update can be easily derived using the likelihood function of the distribution. Finally, it takes the following form using the initial statistics $(S_0)_i$

$$(S_t)_i = (S_{t-1})_i + w_{i;t}y_t^2, \quad (13)$$

which is then used for re-computing the point estimate of the parameter σ_i, see, e.g., [24]

$$(\hat{\sigma}_t)_i = \frac{(S_t)_i}{2(\kappa_t)_i}. \quad (14)$$

With the help of the above relations, the clustering algorithm can be summarized in the following form.

<u>The initialization:</u>
For the time $t = 1$

1) Set the number of clusters n_c.
2) $\forall i$
   a) Set the initial statistics $(V_0)_i$, $(\kappa_0)_i$, $(\nu_0)_i$ and $(S_0)_i$.
   b) Using the initial statistics, compute the point estimates $\hat{\theta}_{i;0}$ and $\hat{r}_{i;0}$ with the help of partitioning the information matrix [22] and $\hat{\alpha}_0$ by the normalization of the statistics ν_0 [21].

<u>The clustering and estimation:</u>
For the time $t = 2, 3, \ldots, n_t$

1) Measure data x_t, y_t.
2) $\forall i$
   a) Obtain the proximities m_i.
   b) Compute the weights $w_{i;t}$ according to (12).
   c) Update the statistics of the GiW and Dirichlet pdfs according to (9), (10) and (11).
   d) Re-compute the point estimates $\hat{\theta}_{i;t}, \hat{r}_{i;t}$ and $\hat{\alpha}_t$, see [20]–[22].
3) Get the point estimate of the pointer c_t according to the index of the maximum entry of the weighting vector w_t.
4) For the i-th cluster corresponding to the value of c_t
   a) Update the statistics of the Rayleigh pdf (4) according to (13).
   b) Re-compute the point estimate of the parameter σ_i using (14).

2) Prediction: Using the above clustering algorithm, all of the models are learned using the data measured at each time instant t for time $t \le n_t$. Then for time $t > n_t$, the estimated model (4) can be used for the prediction purpose using the permanently available data x_t for detecting the clusters. The prediction algorithm is given below.

<u>Prediction:</u>
For the time $t = n_t + 1, n_t + 2, \ldots$

1) Measure data x_t.
2) $\forall i$, obtain the proximities m_i using the last point estimates $\hat{\theta}_{i;n_t}$ and $\hat{r}_{i;n_t}$ and the current data item x_t.
3) Using the proximities and the last available point estimate $\hat{\alpha}_{n_t}$, compute the weighting vector w_t via (12).
4) Determine the point estimate of the pointer c_t according to the index of the maximum probability in the weighting vector, which labels the cluster i generating the data at time t.
5) For $c_t = i$
   a) Predict the expectation of the variable y_t using
   $$(\hat{\sigma}_{n_t})_i\sqrt{\frac{\pi}{2}}. \quad (15)$$
   b) Obtain the distribution by generating values of y_t from the i-th model (4) with the substituted point estimate $(\hat{\sigma}_{n_t})_i$.

The proposed approach was tested in Scilab (www.scilab.org), which is a free and open source software for engineering computations. Figures were made in MATLAB®. The experiments are demonstrated in the subsequent section.

III. Experiments

The target variable in the experiments, i.e., the variable y_t, was the number of boarding and/or disembarking passengers at a tram station. As an explanatory variable, i.e., the variable x_t, the time difference between two tram arrivals was used. A data set with 2,300 data items from a real tram network was provided by the public transport organizer. The number of boarding and disembarking passengers is a nonnegative discrete random variable. The Poisson distribution may be assumed as adequate for describing it [18]. Here, the described experiments aim to observe whether the Rayleigh distribution is suitable for modeling such data. Results are expected to be helpful in the passenger demand prediction task, which consists in predicting the number of boarding and disembarking passengers [25].

Models of the number of boarding and disembarking passengers are contextually identical [18], therefore only the number of disembarking passengers is modeled in this paper. At first, this variable is modeled for one station only. Secondly, a model of a short line consisting of two stations is presented.

A. Experiments for One Station

Fig. 1 compares the histograms of original and predicted numbers of disembarking passengers for a selected tram station. The original measurements are clearly multimodal and the same can be seen for their predictions. Three distributions with the mean near 1, 4, and 7 passengers can be guessed in the original data (top). The predicted values are also multimodal with the Rayleigh distributions with the mean value about 6, 9, and 14 passengers (bottom). The bottom plot in Fig. 1 shows the predictions around the value of 21 disembarking passengers unlike the original data.

To compare the proposed method with alternative ones in the same field, two other approaches were chosen. First, the Poisson distributions learned on clusters [25] were used, and secondly, the single Rayleigh distribution in its basic form was estimated using the data set. Results of these two methods are presented in Fig. 2 for the same station as shown in Fig. 1.

It can be clearly seen in Fig. 2 (top) that the use of the Poisson distributions learned on clusters does not provide adequate results. Almost no multimodality can be seen and the peak of the distribution near 11 passengers does not correspond to the original data set. In Fig. 2 (bottom), the single Rayleigh distribution provides the prediction closer to the real data than the Poisson distribution. However, the multimodality of the data is not caught correctly, which can be seen comparing Fig. 2 (bottom) and Fig. 1 (top).

The data prediction obtained with the help of the proposed method and the two theoretical counterparts is presented in Fig. 3. The results of all the approaches are visually close to each other. However, Table I comparing the mean absolute percentage error (MAPE) for all of them shows that the proposed prediction with the Rayleigh pdfs learned on clusters has the lowest prediction error among the compared approaches.

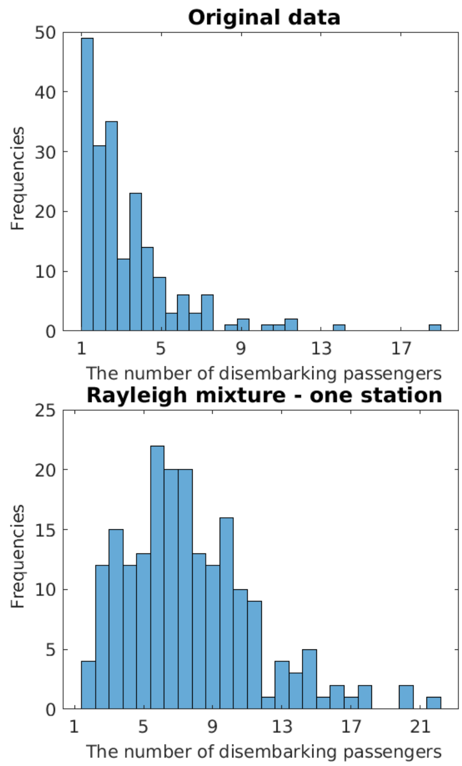

Fig. 1. The comparison of histograms of real data (top) and their predictions (bottom) at one station

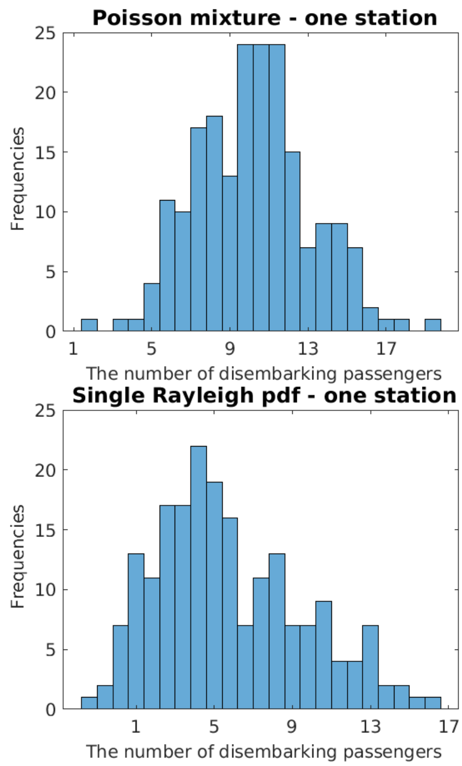

Fig. 2. The comparison of histograms for the Poisson distributions learned on clusters (top) and the single Rayleigh distribution (bottom)

TABLE I
THE MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) FOR ONE STATION

	MAPE
The Rayleigh mixture – one station	0.081
The Poisson mixture – one station	0.086
The single Rayleigh distribution – one station	0.084

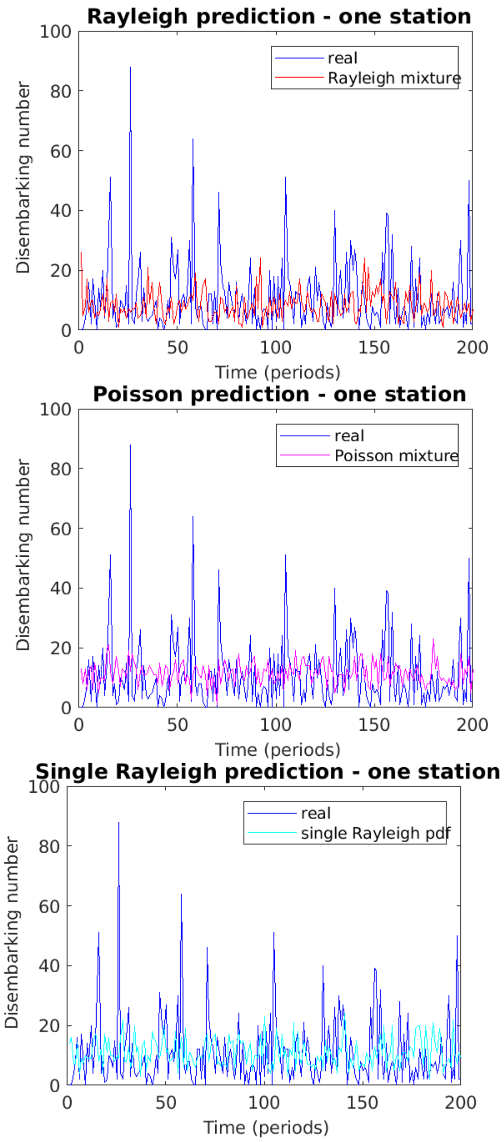

Fig. 3. The prediction with the Rayleigh mixture (top), the Poisson mixture (middle) and the single Rayleigh distribution (bottom) for a single station

B. Experiments for a Tram Line

At the second stage of experiments, a line consisting of two stations was modeled, which is an important step in predicting passenger numbers in larger networks [25]. In this case, the explanatory variable is two-dimensional. Its first entry is the time difference between the arrival time of two tram trips similarly to the previous case. The second explanatory variable is the predicted number of disembarking passengers at the previous station.

Fig. 4 compares the histograms of the predicted data using the mixtures of Rayleigh (top) and Poisson distributions (bottom) for a tram station being a part of a line. Similarly to Fig. 1, the prediction accuracy of the obtained histograms is acceptable and the data multimodality can be clearly observed. Three Rayleigh distributions can be observed near values 6, 9, and 16. Three Poisson distributions near values 5, 9, and 13 could be identified. It can be stated that both the methods provide the similar accuracy.

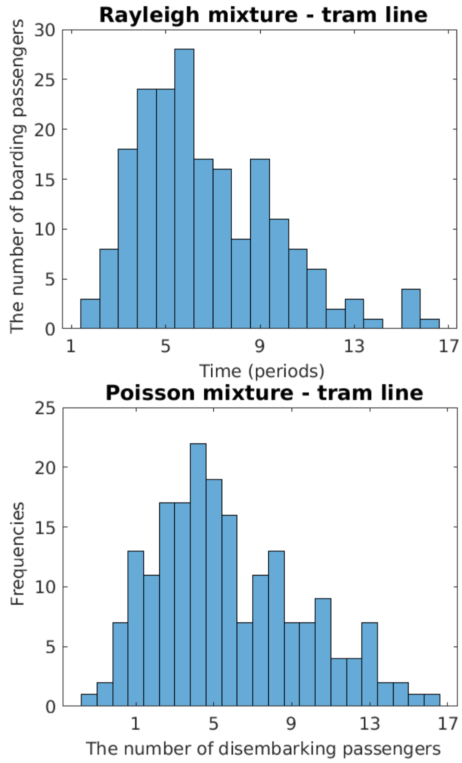

Fig. 4. Histograms of the prediction with the Rayleigh mixture (top) and the Poisson mixture (bottom) for a tram line

The data prediction with the compared methods is not shown here to save space. The MAPE values provided in Table II demonstrate the insignificant difference in the obtained prediction in favor of the Poisson distributions learned on clusters of the explanatory variables.

TABLE II
THE MEAN ABSOLUTE PERCENTAGE ERROR (MAPE) FOR A TRAM LINE

	MAPE
The Rayleigh mixture – tram line	0.091
The Poisson mixture – tram line	0.084

C. Discussion

The main purpose of this paper, i.e., fitting nonnegative discrete data using the Rayleigh distributions learned on clusters of the explanatory variables, was successfully achieved. The results of the conducted experiments show that despite the Rayleigh pdf is originally intended for modeling continuous data, it can be also used to describe specific discrete variables with a higher number of possible nonnegative realizations. Obviously the presented approach can be applied in the case of a similar form of the original pdf describing the data (here the Poisson one).

In this way, the main contributions of the presented approach are: (i) the use of the discretized knowledge of clusters of the explanatory variables used for learning models and (ii) the description of nonnegative discrete data by the Rayleigh distributions estimated on the obtained clusters. It should be also noted that due to the proposed algorithm, the prediction can be computed online using the measurements of the explanatory variables.

Although the results of the prediction with the Rayleigh distributions were generally adequate, there is still a room for an improvement. For example, low values of the number of disembarking passengers near 1 passenger were not almost captured in the prediction while they were significantly present in the original data set. Moreover, in spite of achieving the promising results for modeling a single tram station, the results for modeling a tram line using the proposed method were slightly worse than with the alternative counterparts.

The potential application of the proposed method is not limited by the passenger demand prediction task. It can be also used in other areas where nonnegative discrete data with a high number of possible realizations are common, such as medicine, economics, accident analysis, etc.

As regards the limitations of the approach, they include the assumption of the data multimodality along with available observations of the explanatory variables.

IV. Conclusion

The study proposed the use of the discretized information from clusters of the explanatory variables, which can be obtained using the Bayesian mixture estimation methodology, for the prediction of the target discrete variable. Based on the predictive model learned on these clusters, a mixture of Rayleigh distributions was fitted over the discrete data. The prediction of the target variable, here the number of disembarking passengers, was provided online. The promising results have been demonstrated during the validation of the approach.

The future work include the development of the algorithms for covering a tram line. They are expected to be based on the mixture of the Rayleigh and Poisson pdfs.

References

[1] E.T. Maziriri and E. Chinomona, “Modeling the Influence of Relationship Marketing, Green Marketing and Innovative Marketing on the Business Performance of Small, Medium and Micro Enterprises (SMMES) ,” Journal of Economics and Behavioral Studies, vol. 8, no. 3, 2016, pp. 127-139.

[2] D. Barbolosi, J. Ciccolini, B. Lacarelle, F. Barlési and N. André, “Computational oncology–mathematical modelling of drug regimens for precision medicine,” Nature reviews Clinical oncology, vol. 13, 2016, pp.1-14.

[3] Y. Han, A. Hegyi, Y. Yuan, S. Hoogendoorn, M. Papageorgiou and C. Roncoli, “Resolving freeway jam waves by discrete first-order model-based predictive control of variable speed limits,” Transportation Research Part C: Emerging Technologies, vol. 77, 2017, pp. 405-420.

[4] S. Haar and S. Theissing,“Forecasting passenger loads in transportation networks,” Electronic Notes in Theoretical Computer Science, vol. 327, 2016, pp.49-69.

[5] J. Roos, G. Gavin and S. Bonnevay, “A dynamic Bayesian network approach to forecast short-term urban rail passenger flows with incomplete data,” Transportation Research Procedia, vol. 26, 2017, pp. 53-61.

[6] P. Michalaki, M.A. Quddus, D. Pitfield and A. Huetson, “Exploring the factors affecting motorway accident severity in England using the generalised ordered logistic regression model,” vol. 55, 2015, pp. 89-97.

[7] A. Maydeu-Olivares and Y. Liu, “Item diagnostics in multivariate discrete data,” Psychological Methods, vol. 20, 2015, pp. 276-292.

[8] D. Tran, K. Vafa, K.K. Agrawal, L. Dinh and B. Poole, ‘Discrete flows: invertible generative models of discrete data,” in 33rd Conference on Neural Information Processing Systems (NeurIPS), 2019, pp.14692-14701.

[9] S. Harrod, G. Pournasas and B.F. Nielsen, “Distribution fitting for very large railway delay data sets with discrete values,” Danish Journal of Transportation Research, vol. 1, 2019, pp. 1-12.

[10] R. Anwar and M. Hamilton, “Direct ellipsoidal fitting of discrete multi-dimensional data,” SMU Journal of Undergraduate Research, vol. 5, 2020, 4.

[11] M.H. Alamatsaz, S. Dey, T. Dey and S.S. Harandi, “Discrete generalized Rayleigh distribution,” Pakistan journal of statistics, vol. 32, 2016, pp. 1-20.

[12] A.S.A. Ghani and N.A.M. Isa, “Underwater image quality enhancement through integrated color model with Rayleigh distribution,” Applied Soft Computing, vol. 27, 2015, pp. 219-230.

[13] P. Imrey, “Poisson regression, logistic regression, and loglinear models for random counts”, in Handbook of Applied Multivariate Statistics and Mathematical Modeling, 2000, pp.391-437.

[14] S.Yang and G. Berdine, “Poisson Regression,” The Southwest Respiratory and Critical Care Chronicles, vol. 3(9), 2015, pp. 61–64.

[15] P. Papastamoulis, M.-L. Martin-Magniette and C.Maugis-Rabusseau, “On the estimation of mixtures of Poisson regression models with large number of components”, Computational Statistics & Data Analysis, vol. 93, 2016, p. 97-106.

[16] I. Nagy, E. Suzdaleva and P. Pecherková, “Comparison of various definitions of proximity in mixture estimation, ” in Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Lisbon, Portugal, July, 29 – 31, 2016, pp. 527-534.

[17] R. Likhonina, E. Suzdaleva and I. Nagy, “Comparison of Mixture-Based Classification with the Data-Dependent Pointer Model for Various Types of Components”, Research report 2355, Prague, 2016.

[18] M. Petrouš, E. Suzdaleva and I. Nagy, “Modeling of passenger demand using mixture of Poisson components,” in Proceedings of the 16th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2019), Prague, 2019, p. 617-624.

[19] S. Haykin. Neural Networks: A Comprehensive Foundation. Macmillan, 1994, New York.

[20] M. Kárný, J. Kadlec and E.L. Sutanto, “Quasi-Bayes estimation applied to normal mixture,” in 3rd European IEEE Workshop on Computer-Intensive Methods in Control and Data Processing, Rojíček, J., Valečková, M., Kárný, M., Warwick, K. (Eds.), 1998, September, Prague, Czech Republic, pp. 77-82.

[21] M. Kárný, J. Böhm, T.V. Guy, L. Jirsa, I. Nagy, P. Nedoma and L. Tesař, Optimized Bayesian Dynamic Advising: Theory and Algorithms. Springer-Verlag, London, 2006.

[22] V. Peterka, Bayesian system identification, in Eykhoff, P. (Ed.), Trends and Progress in System Identification. Oxford, Pergamon Press, 1981, pp. 239-304.

[23] 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, 2017.

[24] M. M. Siddiqui, “Statistical inference for Rayleigh distributions”, The Journal of Research of the National Bureau of Standards, Sec. D: Radio Science, vol. 68D, no. 9, 1964, p. 1007.

[25] M. Petrouš and E. Suzdaleva, “Modeling of mixed data for Poisson prediction”, in the IEEE 14th International Symposium on Applied Computational Intelligence and Informatics, 2020, accepted.