Assessment of Poisson Naive Bayes Classifier with Fuzzy Parameters Using Data from Different Statistical Distributions

Elaine A. M. G. Soares and Ronei M. Moraes

Laboratório de Estatística Aplicada ao Processamento de Imagens e Geoprocessamento,
Universidade Federal da Paraíba,
58051-900, João Pessoa-PB, Brasil
elaineanita1@gmail.com,ronei@de.ufpb.br

Abstract. This paper means to assess the accuracy of the Poisson Naive Bayes classifier with Fuzzy Parameters (PNB-FP) for classification tasks, aiming to determine the effectiveness of this method when classifying specific data. This was achieved using data from five different statistical discrete distributions . After the assessment was completed, results were collect. These are shown and analyzed in the result sections according to the statistical distribution of data.

Keywords: Poisson Naive Bayes Classifier, Fuzzy Parameters, Accuracy Assessment

1 Introduction

Since the inception of the use of computers for classification tasks, there has been extensive research attempting to make the processing of data more efficient and accurate. Intending to make data processing more efficient, scientists and researchers have been studying improved classification methods.

There are many different classifiers built for other distributions, such as Beta [15], Binomial [2], and Gaussian [9]. The Poisson distribution expresses the chance of a determined event happen in a given time, space, volume, etc. This distribution is significant because it is applicable to real-life scenarios, such as the number of registered cases of a particular disease, as shown by Feller [7]. Poisson based methods can be found in the scientific literature, such as the Poisson Naive Bayes [1], Poisson mixture models [11], Fuzzy Poisson Naive Bayes [16]. This last one had its uncertainty modeled by the probability proposed by Zadeh [18], which weights the classic probability by the membership function of a fuzzy event.

Within this work, a new classifier is proposed that also follows a Poisson Naive Bayes approach. Additionally, the method used in this work uses a fuzzy parameterization, in opposition to [16] and a standard Naive Bayes approach. For reference purposes, this method will be cited as PNB-FP. Furthermore, in this paper, the proposed new classifier will be assessed.

For this study, data from five different statistical distributions were generated, these include Binomial, Negative Binomial, Poisson, Gaussian, and Discrete Uniform. Each distribution represent one or more possible applications for this method. From each distribution, sample sets were assembled of training and testing data. The training sets were used to estimate the fuzzy parameters, while the testing sets were used in the classification part of the method. The information gathered from the results include execution times, Kappa coefficients and variances, in addition to classification matrices.

2 Definitions

2.1 Fuzzy Sets and Fuzzy Numbers

Let there be a space of objects X with a generic element x. Given that, $X = \{x\}$. It can be defined a fuzzy set A in X characterized by a membership function $\mu_A(x)$ which correlates each point x in X to a real number in the interval $[0, 1]$. The value of $\mu_A(x)$ represents the degree of membership of x in A[17]. For example, if $\mu_A(x_0) = 0$, it is said that x_0 does not belong to A; if $\mu_A(x_1) = 1$, it is said that x_1 belongs to A; and if $\mu_A(x_2) = 0.7$, it is said that the membership degree of x_2 in A is 0.7.

Furthermore, a fuzzy set A with membership function $\mu_A(x)$ can be expressed by the set of its $\alpha-$cuts. Then, it is denoted by A_α and the following is true:

$$A_\alpha = \{x \in X | \mu_A(x) \leq \alpha\} \tag{1}$$

The membership function $\mu_A(x)$ can also be represented in terms of its $\alpha-$cuts[5]:

$$\mu_\alpha = sup_{\alpha \in [0,1]} min\{\alpha, \mu_{A_\alpha}(x)\} \tag{2}$$

A fuzzy number is defined by a convex normalized fuzzy set, which is commonly represented by triangular or trapezoidal shapes. In this work, a triangular shape was used to represent fuzzy numbers. Given that, a triangular fuzzy number $\bar{N}$ is represented by three real numbers a, b, and c; where $[a, c]$ is the base of a triangle and b is its vertex, which is where the membership function reaches its maximum value. Therefore, a fuzzy number is defined by [3] as:

$$\bar{N} = [a/b/c], \quad a \leq b \leq c \tag{3}$$

Now, for any fuzzy number $\bar{N}$, it is known that $\bar{N_\alpha}$ is a closed interval for $0 \leq \alpha \leq 1$, which can be described as:

$$\bar{N_\alpha} = [n_1(\alpha), n_2(\alpha)], \quad n_1(\alpha) \leq n_2(\alpha) \tag{4}$$

with the following properties:

- $n_1(\alpha)$ will be a continuous, monotonically increasing function of α in $[0, 1]$;
- $n_2(\alpha)$ will be a continuous, monotonically decreasing function of α in $[0, 1]$;
- $n_1(1) = n_2(1)$.

2.2 Fuzzy Probability

According to Buckley [3], for an event A in a sample space ϕ, a fuzzy probability can be defined in which the parameters are fuzzy numbers. Thus, distributions of probability can be generalized and their parameters estimated by fuzzy numbers, which, as stated previously, are represented as triangular or trapezoidal numbers. This approach is different from the one proposed by Zadeh [18] because it models the uncertainty on its own parameter of the distribution.

Let X be in R^n a crisp random variable with density of probability D, which function is given by $f_x(B, \theta)$, where θ is the parameters vector for this density of probability function with $\theta = \theta_1, \theta 2, \ldots$. The probability of an event B in ϕ is given by [13] as

$$\bar{P}(x)[\alpha] = \left\{ \int_{R^n} f_X(B, \theta)dx \quad | \quad \theta \in \theta_f[\alpha] \quad | \quad \mathbf{R} \right\} \tag{5}$$

where $0 < \alpha \leq 1$, θ_f are fuzzy numbers, $\theta_f[\alpha]$ are their respective $\alpha-$cuts, and $\mathbf{R}$ is a restriction in order to preserve the following probability property [3]

$$\mathbf{R} = \int_{R^n} f_X(B, \theta)dx = 1, \tag{6}$$

and it is required for all event in the sample space ϕ.

Since the vector of parameters θ is unknown, Confidence Intervals (CI) were used to estimate it, varying the confidence levels [3].

2.3 Fuzzy Parameters Estimation

The fuzzy Poisson distribution is a specialization of equation (5), where the parameter θ is given by λ. Thus, its probability function is expressed by:

$$\bar{P}(x)[\alpha] = \left\{ \frac{\lambda^x e^{-\lambda}}{x!} \quad | \quad \lambda \in \bar{\lambda}[\alpha] \quad | \quad \mathbf{R} \right\} \tag{7}$$

From this, it can be obtained the fuzzy Poisson probability mass function f_x, represented by:

$$f_x(x, \bar{\lambda}) = \frac{\bar{\lambda}^x e^{-\bar{\lambda}}}{x!} \tag{8}$$

where $\bar{\lambda} \in \bar{\lambda}[\alpha]$ and $x \geq 0$.

Additionally, the $\bar{\lambda}$ parameter is a fuzzy number calculated by [3] using the confidence level $(1-\beta)\%$, for $0.001 \leq \beta \leq 1$, which represents each α-cut of $\bar{\lambda}$. The confidence level $(1-\beta)\%$ is a probability calculated from a inverse normal distribution. Thus, $\bar{\lambda}[\alpha]$ is given by:

$$\bar{\lambda}[(1-\beta)\%] = \left[\frac{\sqrt{n} * Mean(x)}{\sqrt{n} + z_{\beta/2}}, \frac{\sqrt{n} * Mean(x)}{\sqrt{n} - z_{\beta/2}} \right] \tag{9}$$

where $\beta \in [0.001; 1]$ and Mean(x) is an estimator given by [11] as:

$$Mean(x) = \frac{c1 + \sum_{k=1}^{dim(D)} (X_k, w_i)}{c2 + dim(D)} \tag{10}$$

where c_1 and c_2 are smoothing constants, $dim(D)$ is the length of sample data D for which the class is w_i, and $\sum_{k=1}^{dim(D)} (X_k, w_i)$ is the counting of events in D, in which the value of X_k is associated to the class w_i. It should be noted that there are many ways of estimating Mean(x). However, the estimator proposed by Ogura [11] helps with preventing estimations with a zero value for $\bar{\lambda}_{ki}$ through the use of constants c_1 and c_2.

2.4 Poisson Naive Bayes with Fuzzy Parameters

Let there be $\Omega = 1, ..., M$ classes of performance, where Ω is the space of decision and M is the total number of classes on it. Let there be vector of data $X = \{X_1, X_2, ..., X_n\}$ in the sample data D, where n is the total of distinct features in X and w_i, $i \in \Omega$, is the class for the vector X. Furthermore, it is assumed that each one of those features X_i are independent between each other. Thus, the Naive Bayes (NB) hypothesis is applied to estimate the probability of X belonging to w_i, which is expressed by:

$$P(w_i|X) = \frac{1}{S} P(w_i) \prod_{k=1}^{n} P(X_k|w_i) \tag{11}$$

where S is a scale factor, which depends on X.

The classification rule for NB is given by:

$$X \in w_i \quad if \quad P(w_i|X) > P(w_j|X) \tag{12}$$

for all $i \neq j$.

Assuming that each X_i belongs to the Poisson distribution, the following is true:

$$P(X_k = v|w_i) = \frac{\bar{\lambda}_{ki}^{v} e^{-\bar{\lambda}_{ki}}}{v!} \tag{13}$$

where v is described as a positive integer, v! is the factorial of v, and λ is the fuzzy parameter, as given by equation (9).

Given the high level of computational complexity of equation (13), Moraes [16] proposed the use of the natural logarithm function on equation (11) in order to simplify it and reduce processing complexity. Hence, equation (11) can be rewritten as:

$$log[P(w_i|X)] = log\left(\frac{1}{S}\right) + log[P(w_i)] + \sum_{k=1}^{n} log[P(X_k|w_i)] \tag{14}$$

The $log[P(X_k|w_i)]$ in equation (14) is given by:

$$log[P(X_k|w_i)] = v * log(\bar{\lambda}_{ki}) - \bar{\lambda}_{ki} - log(v!) \quad (15)$$

Thus, the classification rule for the PNB classifier is:

$$X \in w_i \quad if \quad f(w_i, X) > f(w_j, X) \quad (16)$$

where $f(w_i, X)$ is $log[P(w_i|X)]$.

3 Assessment Methodology

Several classification methods found in the literature can obtain better performance when used with a specific statistical distribution of data[12]. PNB-FP is a recently proposed method consequently, it is not clear its performance with respect to different statistical distributions of data. In order to explore the features of this model, in this paper it is studied the behavior of PNB-FP for classification tasks, using Monte Carlo simulated data [8] from five different statistical distributions: Binomial, Negative Binomial, Poisson, Discrete Uniform and Gaussian. Two sets of data samples were generated for each of the five distributions, described as:

- the first set contained 1000 samples per class for the PNB-FP method training. In fact, it were generated 5000 samples, but the first 4000 were discarded in order to prevent unwanted oscillations in the probability distributions; and
- the second with 4000 samples by class for all classification tests. For this case, it were generated 5000 samples, but the first 1000 were discarded.

Each sample had four different classes, with up to four dimensions. The PNB-FP classifier was executed varying from one through four dimensions and the respective classification matrices were stored for future reference.

Using simulated data, it is possible to evaluate many situations, which would not be conceivable with a study case using real data. Furthermore, benchmark datasets would also present limitations for this analysis.

Figure 1 displays the parameters for the generation of sample data from the Poisson distribution used on training the PNB-FP classifier, in blue, and the sample data used for testing, in red. Furthermore, this figure has 8 lines per 4 columns, each line represents a dimension and each column represents a class. An intentional intersection was made between some data classes in order to make the simulations as similar as possible to real conditions. All distributions had their data generated using the same methodology, but different parameters.

3.1 Kappa Coefficient

The Kappa Coefficient K is widely used in the literature of pattern classification [6]. This coefficient was proposed by Cohen [4] and it is a powerfully weighted measure which takes into account agreements and disagreements between two sources of information. From a classification matrix:

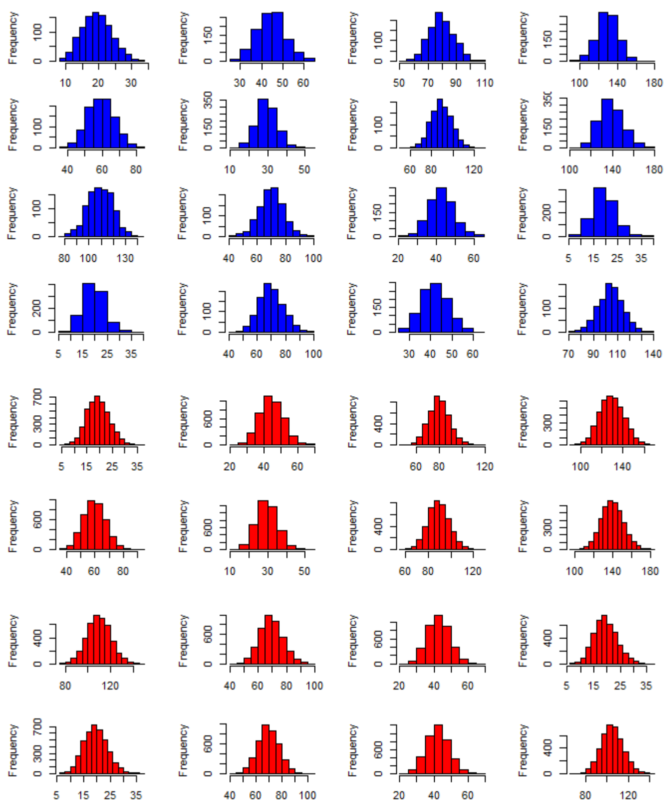

Fig. 1. Random numbers generated for the Poisson distribution (training and test).

$$K = \frac{P_0 - P_c}{1 - P_c} \tag{17}$$

with P_0 and P_c as:

$$P_0 = \frac{\sum_{i=1}^{M} n_{ii}}{N} \quad and \quad P_c = \frac{\sum_{i=1}^{M} n_{i+} n_{+i}}{N^2} \tag{18}$$

where n_{ii} is the total of the main diagonal, n_{i+} is the total of line i, n_{+i} is the total of column i, M is the total number of classes, and N is the total of possible decisions in the classification matrix.

The variance of the Kappa Coefficient K, denoted by σ_K^2 is described by [14] as:

$$\sigma_K^2 = \frac{P_0(1-P_0)}{N(1-P_c)^2} + \frac{2(1-P_0) + 2P_0P_c - \theta_1}{N(1-P_c^3)} + \frac{(1-P_0)^2\theta_2 - 4P_c^2}{N(1-P_c)^4} \tag{19}$$

where θ_1 and θ_2 are given by:

$$\theta_1 = \frac{\sum_{i=1}^{M} n_{ii}(n_{i+} + n_{+i})}{N^2} \quad and \quad \theta_2 = \frac{\sum_{i=1}^{M} n_{ii}(n_{i+} + n_{+i})^2}{N^3} \tag{20}$$

All Kappa coefficients and respective variances were computed, which are presented in the Results section of this paper. Additionally, according to Landis and Koch nomenclature [10], the Kappa coefficient can be interpreted as presented in the Table 1.

Table 1. Interpretation of Kappa Coefficient [10]

Kappa Coefficient	Agreement Degree
< 0.0	Poor
$0.00 \mid\!- 0.20$	Slight
$0.20 \mid\!- 0.40$	Fair
$0.40 \mid\!- 0.60$	Moderate
$0.60 \mid\!- 0.80$	Substantial
$0.80 \mid - \mid 1.00$	Almost Perfect

4 Results for Poisson Distribution

The Poisson Distribution is used to model the occurrence of discrete particular events in continuous space or time or both. Since this is the distribution used in the body of the proposed method, almost perfect results were expected. With the use of only one dimension, the method returned Kappa coefficient equals to 97.89% with variance 2.30736×10^{-7}. For two and three dimensions, the Kappa coefficient obtained from the classification matrices were 99.92% and 99.98%, respectively, with variances 9.01×10^{-9} and 2.47×10^{-9}. Furthermore, when classifying data with four dimensions, this method returned a perfect Kappa coefficient of 100% without variance.

The results presented in this section helped confirm the hypothesis that this data set would have almost perfect Kappa coefficients. For instance, the worst performing sample set of data still classified as almost perfect according to Table 1. Thus, as expected, the Poisson method classified samples of this distribution almost perfectly.

Table 2. Kappa Results for Poisson distribution

Dimensions	Kappa	Variance
1	97.89%	2.30736×10^{-7}
2	99.9189%	9.00686×10^{-9}
3	99.9778%	2.46872×10^{-9}
4	100%	0

5 Results

5.1 Binomial Distribution

The Binomial distribution models a discrete random variable, which is the counting of the occurrence of an determined event. In a simulation using only one dimension of data, the Kappa coefficient was 96.00% with variance 4.31x10^{-7}. When dimensional data was two, the Kappa coefficient resulted in 99.90% with variance 1.06x10^{-8}. Having both three and four dimensions, the Kappa coefficient was 100% with variances 0 and 1.23x10^{-10}, misclassifying only one wrong on the second case.

From those Kappa coefficients, it is correct to conclude that PNB-FP reached an almost perfect for any amount of dimensions. Thus, it is possible to classify data from the Binomial distribution with the proposed method.

Table 3. Kappa Results for Binomial distribution

Dimensions	Kappa	Variance
1	96.0044%	4.30539×10^{-7}
2	99.9044%	1.06097×10^{-8}
3	100%	0
4	99.9989%	$1.23456e \times 10^{-10}$

5.2 Negative Binomial Distribution

The Negative Binomial is a discrete distribution, in which are considered some conditions: the experiment must consists on an undetermined amount of repeated attempts, the probability of success is the same in each event and the events are independent. In the simulation for this distribution with four dimensions, it carried a perfect result with a nil Kappa variance. The results for three and four dimensions are comparable however, it missed an element and carried a variance of $1.23\text{x}10^{-10}$. Using data from one and two dimensions resulted in 92.90% and 99.76%, respectively, with a variances of $7.46\text{x}10^{-7}$ for one and $7.46\text{x}10^{-7}$ for two.

According to Table 1, all classification tasks that used data from the Negative Binomial Distribution presented an almost perfect Kappa coefficient. From this, it is noticeable that data from this distribution is well classified when using the introduced method.

Table 4. Kappa Results for Negative Binomial distribution

Dimensions	Kappa	Variance
1	92.8967%	7.46497×10^{-7}
2	99.76%	2.66186×10^{-8}
3	99.9989%	1.23456×10^{-10}
4	100%	0

5.3 Discrete Uniform Distribution

The Discrete Uniform distribution is used when all possible results have the same probability of occurrence in discrete space. The simulation for the Discrete Uniform distribution produced results that were almost perfect in relation to Kappa coefficient. For two dimensions, the Kappa coefficient was 97.31% with a variance of $2.93\text{x}10^{-7}$. Similarly, for three dimensions produced a Kappa coefficient of 99.02% accompanying a variance of $1.08\text{x}10^{-7}$. Likewise, four dimensions resulted in a Kappa coefficient of 98.05%, which included a variance of $2.13\text{x}10^{-7}$. Whereas, for only one dimension presented a substantial Kappa coefficient, which was 73.18% in addition to a variance of $2.38\text{x}10^{-6}$.

It is worth noticing that, with increment of dimensions, the classifier performed better. However, PNB-FP is only able to classify data from the Discrete Uniform distribution with almost perfect agreement degree, if dimension of data is greater than two. The results for dimension 1 were substantial and even though dimension 2 to 4 performed outstanding, it is clear that this method did not performed as well for this distribution's data as for the rest of the results presented in this paper.

Table 5. Kappa Results for Discrete Uniform distribution

Dimensions	Kappa	Variance
1	73.1811%	2.38383×10^{-6}
2	97.31%	2.92856×10^{-7}
3	99.0178%	1.08331×10^{-7}
4	98.0522%	2.13205×10^{-7}

5.4 Gaussian Distribution

The Gaussian or Normal Distribution is the most typical statistical distribution. It is widely used in several classification problems, even when data do not follow this distribution. It is due to the fact that several robust classifiers can produce almost perfect result in this situation, which is not the ideal situation. The Gaussian distribution simulation produced results that overall can be classified as almost perfect akin to the Kappa Coefficient. For only one dimension, it produced a Kappa coefficient of 96.65% with a variance of $3.63\text{x}10^{-7}$. Correspondingly, two dimensions formed a Kappa coefficient of 99.31% followed by a variance of $7.65\text{x}10^{-8}$. Furthermore, three dimensions delivered a Kappa coefficient of 99.82% which occurred with a variance of $2.67\text{x}10^{-8}$. Moreover, four dimensions coincided with a Kappa coefficient of 98.71% that accompanied a variance of $1.42\text{x}10^{-7}$.

From those results, it is worth noticing that PNB-FB is able to classify data from Gaussian distribution with good accuracy, since all Kappa coefficients were greater than 96.0%.

Table 6. Kappa Results for Gaussian distribution

Dimensions	Kappa	Variance
1	96.6467%	3.63211×10^{-7}
2	99.3078%	7.65126×10^{-8}
3	99.8192%	2.67417×10^{-8}
4	98.7067%	1.42282×10^{-7}

6 Result Analysis

Overall, this method presents good results for all distributions analyzed in this paper. Table 2 displays a summary of the best results obtained by each statistical distribution in these simulations. Most of the distributions obtained an almost perfect agreement degree when two or four dimensions were used. For the Binomial, Negative Binomial, and Poisson distributions, using the PNB-FB classifier, it was possible to achieve more than 99% of agreement classifications for two or more dimensions, according to the Kappa Coefficient. Similarly, for the Discrete Uniform distribution, it performed almost perfectly with a general

Kappa of 97%. While for the Gaussian sample data, a rate of 96% for one or more dimensions.

Table 7. Summary of the best result, by statistical distribution according to the Kappa coefficient.

Distribution	Number of Dimensions	Kappa Coefficient
Binomial	2 or more	$\geq 99\%$
Negative Binomial	2 or more	$\geq 99\%$
Poisson	2 or more	$\geq 99\%$
Discrete Uniform	2 or more	$\geq 97\%$
Gaussian	1 or more	$\geq 96\%$

7 Conclusion

Within this work, it was proposed a new classifier named Poisson Naive Bayes with Fuzzy Parameters (PNB-FP), which is based on fuzzy probability proposed by Buckley [3]. Furthermore, an assessment was presented for the PNB-FP accuracy for classification tasks using data with different statistical distributions. Simulations were made with five different statistical distributions: Binomial, Negative Binomial, Poisson, Discrete Uniform and Gaussian. For each statistical distribution four different dimensions were analyzed according to the Kappa coefficient and its variance. Accordingly to the results obtained, PNB-FP could be used to classify data from all distributions studied in this paper.

As future works, it is intended to study data with more than four dimensions additionally with data from different distributions than the ones presented in this paper. Furthermore, it is possible to explore the behavior of the PNB-FP when using a combination of samples derived from different distributions, and to compare with previous works.

References

1. Altheneyan, A.S., Menai, M.E.B.: Naive Bayes classifiers for authorship attribution of arabic texts. Journal of King Saud University Computer and Information Sciences, 26, 473–484 (2014)
2. Bielza, C.; Larranaga, P.: Discrete Bayesian network classifiers: A survey. ACM Computing Surveys, 47, Article 5 (2014)
3. Buckley, J. J.: Simulating Fuzzy Systems. Springer (2005)
4. Cohen, J.: A coefficient of agreement for nominal scales. Education and Psychology Measurement, 20, 37–46 (1960)
5. Dubois D.; Prade H.: Possibility Theory. Plenum Press (1988)
6. Duda, R. O.; Hart, P. E.; Stork, D. G.: Pattern Classification, 2nd edition, Wiley Interscience (2000)
7. Feller, W.: An Introduction to Probability Theory and its Applications. Wiley, 2nd edition (1971)

8. Gentle, J.E.: Elements of Computational Statistics. Springer (2005)
9. Johnson, R. A.; Wichern, D. W.: Applied Multivariate Statistical Analysis. Pearson, 6th edition (2007)
10. Landis, J. R.; Koch, G.: The measurement of observer agreement for categorical data. Biometrics, 33, 159–174 (1977)
11. Ogura, H.; Amano, H.; and Kondo, M.: Classifying documents with Poisson mixtures. Transactions on Machine Learning and Artificial Intelligence 2, 48–76 (2014)
12. Moraes, R. M.: Performance Analysis of Evolving Fuzzy Neural Networks for Pattern Recognition. Math. Soft Comput. Mag. 20, 63–69 (2013)
13. Moraes, R. M.: A new generalization for naive Bayes style fuzzy probabilistic classier. XV Safety, Health and Environment World Congress (2015)
14. Moraes, R. M.; Machado, L. S.: Psychomotor Skills Assessment in Medical Training Based on Virtual Reality Using a Weighted Possibilistic Approach. Knowledge-Based Systems, v. 70, p. 97-102 (2014)
15. Moraes, R. M., Rocha, A. V., and Machado, L. S.: Intelligent assessment based on beta regression for realistic training on medical simulators. Knowledge Based Systems, 32, 3–8 (2012)
16. Moraes, R. M.; Machado, L. S.: A Fuzzy Poisson Naive Bayes Classifier for Epidemiological Purposes. 7th International Conference on Fuzzy Computation Theory and Applications, 2, 193–198 (2015)
17. Zadeh, L. A.: Fuzzy Sets. Information and Control 8, 338–353 (1965)
18. Zadeh, L. A.: Probability measures of fuzzy events. J. Math. Anal. Applic., 10, 421–427 (1968)