Journal of Data Science and Modeling, Vol. 1, No. 2, 191-203, June 2023

Research Manuscript

Semiparametric Models in Finite Mixture of Negative Binomial Responses

Sima Naghizadeh Ardebili*[1]

[1] Assistant Professor, NationalOrganization for Educational Testing (NOET), Tehran, Iran.

Received: 04/02/2022 Accepted: 04/03/2022

Abstract:

So many natural phenomena of determining relationship and the effect of input variables on response variable in statistical studies may be different from the suggested model that the researcher selects for his study due to the occupant exists in the structure of data. It may be so influential on different distributions considered for response variables. The optimal properties of estimators were evaluated and studied for two statistical variables considered for response variable and input variables in the suggested model. It has been simulated for study, and real data has also been investigated. The results confirmed the superiority of a model close to the structure of the data.

Keywords: Estimator, Optimal properties, Semiparametric models, Negative binomial.

Mathematics Subject Classification (2010): 62J12, 62F12 .

*Corresponding Author: naghizadeh@sanjesh.org

1. Introduction

Determining the relationships between input variables and response variables is one of the most important concepts of statistics which must be considered by researchers in every statistical study. Investigating the effect of input variables on response variable and determining the selection of effective variables is also another part of a study, and, as a result, a model which must be considered. Nowadays, due to the wide volume of data in different fields of sciences such as data related to engineering and medicine and things like, it is needed for models that represent suitable interpretation based on them for data. The closer the represented model is to the data structure and is farther from selected artificial constraints, the more comprehensive the results will be. It may be possible to accelerate calculations by selecting a distribution for response variables, definitely, the achieved result will be less accurate. Maybe the structure of data needs a combination of distributions and or a mixture of distributions. It is possible to analyze developing complex data using these distributions, particularly when data is taken from a heterogeneous society. Therefore, there is sometimes inhomogeneous and inconsistent data, then a finite mixture of statistical models is used as a flexible instrument to investigate such data. Poisson distribution is usually one of the most famous distributions, which is considered as a response variable to analyze data. Equal mean and variance in theory and unequal in practice may be considered as one of the weak points of this distribution. When the number of quantitative variables is high, and a finite mixture of distributions is considered for the response variable, using a mixture of Poisson distributions may increase problems for data analysis due to unequal mean and variance in practice.

The negative binomial distribution is one of the substituent distributions instead of Poisson distribution in a finite mixture of distributions when there is over-dispersion. This may be a suitable research subject when the data structure is considered for response variable in a finite mixture of distributions which has been considered theoretically and practically in this study. A finite mixture of distributions has been considered in previous studies when the distribution of response variables is Poisson, but the current study aims at investigating the changes in theoretical discussions and data analysis when distribution change is done.

It is worth mentioning that using valid methods like penalty function in selecting variables enhances accuracy, in conclusion, bias reduction, and the efficiency of estimators. In order to reduce the possible biases of modeling, usually, so many covariates are removed from the initial stages of modeling. On the other hand, in order to increase prediction capabilities and selecting significant variables, the statisticians mostly use step-by-step deletion and selecting the best subsystem of covariates. This will be discussed and investigated in this study based on response variable change in two states.

What has created the current study is finite mixture integration from statistical models as well as equating concept. According to the approved regulation of the Ministry of Science, Research and Technology, students admitted for Ph.D. grade based on a national-wide exam held by National Organization for Educational Testing, students introduced several times more than the admission capacity to the universities and higher education institutions to send the volunteers' score to National Organization for Educational Testing based on scientific interview for final admission of volunteers. Also, it has been regulated that this exam has been constantly held every other six months, and those volunteers with sufficient scores are introduced to universities for an interview and things like that. Since individuals who are referred to universities for an interview during a specific period are those who have taken part in different exams of the National Organization for Educational Testing, thus scores distribution will be different for them. As a result, it is possible to say that there is a finite mixture of several societies. Equalization or equating happens when there is a combination of distributions. The considerable point here is that the number of questions (samples) that one has to correctly answer to achieve the maximum score to be introduced for the interview is a random variable, which may obey negative binomial distribution due to the bi-state response (true-false). Here, in fact there is a finite mixture of statistical models with negative binomial distribution, which is necessary to investigate the theoretical basis and the method of parameters estimation and tests related to effectiveness and ineffectiveness of contributing factors on the response. The penalty function has also been used to increase accuracy as a new idea. Therefore, a literature of the works conducted on the finite mixture of statistical models represented, then works on equating are discussed.

Although using these methods is useful in practice, they ignore intrinsic random errors at the stage of variables selection. Therefore, their theoretical properties have specific complexity, which must be considered by the researcher. Furthermore, selecting the best subset of variables has other specific features that are the most important of all. Being time-consuming and considering computational discussions are other problems of these methods which must be considered. However, the considerable point in selecting variables is correctly detecting the suggested model to achieve real response, which is considered important in this study. Doing studies and achieving suitable response based on selecting effective covariates is a concept that, if it has been done correctly, it will provide researchers with a better comprehension of changes existing in their surroundings. It must be considered more important when variables structure and the relationship between them have special complexity like semiparametric structure.

2. A Finite Mixture of Statistical Distributions

It is possible to introduce a statistical strategy based on a finite mixture of models in many cases for natural phenomena to criticize and investigate a wide range of statistical data. Such distributions are used when the structure of data is complex, and selecting a specific distribution usually faces with problems. Flexibility for modeling in the finite mixture models is more practical than determining a unit distribution for data. These distributions are successfully practical in different fields like biological, economic, and social sciences and also in different fields including genetics, medicine, psychology, engineering, and marketing.

Definition 2.1. *Consider $Y_1, \cdots, Y_n$ a random sample as much as n, so that Y_j is a p dimensional random vector with probability density function $f(y_j;\theta_i)$ for $(i = 1, \cdots, g)$ on R^p space. if $f_i(y_j;\theta_i)$ is the density function of the i^{th} variable in the i^{th} society, then finite mixture of density function of random variable Y is written as follows:*

$$f(y_j;\Psi) = \sum_{i=1}^{g} \pi_i f(Y_j;\theta_i), \qquad j = 1, 2, \cdots, p \tag{2.1}$$

Ψ is a vector including all unknown parameters in mixture model and is defined as$\psi = (\pi_1, \cdots, \pi_{g-1}, \xi^T)^T$ And $\pi_1, \cdots, \pi_g$ are non-negative quantities which are conisdered as weights and have values between 0 and 1so that:

$$\sum_{i=1}^{g} \pi_i = 1 \tag{2.2}$$

ξ is a vector including all unknown parameters $\theta_1, \cdots, \theta_g$.

Statistical distributions have been used for finite mixture, Ormoz and Eskandari (2016) hypothesized that response variable obeys generalized semiparametric regression model in relation to the covariate. According to this study, response variable Y with possible values of $Y \subset \mathrm{R}$ and a vector of covariates is as (u, x, z) with $x = (x_1, x_2, \cdots, x_q)^T$ as real variables and nonparametric coefficients, $z = (z_1, z_2, \cdots, z_p)^T$ is parametric coefficient of the model and u is a single variable. Therefore, a finite mixture of semiparametric regression model is defined as follows:

Definition 2.2. *consider* $G = \{f(Y;\theta,\Phi);(\theta,\Phi) \in \Theta \times (0,\infty)\}$ *a family of parametric density functions* Y*, where* $\Theta \subset \mathrm{R}$ *and* Φ *are scattering parameters. It has been said that* (u,x,Y) *is finite mixture of regression models with* K *order, when conditional density function* Y *is represented if* (u,x,z)*:*

$$f(y;u,x,z,\Psi) = \sum_{k=1}^{K} \pi_k f\left(y;\theta(u,x,z).\Phi_k\right) \tag{2.3}$$

With the following conditions:

a- $\theta_k(u,x,z) = h(x^T\alpha_k(u) + z^T\beta_k)$

b- $\alpha(\cdot)$ *is a vector including unknown functions of smooth regression coefficients.*

c- parametric vector Ψ *as:*

$\Psi = (\alpha_1,\cdots,\alpha_K,\beta_1,\cdots,\beta_K,\Phi,\pi$ *including* $\alpha_k = (\alpha_{k1},\cdots,\alpha_{kq})^T$, $\beta_k = (\beta_{k1},\cdots,\beta_{kq})^T$, $\Phi = (\Phi_1,\cdots,\Phi_K)^T$ *and* $\pi = (\pi_1,\cdots,\pi_{K-1})^T$.

A finite mixture of semiparametric models introduced by Ormoz and Eskandari (2016) as a member of the exponential family, definition 2.2 provides a common method to model such invisible heterogeneous relations.

However, there is a big problem in the structure of such a family that when linearity happens, the resulted estimates face instability; that is, the estimations considerably change based on different samples. To solve such a problem, it is possible to increase bias a little and consequently decrease variance by zeroing or contracting some of the coefficients of estimators. As a result, it may improve the accuracy of the total prediction. To solve such a problem, Santarelli *et al.* (2016) have used Conway–Maxwell–Poisson distribution (CMP) as a mixture and analyzed Gamma ray's data based on Poisson distribution, and have directly estimated the parameters of the suggested model. The important point is that variable selection and the effect of response change have not been considered in this study. Sometimes there is data that is not homogenous and consistent. Then, finite mixture models are used as a flexible instrument to model such data. Ormoz and Eskandari (2016) introduced variable selection using penalized methods in a combination of the generalized semiparametric model by Li and Liang (2008) and a finite mixture of regression models by Khalili and Chen (2007) that penalized methods have not been investigated about the combination of these two models at the same time. Cho and Fryzlewicz (2012) have also studied variable selection. In fact, they have generally investigated variable selection in exponential distributions family in the introduced model. Then they introduced a finite mixture of generalized semiparametric models using penalized estimation. In fact, they expanded the model in semiparametric state and considered nonparametric function multidimensional.

3. The Finite Mixture of Generalized Models (FMGM) with Negative Binomial Distribution Response

Finite mixture of Poisson (FMP) distributions is a famous method to analyze enumerated types. However, since it is covariant (equal mean and variance), it is limited in use in this distribution due to over-dispersion. As a result, due to the over-scattered structure, using such substituent methods has been highly suggested. As a substituent for finite mixture, finite binomial distributions are suggested. Negative binomial regression (NBR) is a suitable selection to model the relationship between explanatory variables and a dependent enumerated variable. Then, Poisson regression generalized state is considered, because the mean of negative binomial distribution has a similar structure with Poisson regression. Also it has an extra parameter to model over-dispersion. According to the definition of negative binomial distribution:

$$f(y_j;\alpha,\beta)=\binom{y_j+\alpha-1}{y_j}\left(\frac{\beta}{\beta+1}\right)^{\alpha}\left(\frac{1}{\beta+1}\right)^{y_j}I_A(y_j) \tag{3.4}$$

Negative binomial distribution (3.4) is denoted with $NB(\alpha, \frac{\beta}{\beta+1})$.

4. Variable Selection in A Finite Mixture of Negative binomial Semiparametric Models

Consider Y an idea response variable, and $(X,\ U,\ Z)$ is a vector of covariates effective on the response variable. Then, a finite mixture of models with negative binomial response has been suggested as follow:

$$f(y_i; x_i, u_i, z_i, \Psi) = \sum_{k=1}^{K} \pi_k NB(\mu_{ik}, \Phi_k)$$

$$NB(\mu_{ik}, \Phi_k) = \left[\frac{\Gamma(y_i + \Phi_k)}{\Gamma(y_i + 1)\Gamma(\Phi_k)} \left(\frac{\mu_{ik}}{\mu_{ik} + \Phi_k} \right)^{y_i} \left(\frac{\Phi_k}{\mu_{ik} + \Phi_k} \right)^{\Phi_k} \right]$$

$$i = 1, \cdots, n; k = 1, \cdots, K$$

According to relation (3.4), parametric vector Ψ is shown as $\Psi = (\beta_1, \beta_2, \cdots, \beta_k, \alpha_1, \alpha_2, \cdots, \alpha_k, \Phi)$ inclduing $\beta_k = (\beta_{k1}, \cdots, \beta kp)^T$, $\alpha_k = (\alpha_{k1}, \cdots, \alpha)^T$ and $\pi = (\pi_1, \cdots, \pi_{k-1})^T$, so that shows the dimensions of covariates. $\pi_k > 0$ and $\sum_{k=1}^{K} \pi_k = 1$ are possible for vector $\pi \cdot \mu_{ik}$ is the mean of negative binomial distribution for i^{th} observation and k^{th} parameter in relation (4.5) as follows:

$$\mu_{ik}(x_i, u_i, z_i) = \exp\left(x_i^T \alpha_k(u_i) + z_i^T \beta_k\right) \tag{4.5}$$

Variable selection is expressed using penalized likelihood approach and considering negative binomial distribution finite mixture for response variable in three main steps and based on EM algorithm.

First Step: Calculating Non-Parametric Coefficients Local Estimation

Log-likelihood function on the condition of parameter Ψ based on a finite mixture of negative binomial distributions is defined as follows:

$$\mathrm{l}_n(\Psi) = \sum_{i=1}^{n} \log \left\{ \sum_{k=1}^{K} \pi_k f(y_i; \mu_{ik}(x_i, u_i), \Phi_k) \right\} \tag{4.6}$$

As a result, to estimate unknown coefficients based on complete data at the presence of hypothetical marker variable ν_{ik}, the complete log-likelihood function is defined as:

$$\begin{aligned} \mathrm{l}_n^c(\Psi) &= \sum_{i=1}^{n} \sum_{k=1}^{K} \nu_{ik} \left\{ \log \pi_k + \log\{ f(y_i; \mu_{ik}(x_i, u_i), \Phi_k)\} \right\} \\ &= \sum_{i=1}^{n} \sum_{k=1}^{K} \nu_{ik} \left\{ \log \pi_k + \log \left(\frac{\Gamma(y_i + \Phi_k)}{\Gamma(y_i + 1)\Gamma(\Phi_k)} \right) - (\Phi_k + y_i) + y_i \log(\mu_{ik}) + \Phi_k \log(\Phi_k) \right\} \end{aligned} \tag{4.7}$$

Due to using a finite mixture model of negative binomial distributions and non-awareness of nonparametric function $\alpha(\cdot)$, in order to estimate nonparametric section, the method introduced by Lee and Liang (2008) using First-Order Taylor

Linear Approximation, $\alpha_{kj}(\nu)$ is as follows for ν and in the neighborhood of a variable like u:

$$\alpha_{kj}(\nu) \approx \alpha_{kj}(u)+\alpha'_{kj}(u)(\nu-u) \equiv \alpha_{kj}+b_{kj}(\nu-u) \quad k=1,2,\cdots,K; j=1,2,\cdots,P \tag{4.8}$$

Fuction $\alpha(\cdot)$ is represented for simplification for variable u and k^{th} parameter and dimension j^{th} with variable α_{kj} and its first ordered derivative with variable b_{kj} in relation (4.8).

α_{kj} and b_{kj} are Taylor expansion parametric coefficient. Only second order Taylor expansion is considered in relation (4.8) and this is while the more the number of orders increases, the more accurate it will be. In order to smooth the nonparametric section, Taylor series expansion is used to estimate its coefficients.

Therefore, likelihood function based on nonparametric parameter and kernel function $k_h(u_i-u)=\frac{1}{h}k\left(\frac{u_i-u}{h}\right)$ is defined for local and approximate estimation a, b and β as follows:

$$\mathrm{l}_n=\sum_{i=1}^{n}\log\sum_{k=1}^{K}\left\{\pi_k\left[\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\left(\frac{\tilde{\mu}_{ik}}{\tilde{\mu}_{ik}+\Phi_k}\right)^{y_i}\left(\frac{\Phi_{ik}}{\tilde{\mu}_{ik}+\Phi_k}\right)^{\Phi_k}\right]k_h(u_i-u)\right\} \tag{4.9}$$

Where $\tilde{\mu}_{ik}(u_i,x_i,z_i))$ is defined based on a and b as follows:

$$\tilde{\mu}_{ik}(u_i,x_i,z_i))=\exp(x_i^T\alpha_k+x_i^Tb_k(u_i-u)+z_i^T\beta_k) \quad i=1,2,\cdots,n; k=1,2,\cdots,K \tag{4.10}$$

In order to maximize and calculate an optimal value based on relation (4.10), the complete likelihood function is defined as follows:

$$\begin{aligned}\mathrm{l}_n^c(\Psi)=\sum_{i=1}^{n}\sum_{k=1}^{K}\nu_{ik}\{\log\pi_k+\log\left(\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\right)-(\Phi_k+y_i)\log(\tilde{\mu}_{ik}+\Phi_k)\\ +y_i\log(\tilde{\mu}_{ik})+\Phi_k\log(\Phi_k)+\log k_h(u_i-u)\}\end{aligned} \tag{4.11}$$

$k_h(u_i-u)$, h is the width of band in kernel function. After specifying $\tilde{\mu}_{ik}$, optimal value of relation (4.10) is calculated using logarithm EM and estimating parameters and nonparametric section coefficients.

step E:

In this step, conditional expectation $\mathrm{l}_n^c(\Psi)$ on the condition of ν_{ik} based on observations (u_i, x_i, z_i, y_i) is defined as follows:

$$Q\left(\Psi; \Psi^{(m)}\right) = \sum_{i=1}^{n}\sum_{k=1}^{K}\omega_{ik}^{(m)}[\log\left(\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\right) - (\Phi_k + y_i)\log(\tilde{\mu}_{ik}+\Phi_k)$$
$$+ y_i \log(\tilde{\mu}_{ik}) + \Phi_k \log(\Phi_k)] + \sum_{i=1}^{n}\sum_{k=1}^{K}\omega_{ik}^{(m)} \log \pi_k$$
$$+ \sum_{i=1}^{n}\sum_{k=1}^{K}\omega_{ik}^{(m)} \log k_h(u_i - u) \tag{4.12}$$

So that $\omega_{ik}^{(m)}$s are conditional expectation of ν_{ik}s on the condition of observations and are available as weight values as follows:

$$\omega_{ik}^{(m)} = \frac{\pi_k^{(m)}\left[\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\left(\frac{\tilde{\mu}_{ik}^{(m)}}{\tilde{\mu}_{ik}^{(m)}+\Phi_k}\right)^{y_i}\left(\frac{\Phi_{ik}}{\tilde{\mu}_{ik}^{(m)}+\Phi_k}\right)^{\Phi_k}\right]}{\sum_{k=1}^{K}\pi_k^{(m)}\left[\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\left(\frac{\tilde{\mu}_{ik}^{(m)}}{\tilde{\mu}_{ik}^{(m)}+\Phi_k}\right)^{y_i}\left(\frac{\Phi_{ik}}{\tilde{\mu}_{ik}^{(m)}+\Phi_k}\right)^{\Phi_k}\right]} \tag{4.13}$$

Step M:

$Q\left(\Psi; \Psi^{(m)}\right)$ is maximized in the $(m+1)^{th}$ step of repetition stage than the components of parametric vector Ψ. When using EM algorithm, maximizing $Q\left(\Psi; \Psi^{(m)}\right)$ based on mixture ratios as well as other parameters has computational complexities. Therefore, it is necessary to calculate mixture rations $\pi_k^{(m+1)}$ in each step based on weights $\omega_{ik}^{(m)}$ as follows:

$$\pi_k^{(m+1)} = \frac{1}{n}\sum_{i=1}^{n}\omega_{ik}^{(m)} \qquad k = 1, 2, \cdots, K \tag{4.14}$$

It is possible to maximize $Q\left(\Psi; \Psi^{(m)}\right)$ on the condition of $\pi_k^{(m+1)}$ than a, b and β. According to relation (4.14), a and b are nonparametric coefficients that using EM algorithm, approximate results are estimated. Thus, we have to solve the following equations to determine the estimations of coefficients related to parametric and non-parametric section:

$$\sum_{i=1}^{n}\omega_{ik}^{(m)}\frac{\partial}{\partial\beta_{kj}}\{\log\left(\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\right) - (\Phi_k + y_i)\log(\tilde{\mu}_{ik}+\Phi_k) + y_i\log(\tilde{\mu}_{ik})$$
$$+ \Phi_k \log(\Phi_k) + \log k_h(u_i - u)\} = 0 \tag{4.15}$$

$$\sum_{i=1}^{n}\omega_{ik}^{(m)}\frac{\partial}{\partial\alpha_{kj}}\{\log\left(\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\right)-(\Phi_k+y_i)\log(\tilde{\mu}_{ik}+\Phi_k)+y_i\log(\tilde{\mu}_{ik})$$
$$+\Phi_k\log(\Phi_k)+\log k_h(u_i-u)\}=0 \tag{4.16}$$

$$\sum_{i=1}^{n}\omega_{ik}^{(m)}\frac{\partial}{\partial b_{kj}}\{\log\left(\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\right)-(\Phi_k+y_i)\log(\tilde{\mu}_{ik}+\Phi_k)+y_i\log(\tilde{\mu}_{ik})$$
$$+\Phi_k\log(\Phi_k)+\log k_h(u_i-u)\}=0 \tag{4.17}$$

According to the above equations, a complex equational system of unknowns has been created, although it is not possible to solve it manually, and it is not possible to achieve a closed form for them explicitly. Then the estimations will be calculated in the following.

Second Step: Calculating Penalized β Coefficient Estimation

First, in order to estimate penalized β coefficient, the following likelihood function is defined:

$$\mathrm{l}(\beta)=\sum_{i=1}^{n}\log\sum_{k=1}^{K}\left\{\pi_k\left[\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\left(\frac{\mu_{ik}^*}{\mu_{ik}^*+\Phi_k}\right)^{y_i}\left(\frac{\Phi_k}{\mu_{ik}^*+\Phi_k}\right)^{\Phi_k}\right]\right\} \tag{4.18}$$

Where μ_{ik}^* is as follows:

$$\mu_{ik}^*(u_i,x_{izi})=\exp\left(x_i^T\tilde{\alpha}_k(u_i)+z_i^T\beta_k\right)\qquad k=1,2,\cdots,K \tag{4.19}$$

μ_{ik}^* and $\tilde{\mu}_{ik}^{(m)}$ are different from each other in that in relation (4.19) instead of unknown function of previous step nonparametric coefficients, the optimal estimation resulted from the first step is used to calculate more accurate estimations for parametric coefficients. Maximizing likelihood function $\mathrm{l}(\beta)$ than β, the estimation of parametric section coefficients are achieved.

Step E:

Like conditional expectation in the first step, the conditional expectation of completely log-likelihood function $\mathrm{l}(\beta)$ is calculated on the condition of marker variables ν_{ik}s and observations (u_i,x_i,z,y_i):

$$Q\left(\Psi;\Psi^{(m)}\right)=\sum_{i=1}^{n}\sum_{k=1}^{K}\omega_{ik}^{(m)}[\log\left(\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\right)-(\Phi_k+y_i)\log(\mu_{ik}^*+\Phi_k)$$
$$+y_i\log(\mu_{ik}^*)+\Phi_k\log(\Phi_k)]+\sum_{i=1}^{n}\sum_{k=1}^{K}\omega_{ik}^{(m)}\log\pi_k \tag{4.20}$$

Where $\omega_{ik}^{(m)}$ is as follows:

$$\omega_{ik}^{(m)} = \frac{\pi_k^{(m)} \left[\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \left(\frac{\mu_{ik}^{*\,(m)}}{\mu_{ik}^{*\,(m)}+\Phi_k} \right)^{y_i} \left(\frac{\Phi_{ik}}{\mu_{ik}^{*\,(m)}+\Phi_k} \right)^{\Phi_k} \right]}{\sum_{k=1}^{K} \pi_k^{(m)} \left[\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \left(\frac{\mu_{ik}^{*\,(m)}}{\mu_{ik}^{*\,(m)}+\Phi_k} \right)^{y_i} \left(\frac{\Phi_{ik}}{\mu_{ik}^{*\,(m)}+\Phi_k} \right)^{\Phi_k} \right]} \tag{4.21}$$

Now it is possible to use the local second-order approximation of $P_n(\Psi)$ instead of itself.

Step M:

$Q\left(\Psi; \Psi^{(m)}\right)$ defined in step E is maximized in spite of nonparametric section local estimation than its unknown parameters in the $(m+1)^{th}$ step of repetition. Like the first step, after updating mixing probabilities, considering π_k fixed in $Q\left(\Psi; \Psi^{(m)}\right)$, it is maximized than β and there will be:

$$\sum_{i=1}^{n} \omega_{ik}^{(m)} \frac{\partial}{\partial \beta_{kj}} \{\log \left(\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \right) - (\Phi_k + y_i) \log(\mu_{ik}^{*\,(m)} + \Phi_k)$$
$$+ y_i \log(\mu_{ik}^{*\,(m)}) + \Phi_k \log(\Phi_k)\} = 0 \tag{4.22}$$

In fact, it is possible to use penalty functions $p_n^L\left(\Psi; \Psi^{(m)}\right)$ and or $p_n^S\left(\Psi; \Psi^{(m)}\right)$ for values $k = 1, 2, \cdots, K$ and $j = 1, 2, \cdots, P$ separately instead of $p_n(\beta_{kj})$ considering the type of penalty function used in step E in step M. It is important to consider that optimal value is achieved when after the repetition of several stages for steps E and M, the difference of Euclidean norm for estimation values of parametric coefficients for two sequential stages as $\| \beta_{11}^{(m+1)} - \beta_{11}^{(m)} \|$ is less than desirable small value like δ. In such a step, the penalized parametric coefficients of $\hat{\beta}$ are achieved, and the homogeneity is confirmed.

Third Step: Calculating Nonparametric Coefficient Accurate Estimation

It is possible to use the penalized estimations of parametric coefficient $\hat{\beta}$ of the previous step and put them instaed of nonpenalized local estimations of the first step. In fact, in this step, $\tilde{\mu}_{ik}^{*}$ is substituted with $\tilde{\mu}_{ik}$ in the first step and is defined as follows:

$$\tilde{\mu}_{ik}^{*} = \exp\left(x_i^T \alpha_k + x_i^T b_k(u_i - u) + z_i^T \hat{\beta}_k \right) \tag{4.23}$$

In fact, in the definition of $\tilde{\mu}_{ik}^{*}$, the nonparametric coefficients a and b are considered unknown, and the parametric estimation of the second stage regression coefficients is replaced. This step mainly aims at accurately calculating a and b instead of local estimations of $\tilde{a}$ and $\tilde{b}$ of the first step at the presence of regression coefficients penalized estimations. Likelihood function in this step based on $\tilde{\mu}^*_{ik}$ is as follows:

$$\sum_{i=1}^{n} \log \sum_{k=1}^{K} \left\{ \pi_k \left[\frac{\Gamma(y_i + \Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \left(\frac{\tilde{\mu}^*_{ik}}{\tilde{\mu}^*_{ik} + \Phi_k} \right)^{y_i} \left(\frac{\Phi_k}{\tilde{\mu}^*_{ik} + \Phi_k} \right)^{\Phi_k} \right] \right\} \tag{4.24}$$

As a result, the complete likelihood function at the presence of variables ν_{ik} is as follows:

$$\begin{aligned} \mathrm{l}^c_n(\Psi) = \sum_{i=1}^{n}\sum_{k=1}^{K} \nu_{ik} \{ \log \pi_k + \log \left(\frac{\Gamma(y_i + \Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \right) - (\Phi_k + y_i) \log(\tilde{\mu}^*_{ik} + \Phi_k) \\ + y_i \log(\tilde{\mu}^*_{ik}) + \Phi_k \log(\Phi_k) + \log k_h(u_i - u) \} \end{aligned} \tag{4.25}$$

Step E:

Conditional expectation $\mathrm{l}^c_n(\Psi)$ on the condition of invisible marker variables ν_{ik} and observations (u_i, x_i, z_i, y_i) is defined as follows:

$$\begin{aligned} Q\left(\Psi; \Psi^{(m)}\right) = \sum_{i=1}^{n}\sum_{k=1}^{K} \omega^{(m)}_{ik} [\log \left(\frac{\Gamma(y_i + \Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \right) - (\Phi_k + y_i) \log(\tilde{\mu}^*_{ik} + \Phi_k) \\ + y_i \log(\tilde{\mu}^*_{ik}) + \Phi_k \log(\Phi_k)] \\ + \sum_{i=1}^{n}\sum_{k=1}^{K} \omega^{(m)}_{ik} \log \pi_k + \sum_{i=1}^{n}\sum_{k=1}^{K} \omega^{(m)}_{ik} \log k_h(u_i - u) \end{aligned} \tag{4.26}$$

Where $\omega^{(m)}_{ik}$s are calculated as follows:

$$\omega^{(m)}_{ik} = \frac{\pi^{(m)}_k \left[\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \left(\frac{\tilde{\mu}^{*\,(m)}_{ik}}{\tilde{\mu}^{*\,(m)}_{ik} + \Phi_k} \right)^{y_i} \left(\frac{\Phi_{ik}}{\tilde{\mu}^{*\,(m)}_{ik} + \Phi_k} \right)^{\Phi_k} \right]}{\sum_{k=1}^{K} \pi^{(m)}_k \left[\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \left(\frac{\tilde{\mu}^{*\,(m)}_{ik}}{\tilde{\mu}^{*\,(m)}_{ik} + \Phi_k} \right)^{y_i} \left(\frac{\Phi_{ik}}{\tilde{\mu}^{*\,(m)}_{ik} + \Phi_k} \right)^{\Phi_k} \right]} \tag{4.27}$$

Step M:

$Q\left(\Psi; \Psi^{(m)}\right)$ is maximized than unknown nonparametric coefficients in the $(m+1)^{th}$ step of repetition. In fact, it is possible after updating mixing probabilities, so considering π_k fixed and scolding equations (4.27) and (4.27) there will be:

$$\begin{aligned} \sum_{i=1}^{n} \omega^{(m)}_{ik} \frac{\partial}{\partial \alpha_{kj}} \{ \log \left(\frac{\Gamma(y_i + \Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)} \right) - (\Phi_k + y_i) \log(\tilde{\mu}^{*\,(m)}_{ik} + \Phi_k) + y_i \log(\tilde{\mu}^{*\,(m)}_{ik}) \\ + \Phi_k \log(\Phi_k) + \log k_h(u_i - u) \} = 0 \end{aligned} \tag{4.28}$$

$$\sum_{i=1}^{n}\omega_{ik}^{(m)}\frac{\partial}{\partial b_{kj}}\{\log\left(\frac{\Gamma(y_i+\Phi_k)}{\Gamma(y_i+1)\Gamma(\Phi_k)}\right)-(\Phi_k+y_i)\log(\tilde{\mu}_{ik}^{*\,(m)}+\Phi_k)+y_i\log(\tilde{\mu}_{ik}^{*\,(m)})$$
$$+\Phi_k\log(\Phi_k)+\log k_h(u_i-u)\}=0 \tag{4.29}$$

After solving equation (4.29) in the step M of the second stage, in order to estimate unknown coefficients of the nonparametric section more accurately against local estimations of the nonparametric coefficients at the first stage, equations (4.28) and (4.29) will be solved. Like the first stage, there is another complex equational system here that is impossible to achieve a closed form for than explicitly. As a result, the optimal response of the equations system is achieved with the necessary number of repetitions. When the EM algorithm becomes homogenous after passing stages E and M, the main estimations of $\hat{a}$ and $\hat{b}$ are achieved. Therefore, nonparametric coefficients accurate estimations of the third stage and parametric coefficients penalized estimations of second stage are resulted as $\{\hat{a},\hat{b},\hat{\beta}\}$. Calculations are done generally in four steps:

a- Considering an initial value for β called β^0.

b- Calculating local estimation of nonparametric function as $\hat{\alpha}(u)=\hat{\alpha}$.

c- Calculating the estimation of parametric coefficient β called β^0 at the presence of nonparametric function $\hat{\alpha}(u)$.

d- Calculating the accurate estimation of nonparametric function based on coefficients a and b of Taylor expansion at the presence of coefficient β called $\hat{a}$ and $\hat{b}$.

The final results are considered after repeating steps E and M and EM algorithm homogeneity in each of the computational steps.

5. Simulation Study

In order to compare the results of using negative binomial distribution instead of Poisson distribution in a finite mixture of models, a simulation program was used to estimate parameters. The following steps have been done:

1- First as primary value, parameters vector considered as $\beta^0=\begin{pmatrix}1 & 0.7\\ 0.5 & -0.8\end{pmatrix}$ and simulation algorithm is repeated for 100 times. Since it is possible to investigate both models based on primary value, so, comparing these two models does not affect on the results.

2- First, two Poisson distributions are written as a linear combination, so that the weight of both Poisson distributions will be (0.450146&0.549854). This will be (0.492196&0.507804) for negative binomial distribution.

3- Then, the log-likelihood function for observations is achieved based on a mixture of two Poisson distributions and a mixture of two negative binomial distributions. According to the results in table 1 based on the log-likelihood function, a mixture of two negative binomial distributions is more valid than a mixture of two Poisson distributions. The values are considerably different.

Table 1: estimating the logarithm values of the likelihood function for a mixture of distributions

Mixture type of distributions	The estimation of log-likelihood function
Poisson two distribution combination	-307.3718
Negative binomial two distribution combination	-289.276

According to Table 1, the statistic of logarithm values of likelihood function is bigger in using negative binomial distribution than Poisson distribution. This shows that the fitness of the model is better in using a finite mixture of generalized linear models based on negative binomial distribution than a finite mixture of generalized linear models based on Poisson distribution. Also, Table 2 shows the estimates of parameters and the estimates of the variance in both mixture distributions. Wald statistics for estimators determined in both mixture distributions still confirms the related hypotheses for the finite mixture of two negative binomial distribution. In

Table 2: the estimation of parameters and determining statistic based on Poisson two distribution and negative binomial distribution

The type of distribution	The estimation of parameters	The estimation of estimator variance	Wald statistic	Sig
first Poisson distribution	3.88	1.067	14.109	The hypothesis is rejected
second Poisson distribution	0.499	0.00005	4.98	The hypothesis is rejected
First negative binomial distribution	0.588	0.356	0.874	The hypothesis is approved
second negative binomial distribution	0.541	0.00006	4.87	The hypothesis is approved

order to investigate that change of distribution from a finite mixture of two Poisson distributions to a finite mixture of two negative binomial distributions is suitable, the statistic of Weighted Generalized Mean Squared Error will be used, which is shown as WGMSE and is defined as follows:

$$WGMSE = \gamma\Lambda_1 + (1-\gamma)\Lambda_2$$

So that for $i = 1, 2$ there is:

$$\Lambda_i = (\hat{\beta_{i0}}, \hat{\beta_{i1}})^T \begin{pmatrix} Var(\hat{\beta_{i0}}) & Cov(\hat{\beta_{i0}}, \hat{\beta_{i1}}) \\ Cov(\hat{\beta_{i0}}, \hat{\beta_{i1}}) & Var(\hat{\beta_{i1}}) \end{pmatrix} (\hat{\beta_{i0}}, \hat{\beta_{i1}})$$

Considering that estimated weights in a finite mixture of Poisson distributions are:

$$\gamma^{pois} = (0.55, 0.45)$$

and Λ_i values for Poisson distribution are:

$$\Lambda^{pois} = (0.783, 0.736)$$

Therefore, there will be:

$$WGMSE^{pois} = (0.55 * 0.783) + (0.45 * 0.736) = 0.762$$

If a finite mixture of negative binomial distributions is considered for, then, there will be:

$$\gamma^{nbinom} = (0.43, 0.57)$$

And there is $\Lambda^{nbinom} = (0.105, 0.068)$. As a result:

$$WGMSE^{nbinom} = (0.43 * 0.105) + (0.57 * 0.068) = 0.083$$

Regarding that, the mixture of two negative binomial distributions has been replaced with a mixture of two Poisson distributions, and with comparing between them using the mixture of efficiency or (ME_{NB-PO}), we have:

$$ME_{NB-PO} = \frac{0.762}{0.083} = 9.18$$

Figure 1 shows the superiority of negative binomial distributions over Poisson distributions based on 30 times repeating the suggested model simulation. The dotted line in Figure 1 shows values related to Λ in Poisson distribution, and the continuous diagram is related to Poisson distribution after 30 time repetition.

6. Practical Example

The statistical population of this study is related to the number of participants of the doctoral exam during 2019 and 2020, which has been held as a national wide exam by National Organization for Educational Testing. Each of these tests includes 45 questions. The number of 1540 and the number of 1480 individuals

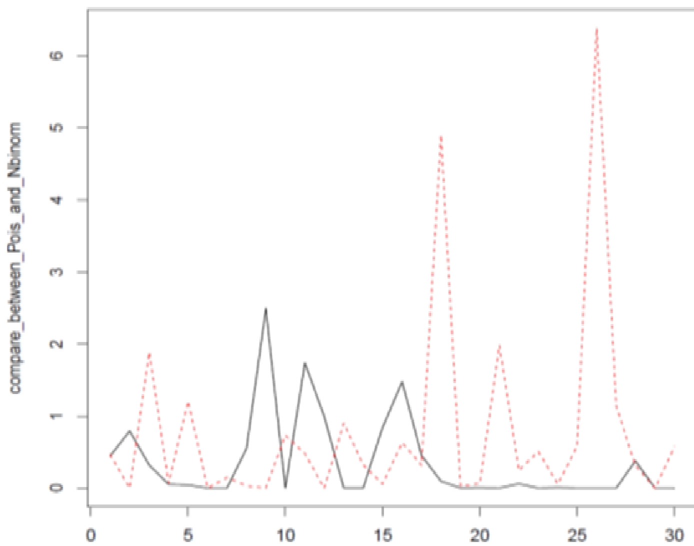

Figure 1: Comparison between Λ values between mixture of negative binomial distribution and Poisson distribution.

have participated in the statistics Ph.D. degree exam in 2019 and 2020, respectively, and the information of all of them has been used in this study. We have the information such as the type of bachelor's and master's degree university (government, Payam Noor, non-profit, free university and scientific-applied), type of admission period (daily, second round, Payam Noor, non-profit, etc.), gender, and year of birth. Using this information, we have formed a number 4 Table in which the name and code of the master's degree university, the number of admissions and rejections, as well as the number of female admissions in 2019 and 2020, are listed. For example, the University of Isfahan in 2019 had 625 participants in the doctoral exam, of which 49 were accepted, one of whom was a woman. However, this university had 586 participants in the doctoral exam in 2020, of which 108 were admitted, and 18 were female. Imam Khomeini International University had 164 and 256 participants in 2019 and 2020, respectively, of which 25 and 96 were accepted, respectively, none of whom were women. The highest number of women admitted is related to the University of Science and Technology, which in 2019 and 2020, respectively 210 and 108 were among those accepted. Now, the finite mixture model of Poisson distributions and the finite mixture of negative binomial distributions are fitted to the data. Therefore, first, the logarithm of both models likelihood function is estimated after fitting both models. Table 3 shows these values. The results of Table 3 show the superiority of the finite mixture of generalized linear models based on negative binomial distribution over the finite mixture of generalized linear models based on Poisson distribution. Table 4 shows

Table 3: the estimation of the logarithm of likelihood function for a mixture of distributions

Mixture type of distributions	The estimation of the logarithm likelihood function
Poisson	-132.7962
Negative binomial	-48.9094

the estimation of model parameters and the estimation of variance. Wald statistic shows that the hypotheses of fixed values not influencing and quantitative unit are effective on the response variable. WGMSE will be used to investigate whether

Table 4: parameters estimation and determining tests statistic based on Poisson distributions and negative binomial distributions

The type of distribution	The estimation of parameters	The variance estimator	Wald statistic	Sig
Poisson first distribution	0.0381	0.001	1.204	The hypothesis is approved
Poisson second distribution	0.022	0.001	0.695	The hypothesis is approved
First negative binomial distribution	0.029	0.001	0.917	The hypothesis is approved
second negative binomial distribution	0.064	0.001	2.02	The hypothesis is approved

finite mixture distribution change from Poisson two distribution combination to a finite mixture of two distributions negative binomial distribution combination is suitable or not and it is:

$$WGMSE = \gamma\Lambda_1 + (1-\gamma)\Lambda_2$$

So that for $i = 1, 2$ there is:

$$\Lambda_i = (\hat{\beta_{i0}}, \hat{\beta_{i1}})^T \begin{pmatrix} Var(\hat{\beta_{i0}}) & Cov(\hat{\beta_{i0}}, \hat{\beta_{i1}}) \\ Cov(\hat{\beta_{i0}}, \hat{\beta_{i1}}) & Var(\hat{\beta_{i1}}) \end{pmatrix} (\hat{\beta_{i0}}, \hat{\beta_{i1}})$$

Considering that estimated weights in a finite mixture of Poisson distributions are:

$$\gamma^{pois} = (0.50, 0.50)$$

and Λ_i values for Poisson distribution are:

$$\Lambda^{pois} = (0.432, 0.697)$$

Therefore, there will be:

$$WGMSE^{pois} = (0.50 * 0.432) + (0.50 * 0.697) = 0.564$$

If it is considered for a finite mixture of negative binomial distributions, there will be:

$$\gamma^{nbinom} = (0.43, 0.57)$$

And there is $\Lambda^{nbinom} = (1.016, 1.163)$ As a result:

$$WGMSE^{nbinom} = (0.995 * 0.105) + (0.005 * 0.068) = 0.1048$$

According to two distributions, negative binomial distribution has been replaced with two distribution Poisson distribution, comparing the statistics of tests showed less value. In fact there is:

The mixture efficiency of negative binomial than the mixture of Poisson distribution

$$\frac{0.564}{0.1048} = 5.42$$

The study conducted on data related to the Ph.D. degree exam shows that using a finite mixture of negative binomial distributions is considerably superior to using a finite mixture of Poisson two distributions. The ability variable has also been determined as effective viable in both societies (2019 and 2020). This has been confirmed by the simulation study.

References

Chen and Holland (2010). New Equating Methods and Their Relationships With Levine Observed Score Linear Equating Under The Kernel Equating Framework. *Psychometrika*, 75, **3**, 542–557.

Cho, H. and Fryzlewicz, P. (2012), High-Dimensional Variable Selection via Tilting, *J. Roy. Stat. Soc. Ser. B, Stat. Method.*, **74**, 593-622.

Du, Y., Khalili, A., Neslehova, J.G., and Steele, R.J., (2013), "Simultaneous Fixed and Random Effects Selection in Finite Mixture of Linear Mixed-Effects Models, *The Canadian Journal of Statistics*, **41**, 596-616.

Eskandari, F., Ormoz, E. (2016), Finite Mixture of Generalized Semiparametric Models: Variable Selection via Penalized Estimation, *Communications in Statistics – Simulation and Computation*,(to appear).

Fan, J. and Li, R. (2001), Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties, *J. Amer. Statist. Assoc.*, **96**, 1348-1360.

Keller, L. A., Skorupski, W. P., Swaminathan, H. and Jodoin, M. G. (2004). *An Evaluation of Item Response Theory Equating Procedures for Capturing Changes in Examinee Distributions with Mixed-Format Tests.* Paper Presented at the Annual Meeting of the National Council on Measurement in Education.

Khalili, A. and Chen, J. (2007), Variable Selection in Finite Mixture of Regression Models, *Journal of American Statistical Association*, **102**, 1025-1038.

Khalili, A. (2011), An Overview of the New Feature Selection Methods in Finite Mixture of Regression Mdels, *JIRSS*, **10**, 201-235.

Kolen, M. J. and Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and 31 Practices*. New York: Springer.

Li, R. and Liang, H. (2008), Variable Selection in Semiparametric Regression Modeling, *Ann. Statist.*, **36**, 261-286.

Lee, W. and Ban, J. (2010). A comparison of IRT Linking Procedures. *Applied Measurement in Education*, **23**, 23–48.

Ma, S., Song, Q. and Wang, L. (2013), Simultaneous Variable Selection and Estimation in Semiparametric Modeling of Longitudinal/Clustered Data, *Bernoulli*, **19**, 252-274.

McLachlan, G. J. and Peel, D. (2000), *Finite Mixture Models*, New York: Wiley.

Nelder, J., and Wedderburn, R.W.M, (1972), Generalized Linear Models, *J. Roy. Statist. Soc. Ser. A.*, **135**, 370-384.

Myint, A. and Htet L., O. (2010), An Application of Linear Test Equating Method in Scoring. *Yangon Institute of Education Research Vol. 2*, **1**, 1-16.

Ormoz, E., and Eskandari, F., (2013), Variable Selection in Finite Mixture of Semi-Parametric Regression models, *Commun Stat-Theorm*, to appear.

Ormoz, E., Eskandari, F., (2016), Variable Selection in Finite Mixture of Semi-Parametric Regression Models, *Communications in Statistics-Theory and Methods*, **3**, 657-670.

Santarelli, M. F., Latta, D.D., Michele Scipioni, M., Positano,V.and Landini, L.(2016). A Conway–Maxwell–Poisson (CMP) model to address data dispersion on positron emission tomography. *Computers in Biology and Medicine*, **77**,90–101.

Skrondal, A. and Rabe-Hesketh, S. (2004), *Generalized Latent Variable Modelling: Multilevel, Longitudinal, and Structural Equations Models*, Chapman and Hall.

Schwarz, G. (1978), Estimating the Dimension of a Model, *The Annals of Statistics*, **6**, 461-464.

Von Davier, A. A., Holland, P. W., and Thayer, D. T. (2004). *The kernel Method of Test Equating*. New York: Springer-Verlag.

Von Davier, A. A. (2011). *A Statistical Perspective on Equating Test Scores. Scaling and Linking*. New York, NY: Springer-Verlag.

Santarelli, M. F., Latta, D.D., Michele Scipioni, M., Positano,V.and Landini, L.(2016). A Conway–Maxwell–Poisson (CMP) model to address data dispersion on positron emission tomography. *Computers in Biology and Medicine*, **77**, 90–101.