

A finite mixture of multiple discrete distributions for modelling heaped count data

Lluís Bermúdez^{a,*}, Dimitris Karlis^b, Miguel Santolino^c

^a*Universitat de Barcelona, Riskcenter-IREA, Departament de Matemàtica Econòmica, Financera i Actuarial, Spain*

^b*Athens University of Economics and Business, Department of Statistics, Greece*

^c*Universitat de Barcelona, Riskcenter-IREA, Departament d'Estadística, Econometria i Economia Espanyola, Spain*

Abstract

A new modelling approach, based on finite mixtures of multiple discrete distributions of different multiplicities, is proposed to fit data with a lot of periodic spikes in certain values. An EM algorithm is provided in order to ensure the models' ease-of-fit and then a simulation study is presented to show its efficiency. A numerical application with a real data set involving the length, measured in days, of inability to work after an accident occurs is treated. The main finding is that the model provides a very good fit when working week, calendar week and month multiplicities are taken into account.

Keywords: Digit preference, EM algorithm, Multiple Poisson, Work disability days

1. Introduction and aim

This paper aims at presenting a new model to fit frequency data with a lot of spikes in certain values. This paper is motivated by the need to model the duration, measured in days, of inability to work (also called work disability days) after a motor accident. In this case, the frequency distribution exhibits spikes at multiples of 5, 7 and 30 days, implying perhaps the different time scales used by doctors when deciding on the number of days of sick leave for workers. For example, a doctor may think on a daily or weekly or monthly

***Corresponding Author.** Departament de Matemàtica Econòmica, Financera i Actuarial, Universitat de Barcelona, Diagonal 690, 08034-Barcelona, Spain. Tel.:+34-93-4034853; fax: +34-93-4034892; e-mail: lbermudez@ub.edu

scale for the number of days of absence for a patient based on the severity of certain injuries. It is rather improbable to say 13 days of absence but more probable to say 14 days (2 weeks).

We emphasize that such periodic peaks, or regularly spaced spikes within the overall data layout, observed in a frequency distribution can occur in many other applications. A typical example is the phenomenon known as digit preference, i.e. the tendency to round outcomes to pleasing digits. For reported count data, this phenomenon is referred to as heaping (or coarsening data), i.e. rounding exact counts to even multiples of reported units. Age misreporting (age heaping) has long been an issue in demography (for a review, see Siegel and Swanson, 2004) with observed peaks in ages ending with 0 or 5. Some other examples of heaped count data include number of cigarettes smoked in a day (Wang et al., 2012), blood pressure measurements (Nietert et al., 2006), unemployment duration (Torelli and Trivellato, 1993; Wolff and Augustin, 2003), age of death (Camarda et al., 2008), or birth weight (Channon et al., 2011).

The literature in digit preference or heaping assumes that data can be interpreted as indirect (or rounded) observations of a latent distribution. In consequence, the basic idea of modelling approaches to address heaped data is to define a model for the latent exact values, usually, by smoothing response heaping. Literature include different approaches to deal with heaped data. For instance, Camarda et al. (2008) used a composite link model, Wang et al. (2012) used the method of multiple imputation, Wang and Wertelecki (2013) proposed a nonparametric bootstrap kernel density estimator, Wang and Heitjan (2008) proposed a Bayesian proportional odds rounding behavior model.

However, the modelling approach proposed in this paper is very different. Planners and managers may be interested in modelling the data itself, not the latent distribution. In such a case, the aim is not to smooth the data to obtain the latent distribution anymore, but to propose count models that fit the data with periodic peaks appropriately. For instance, for the work disability days' application, one may interested in the analysis of the financial cost of disability, which is legally obtained from the days of absence reported. Smoothing models may not reflect how people (doctors, in the application case) behave and, hence, they may underestimate the financial cost.

In this paper, a new modelling approach based on finite mixtures of multiple Poisson (and negative binomial) distribution of different multiplicities are proposed to fit this kind of data. Simultaneously, an EM algorithm is

provided as a standard tool for maximum likelihood estimation of these finite mixture models.

The modelling approach is described in Section 2. In Section 3 an EM algorithm is proposed for maximum likelihood (ML) estimation and a simulation study is conducted to assess its efficiency. A numerical application with a real data set of days of absence from work is presented in Section 4. Finally, some concluding remarks are given in Section 5.

2. The proposed modelling approach

First of all, consider a simple Poisson distribution with parameter λ denoted as $Po(\lambda)$ with probability density function (pdf) given by

$$P(X = x) = \frac{\exp(-\lambda)\lambda^x}{x!}, \quad \lambda \geq 0, \quad x = 0, 1, 2, \dots$$

We extend by defining the multiple Poisson, with multiplicity m and parameter λ , denoted as $P_m(\lambda)$ having pdf:

$$P_m(X = y) = \begin{cases} \frac{\exp(-\lambda)\lambda^x}{x!} & y = mx, \quad x = 0, 1, 2, \dots \\ 0 & \text{otherwise.} \end{cases}$$

Hence this distribution gives positive probability to points $0, m, 2m, \dots$ and 0 elsewhere. The idea for this representation is that we can have data measured in different scales, as for example observations measured in days, weeks and months. In this case, $P_1(\lambda_1)$, $P_7(\lambda_7)$ and $P_{30}(\lambda_{30})$ respectively.

The derivation of the above model has been previously reach in the literature. It can be interpreted as an special case of what is known as stuttering Poisson model (see Patel, 1976). The distribution is also referred as a scaled Poisson distribution in (Jorgensen, 1997, p.66). One can also note that it can be seen as a Poisson random sum (compound Poisson) of the form $X = Y_1 + \dots + Y_N$ where N is a Poisson random variable and Y 's follow a degenerate distribution at m points. Finally, the distribution is related to the power variance function model (see, e.g. Vinogradov, 2004). However, our definitive modelling approach, based on the above model and explained bellow, is new up to our knowledge.

Following the above multiple Poisson distribution with parameters m and λ , we got that

$$E(X) = m\lambda \quad \text{and} \quad Var(X) = m^2\lambda.$$

Also, the probability generating function takes the form

$$G(t) = \exp(\lambda(t^m - 1)).$$

While we showed the definition based on the Poisson distribution this can be extended to any discrete distribution.

2.1. A finite mixture of multiple Poisson model

The idea is to construct a new model to fit count data that exhibits periodic peaks in the frequency distribution. This peculiarity of the data can be attributed to the fact that different scales are used implicitly at the same time. For example, based on the severity of some injury, a doctor may think on a daily or weekly or monthly scale for the number of days of absence for a patient. This fact leads naturally to a finite mixture and, hence, a finite mixture of multiple Poisson distributions of different multiplicities is proposed here to account for such a feature of the data.

Denote by Y_i the count of some event for $i = 1, \dots, n$ observations. Conditional on the fact that the i -th observation belongs to the j -th group or component, the observed counts come from a Poisson distribution with mean λ_j and multiplicity m_j . The probability that a randomly selected individual belongs to group j is $0 < \pi_j < 1$ for $j = 1, \dots, k$ with $\sum \pi_j = 1$. Then the unconditional pdf is

$$P(Y_i = y) = P(y) = \sum_{j=1}^k \pi_j P_{m_j}(y|\lambda_j), \quad y = 0, 1, \dots$$

where k is the number of groups, $\lambda_j > 0$ are the means for each group, m_j are the multiplicities. This is a typical finite mixture model. This implies that many of the ideas developed so far for finite mixtures are applicable here, like derivation of an EM type algorithm or the selection of the number of components. Note that the model allows for different components with the same multiplicity.

An interesting point that arises for this model is that the multiplicities m_j need to be known *a priori*. Therefore they should be set beforehand based on the context and the knowledge about the data. Typically one recognizes from the data the position of peaks on the frequencies implying certain multiplicities.

This simple model can be extended if we allow the mean of each component to depend on some covariates related to the i -th individual, but this is not pursued in the present paper.

However, this model, based on the Poisson distribution, can be extended to any discrete distribution. In the next subsection, an extension to a negative binomial distribution is presented.

2.2. A finite mixture of multiple negative binomial model

One reason for such an extension lies on the fact that for certain multiplicities assuming a single Poisson component is perhaps limited and one needs more components to account for the variability. Two natural choices for that is either to assume more components with the same multiplicity, which is equivalent to attempting to model overdispersion with finite mixtures, or to assume for the components a more general model that allows for greater variability. We describe a multiple negative binomial distribution and its mixture in this section.

Using a similar approach we may define a multiple negative binomial distribution as

$$G_m(X = y) = \begin{cases} \left(\frac{r}{r+\mu}\right)^r \frac{\Gamma(r+x)}{x!\Gamma(r)} \left(\frac{\mu}{r+\mu}\right)^x & y = mx, \quad x = 0, 1, 2, \dots \\ 0 & \text{otherwise.} \end{cases}$$

in a similar manner. Note that we use the mean parametrization and hence μ is the mean of the negative binomial distribution. The variance is $\mu + \mu^2/r$ and hence for $r \rightarrow \infty$ we get back the Poisson distribution. Again this distribution gives positive probability to points $0, m, 2m, \dots$ and 0 elsewhere.

We can extend this to a finite mixture, as for the multiple Poisson, by considering:

$$P(Y_i = y) = \sum_{j=1}^K \pi_j G_{m_j}(y|\mu_j, r_j), \quad y = 0, 1, \dots$$

with $0 < \pi_j < 1$ for $j = 1, \dots, k$ with $\sum \pi_j = 1$. Again the m_j are the multiplicities for the different components and μ_j, r_j are the specific parameters for the j -th component of the mixture. The finite mixture of multiple negative binomial allows for greater flexibility since we can have overdispersion within each components and hence to get a more parsimonious model: for data that

we could need a lot of Poisson components to catch the overdispersion, few negative binomials can do the job.

Finally, note that since the Poisson distribution is a limiting case, it is possible that if the parameter r_j for some components tend to infinity we get Poisson components and hence the final model is a mixture with some negative binomial and some Poisson components.

3. ML estimation via an EM algorithm

3.1. The Algorithm

In this section we describe briefly an EM algorithm for ML estimation. Based on the finite mixture representation, standard results of applying the EM algorithm for finite mixtures can be used.

To start the description of the EM algorithm, consider the simple case when a random sample of size n from a P_m distribution is available. This implies no mixture or simply that $k = 1$.

Then for the multiple Poisson distribution it is easy to show that the ML estimate of λ is simply

$$\hat{\lambda} = \frac{\sum_{i=1}^n x_i}{mn}. \quad (1)$$

Note that if the data come from the P_m then the integer division will be valid in all cases. This will be used in the sequel in the EM algorithm for the finite mixture case at the M-step.

We come back to the finite mixture of multiple Poisson distributions case. Consider the case of k components, with multiplicities m_j , parameters λ_j and mixing proportions π_j , $j = 1, \dots, k$. Note that m_j 's are known, since they were set beforehand, and thus we do not need to estimate them.

Using the standard approach for finite mixtures we assume the unobserved latent variables Z_{ij} which are the component indicator i.e. $Z_{ij} = 1$ if the i -th observation belongs to the j -th group and 0 otherwise. Hence $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ik})$. Note that for our model and an specific multiplicity some observations clearly have $Z_{ij} = 0$ by definition. For example for multiplicity equal to 2 every odd observation cannot be generated from this component. Then the steps of the algorithm are the following:

At the E-step, using the current parameter estimates, calculate

$$w_{ij} = E(Z_{ij}) = \frac{\pi_j P_{m_j}(y_i | \lambda_j)}{P(y_i)}$$

and then update the parameters at the M-step as

$$\hat{\pi}_j = \frac{\sum_{i=1}^n w_{ij}}{n} \quad \text{and} \quad \hat{\lambda}_j = \frac{\sum_{i=1}^n w_{ij} x_i}{m_j \sum_{i=1}^n w_{ij}}.$$

It is easy to verify that the M-step for the λ 's is simply a weighted version of (1). Therefore for this simple case the algorithm is so simple and has only closed form expressions.

All the advantage and disadvantages of the EM for finite mixtures apply here too. Of course typical cautions about EM apply. We need good initial values in order to avoid problems of local maxima. Simple initial values can be derived for considering data from the possible multiplicities and derived from the initial values for λ 's but also for the mixing proportions. For example, for a multiplicity $m = 10$ one can use only the data with value $10x$ and derive from them some estimate for the mean of their distribution. In all our extensive examination of the algorithm such an approach led to quick convergence of the algorithm to the global maximum.

Finally, note that the derivation of the finite mixture and the algorithm does not exclude the case of the same multiplicity for some components. In this case derivation of starting values is equivalent to that of finite Poisson mixtures.

Switching to the case of finite mixtures of multiple negative binomial, the EM algorithm is similar but the main difference is that since negative binomial does not have a closed form ML estimator, the M-step is not in closed form. In fact in the M-step one has to maximize a weighted likelihood, namely to maximize for each components $j = 1, \dots, k$ the weighted log-likelihood function

$$L_j(\mu_j, r_j | \mathbf{y}) = \sum_{i=1}^n w_{ij} \log [G_{m_j}(y_i | \mu_j, r_j)].$$

Now, no closed form expressions are available but the problem is still a rather simple one. For example this is easy with standard software. In all our cases we used `optim` function in R. A final note here applies. When the overdispersion parameter of the negative binomial tends to infinity, the negative binomial becomes a Poisson distribution. In this case, the likelihood

is flat and hence the algorithm may have problems to estimate this parameter, since large differences in its value result to very small changes to the likelihood. To avoid numerical problems one may use an upper limit for this parameter.

3.2. Simulation study

In order to investigate the behavior of the proposed algorithm we conducted some simulation experiments. We used 4 configurations as described below, and 3 sample sizes $n = 200, 500, 1000$. Multiplicities are considered known. The configurations are the following: case A, $m = (1, 2, 5, 7)$, $\lambda = (1, 2, 0.5, 0.5)$ and $\pi = (0.25, 0.25, 0.25, 0.25)$. This seems to be an easy case: the multiplicities are well separated. Case B assumes $m = (1, 2, 4, 8)$, $\lambda = (1, 2, 0.5, 0.5)$ and $\pi = (0.25, 0.25, 0.25, 0.25)$. So now the multiplicities are multiples of 2 and for example all the values of 4 can be obtained from 2 as well, while values from multiplicity $m = 8$ can be also obtained from multiplicity $m = 4$. In this case, we want to examine if the algorithm may be trapped. Case C is similar to B but now some of the components have small mixing probability, namely $m = (1, 2, 4, 8)$, $\lambda = (1, 2, 0.5, 0.5)$ and $\pi = (0.25, 0.5, 0.05, 0.2)$. The aim is to see whether such small components can be recovered. Finally, case D has $m = (1, 2, 4, 8)$, $\lambda = (1, 2, 0.3, 0.2)$ and $\pi = (0.75, 0.15, 0.05, 0.05)$, which implies multiplicities which are related and small components, but now we include small λ 's for the small components. This implies that the peaks related to the smaller multiplicities are small. Figure 1 shows the probability mass functions for all cases.

The EM algorithm was used in all simulations. 1000 replications were run in each experiment. Random initial values were chosen for the λ 's and the π 's. The convergence of the EM was achieved very fast for all cases.

Boxplots in Figures 2 present the findings over the 1000 replications. For all cases the estimated parameters are centered at the true values and, as expected, the variability decreases with the sample size. Even for small sample size $n = 200$ the estimates behave very well. Even for cases C and D, with small components and small lambda's, the algorithm easily identified the correct structure. However for small sample sizes we can see small deviations in such cases since very few observations from these components existed in the data. So, the simulations support that the performance of the algorithm is good and the underlying parameter values are adequately located.

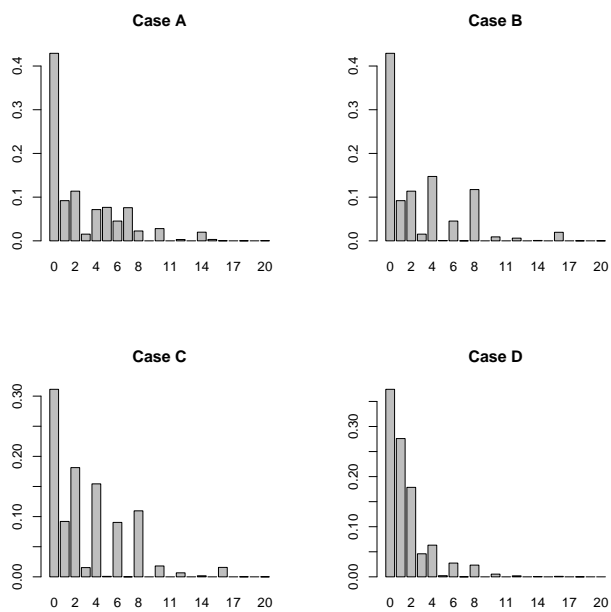
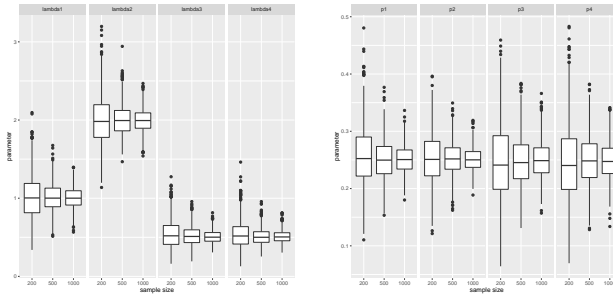
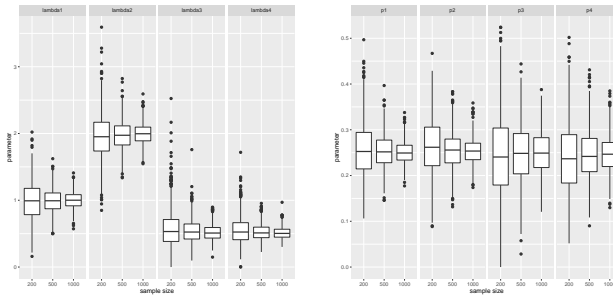


Figure 1: The probability mass functions for all cases

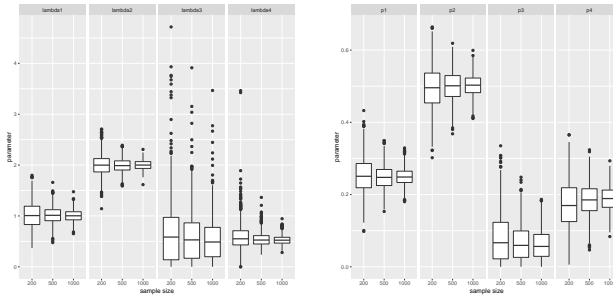
Case A



Case B



Case C



Case D

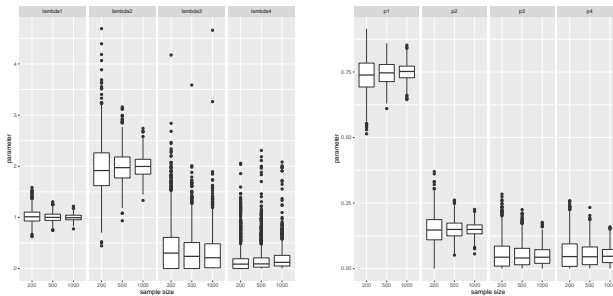


Figure 2: Boxplots with estimated parameters for all cases

3.3. The case of unknown multiplicities

So far we have assumed that the multiplicities are known. This sounds a reasonable approach since for certain phenomena the researcher might have sufficient information about the peaks and the heaped values and hence this knowledge can lead to a reasonable choice of values for the multiplicities. Also in practice a simple look on the data can easily reveals candidate values for the multiplicities. However, it is interesting to discuss the general case when the multiplicities are considered unknown and hence they must be estimated as well.

First of all, to avoid identifiability problems, multiplicities are assumed to be ordered in the components as otherwise interchanging components may lead to the same distribution. To allow for generality, we may assume that there is no knowledge at all about their values. We do not allow the λ to be zero since this implies a points mass at zero regardless the multiplicity and hence leads to identifiability problems.

For such conditions, our EM algorithm can be used to determine the unknown multiplicities: now one can consider a range of multiplicities starting from 1 up to the highest value at the data. The EM algorithm described can be used. Typically, multiplicities that are not useful will be estimated with mixing proportion tending to zero.

To exploit this strategy we run a small simulation experiment. We generated 100 samples of size $n = 1500$ from a mixture with multiplicities at 1, 2, 4, 8, with λ 's equal to (5, 2, 0.55, 0.25) and mixing proportions (0.45, 0.35, 0.15, 0.05). Then we fitted the model assuming multiplicities with values, 1, \dots , 15 but also with the correct multiplicities, 1, 2, 4, 8. Figure 3 shows at the left panel the estimated mixing proportions and λ 's for the 15 multiplicities and the right panel with only the correct 4 multiplicities. From the left panel one can see that the algorithm more or less discards the incorrect multiplicities. Note that since the algorithm stopped when the relative change of the log-likelihood was smaller than 10^{-12} , some mixing proportions are too close to zero and hence implying discarding this multiplicity. Finally, focusing on the difference between the model with 4 multiplicities and the one with 15 multiplicities, the model with 4 multiplicities was always more parsimonious and AIC selected easily this model. Also starting with the model with 15 multiplicities and removing multiplicities with a step AIC approach (i.e. removing the multiplicity that does not improve the AIC every time) ended always to the correct model with 4 multiplicities.

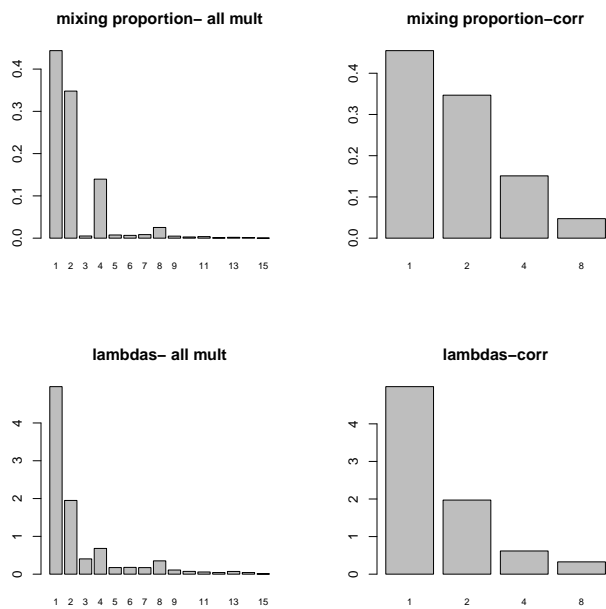


Figure 3: Averages over the 100 replications for the case when the multiplicities are considered unknown (left panel) and when we know them (right panel)

4. Numerical application

4.1. The data

The data refer to the duration of inability to work or the number of work disability days. The database was provided by one of the biggest Spanish motor insurers. We make use of 19,960 observations from non-fatal victims involved in traffic collisions in Spain. Information of the number of work disability days reported by each victim was recorded by the insurer during the processing of claims to track them until settlement. In particular, the number of work disability days is used to evaluate the temporary disability compensation. The data set has been previously used for other purposes in Boucher and Santolino (2010), Santolino et al. (2012) and Ayuso et al. (2016).

The most interesting feature of the data is that there are a lot of spikes in certain values. The data are depicted in Figure 4. A close look on these values shows, for instance, spikes in multiples of 5, 7 and 30, implying perhaps the different time scale used by doctors when deciding on the number of days of sick leave for workers. For example, it is reasonable that a doctor advises re-examination of a patient after 30 days (one month) rather than 29 or 31 days. This creates the spikes in the frequency distribution which we aim at modelling with the models described in section 2.2.

From the plot one can see significant spikes at certain values that apparently coincide with the different time scales used by doctors according to the severity of injuries: 5 days (a working week), 7 days (a calendar week), 10 days, 15 days (a fortnight), 20 days, 30 days (a month) and 90 days (a trimester). Also the variance of the observed values is very large. The variance equals 3,160 and the mean equals 68.22. This implies that even for the simple Poisson case (i.e. if $m = 1$) we need more components to fit this part. The same is perhaps valid for other values of m , where more than one component would be needed (adding components until the increase in the log-likelihood starts to fall away). Also note that all components contribute to the zero cell which is very large as it can be seen in Figure 4.

4.2. Fitting a finite mixture of multiple Poisson distributions

In order to model this data set we assumed as a starting model the finite mixture of multiple Poisson distributions. Later on we also switch to a finite mixture of multiple negative binomial distributions as described in section 2.2.

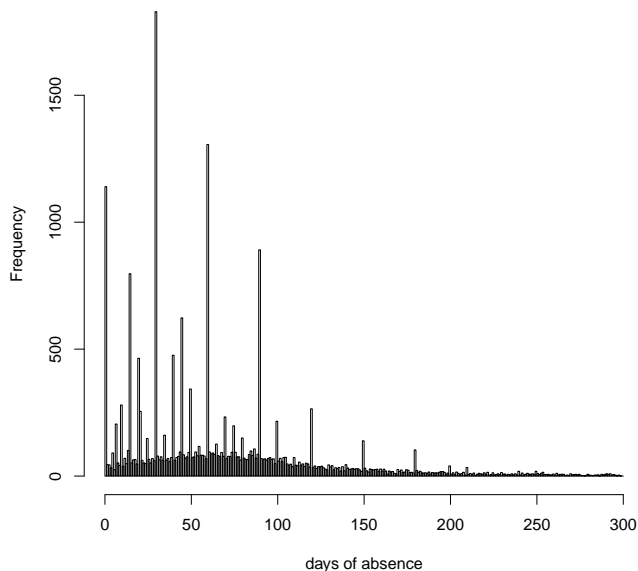


Figure 4: The observed frequencies for the data. Number of work disability days.

For all estimations we used the EM algorithm described in section 3. The stopping criterion was to stop when the relative change in the log-likelihood between two successive iterations was smaller than 10^{-8} . As initial values we used an incremental approach, where each new component was given probability 0.05 (decreasing the probability of the existing components), with mean at the center (i.e. that we would obtain if only this component was used). Based on the spikes that one can see in Figure 4 and the different time scales used by doctors, we assume that the multiplicities are 1, 5, 7, 10, 15, 20, 30 and 90. Alternatively, we might have considered the multiplicities unknown and proceed according to section 3.3. Finally, decisions on whether to add a new component for each multiplicity was determined by adding components until the AIC criterion did not support this addition.

Following the above considerations, for the finite mixture of multiple Poisson model, we used 19 components for this data set. Table 1 reports the estimated components, and their bootstrap standard errors, for this model. Note that since the sample size is large the uncertainty is small.

Table 1: The fitted components with their estimated parameters and bootstrap standard errors

m	λ	std.err.	π	std.err.	Log-lik if removed
1	2.8476	0.3235	0.0090	0.0012	
1	9.9614	0.5581	0.0203	0.0022	
1	20.8443	0.6438	0.0394	0.0032	
1	35.8674	0.9688	0.0540	0.0040	
1	52.7750	1.0976	0.0767	0.0048	
1	72.0698	1.0248	0.0887	0.0048	
1	96.3278	0.7238	0.0947	0.0043	
1	126.8449	0.9223	0.0556	0.0025	
1	161.7984	0.9718	0.0394	0.0021	
1	204.1400	1.0489	0.0268	0.0016	
1	260.6124	1.0334	0.0212	0.0011	
5	4.0235	0.1578	0.0352	0.0041	-91,372.29
5	16.3175	0.7820	0.0177	0.0036	
7	2.1336	0.1458	0.0262	0.0020	-91,460.22
10	4.0051	0.1093	0.0888	0.0043	-91,398.46
15	2.1845	0.0582	0.1302	0.0048	-91,781.10
20	5.6823	0.4302	0.0119	0.0023	-91,292.23
30	2.0783	0.0425	0.1318	0.0057	-91,674.76
90	0.4533	0.0458	0.0324	0.0037	-91,325.08

The first 11 components had multiplicity $m = 1$ and well separated components that try to fit the general picture. This perhaps implies that this component can be replaced from an overdispersed distribution like the negative binomial rather than many simple Poisson ones. The rest of components fits the spikes present in the data that refer to the different time scale implied.

We also report at the last column of Table 1 the log-likelihood when the specific multiplicity was removed as an indication how important is this multiplicity. The log-likelihood of the full model is -91,265.74. One can see some interesting features. For all values of m , if we remove this component the log-likelihood decreases a lot, implying that the existence of this multiplicity is very important for the model. Only for $m = 5$ we need two components to fit this part, implying again perhaps some overdispersion. While we have tried with the other multiplicities, the results coincide with just using one component for each multiplicity.

To have an idea on the goodness of fit we have calculated a chi-square

goodness of fit statistic. Cells with expected frequency lower than 5 were removed from the calculations. In total 275 out of 300 cells were used, and the sum of the expected frequencies not used was only 91 (out of 19960) which is considered as rather unimportant (less than 0.5%). The chi-square test statistic was 27.63 which implies a very good fit for 237 degrees of freedom. Figure 5 shows the difference between observed and expected frequencies. One can see as checked with χ^2 that the fit is very good in a great extend. Such a plot can be useful to identify potential missing multiplicities, i.e. if we see that the model fails to fit certain points which is a multiple of some number this would imply perhaps that we need to add such a component. For example, using a component with multiplicity $m = 15$ contains also part of the information from a component with $m = 30$. Hence such a plot may reveal which one is preferable by looking at which values we have failed to fit in a satisfactory way.

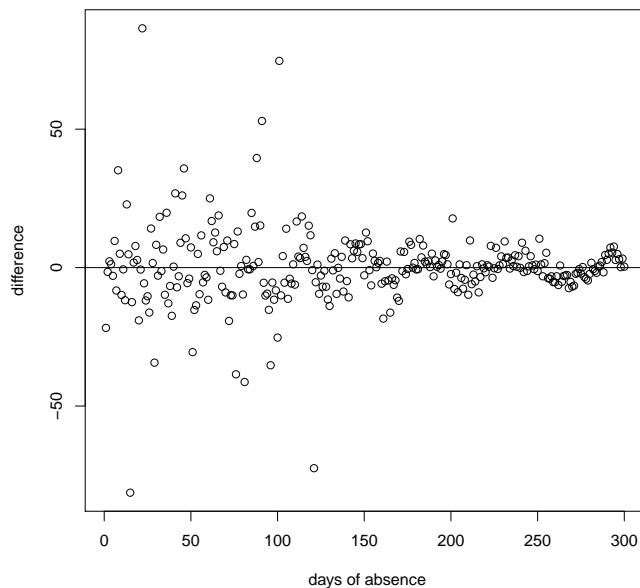


Figure 5: Differences between observed and expected frequencies from the model.

To further show the goodness of the model, in Figure 6, we have plotted the empirical cumulative distribution function with the one implied by the

fitted model. The fit is as already discussed quite close. Looking on the cdf, which does not have the rather unsmooth picture of the probability function, one can see that the model, if used for prediction of say future number of absence days, can be a good model since the fit is quite close.

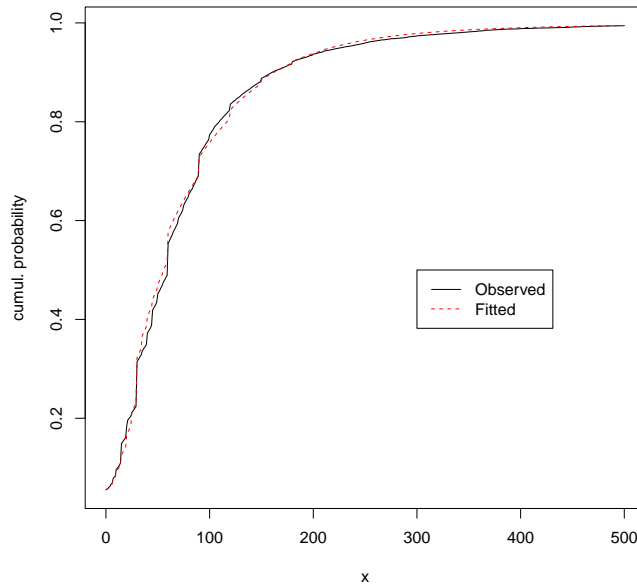


Figure 6: Observed versus fitted distribution function.

4.3. *Fitting a finite mixture of multiple negative binomial distributions*

The big number of components needed for $m = 1$ was clearly an indication for large overdispersion, with respect a Poisson model, at this multiplicity. A standard approach to model overdispersion is to consider a negative binomial model. Hence we fitted finite mixtures with multiple negative binomial distributions. Two major comments apply here.

First, in terms of the log-likelihood is well known that the maximum possible likelihood can be obtained with a finite number of simple Poisson components rather than negative binomial or their mixture. See for example the relevant results on the non-parametric maximum likelihood given in Lind-

say (1995). Hence, fitting mixtures of negative binomial do not necessarily improve the log-likelihood.

Second, since for certain parameters of the negative binomial we reach a Poisson distribution, the finite mixture of multiple negative binomial cannot be worse than that of multiple Poisson components.

Our result showed the interesting issue that for the larger multiplicities, the size parameter of the negative binomial took large values implying Poisson distributions. For $m = 1$ and $m = 5$ where the finite mixture of multiple Poisson needed more than one component the negative binomial fitted this part of the data with fewer components, three for $m = 1$ and one for $m = 5$. Hence, the finite mixture of multiple negative binomial needed less parameters but failed to achieve a better likelihood. Table 2 shows a formal comparison. In this table we use three different models: the finite mixture of multiple Poisson, the finite mixture of multiple negative binomial and, finally, a model between the others (assuming Poisson distributions for components with multiplicity $m > 5$ and negative binomial distributions for $m = 1, 5$). As commented above, the last model gives the same likelihood as the negative binomial (but saving some parameters which we do not need to estimate), and the finite mixture of multiple Poisson outperforms the other models implying better fit even if we need to have more parameters.

Table 2: Comparison between the models.

Model	Log-lik	parameters	BIC
Poisson	-91265.74	37	182897.83
NB	-91408.73	29	183104.60
NB- Poisson	-91408.73	23	183045.19

5. Concluding Remarks

In this paper we developed new models to fit count data with a lot of spikes in certain values. Although linked to digit preference or heaping literature, our approach is very different. The literature in digit preference assumes that data can be interpreted as indirect (or rounded) observations of a latent distribution and hence the usual way to address heaped data is

to define a model for the latent exact values by smoothing response heap-
ing. In contrast, we are interested in modelling the data itself, not the latent
distribution.

For this purpose, based on finite mixtures of multiple discrete distribution
of different multiplicities, we propose count models that fit the data with
periodic peaks suitably. In particular, we propose a finite mixture of multiple
Poisson model and a finite mixture of multiple negative binomial model. An
EM algorithm is derived to fit the models to a real data sets. A simulation
study was conducted to analyze the behavior of this algorithm concluding
that the EM algorithm works fine.

In order to show the applicability of the proposed models, we use them to
fit a real data set referred to the number of work disability days after a motor
accident occurs. A specific feature of such data is the presence of spikes in
certain number of days perhaps due to different calendar scales that are used
by medical doctors depending on the severity of the injury. A doctor may
think on a daily or weekly or monthly scale and hence it is more probable to
say 7 days or 30 days of sick leave than 6 or 29 days.

After fitting the finite mixture of multiple Poisson model to this data set,
the main finding is that the model provides a very good fit when working
week, calendar week, fortnight, month, and other multiplicities are taking
into account. For multiplicities equal to one and five we need more than one
component to fit this part, implying perhaps some overdispersion. To account
for this variability, a finite mixture of multiple negative binomial model was
also fitted to this data set obtaining a model with fewer components for
multiplicities equal to one and five and dispersion parameter of the negative
binomial very large, implying Poisson distributions, for larger multiplicities.
However, although more parameters are needed, the finite mixture of multiple
Poisson model leads to a better fit in terms of BIC than the negative binomial
one.

Finally, we would like to mention a way in which this paper might be ex-
tended. Extension to the case when covariates are present is straightforward
but more cumbersome. Including covariates may help to reduce the number
of components (or multiplicities) that are necessary to fit the data.

Acknowledgements. Research for this paper was initiated while the sec-
ond author was visiting the Riskcenter Research Group at the University
of Barcelona. The authors wish to acknowledge the support of the Spanish

Ministry for grant ECO2012-35584.

The authors gratefully acknowledge the constructive comments made by two anonymous referees, which helped to improve the quality of the paper.

References

- Ayuso, M., Bermúdez, L., Santolino, M., 2016. Copula-based regression modeling of bivariate disability severity of temporary and permanent motor injuries. *Accident Analysis & Prevention* 89, 142–150.
- Boucher, J.-P., Santolino, M., 2010. Discrete distributions when modeling the disability severity score of motor victims. *Accident Analysis & Prevention* 42 (6), 2041–2049.
- Camarda, C. G., Eilers, P. H., Gampe, J., 2008. Modelling general patterns of digit preference. *Statistical Modelling* 8 (4), 385–401.
- Channon, A. A., Padmadas, S. S., McDonald, J. W., 2011. Measuring birth weight in developing countries: does the method of reporting in retrospective surveys matter? *Maternal and Child Health Journal* 15 (1), 12–18.
- Jorgensen, B., 1997. *The Theory of Dispersion Models*. Chapman & Hall, London.
- Lindsay, B. G., 1995. Mixture models: theory, geometry and applications. In: NSF-CBMS Regional Conference Series in Probability and Statistics.
- Nietert, P. J., Wessell, A. M., Feifer, C., Ornstein, S. M., 2006. Effect of terminal digit preference on blood pressure measurement and treatment in primary care. *American Journal of Hypertension* 19 (2), 147–152.
- Patel, Y., 1976. Estimation of the parameters of the triple and quadruple stuttering-poisson distributions. *Technometrics* 18 (1), 67–73.
- Santolino, M., Bolancé, C., Alcañiz, M., 2012. Factors affecting hospital admission and recovery stay duration of in-patient motor victims in Spain. *Accident Analysis & Prevention* 49, 512–519.
- Siegel, J. S., Swanson, D. A., 2004. *The methods and materials of demography*. Elsevier Academic Press, Amsterdam.

- Torelli, N., Trivellato, U., 1993. Modelling inaccuracies in job-search duration data. *Journal of Econometrics* 59 (1), 187–211.
- Vinogradov, V., 2004. On the power-variance family of probability distributions. *Communications in Statistics* 33 (5), 1007–1029.
- Wang, B., Wertelecki, W., 2013. Density estimation for data with rounding errors. *Computational Statistics & Data Analysis* 65, 4–12.
- Wang, H., Heitjan, D. F., 2008. Modeling heaping in self-reported cigarette counts. *Statistics in medicine* 27 (19), 3789–3804.
- Wang, H., Shiffman, S., Griffith, S. D., Heitjan, D. F., 2012. Truth and memory: Linking instantaneous and retrospective self-reported cigarette consumption. *The Annals of Applied Statistics* 6 (4), 1689–1706.
- Wolff, J., Augustin, T., 2003. Heaping and its consequences for duration analysis. *Allgemeines Statistisches Archiv* 87, 1–28.