Chapter 32

Zero-Inflated Count Models and their Applications in Public Health and Social Science

Dankmar Böhning, Ekkehart Dietz and Peter Schlattmann

Department of Epidemiology, Institute for Social Medicine, Free University Berlin

1. Introduction

In many areas of interest including public health, epidemiology, sociology, psychology, engineering, agriculture and others, the analysis of count data is of primary interest. Typically, a Poisson model $p(X = x) = Po(x,\mu)$ is assumed for modelling the distribution of the count observation X or, at least, approximating its distribution. However, it has been observed in various applications that the dispersion of the Poisson model underestimates the observed dispersion. This phenomenon, also called *overdispersion*, occurs because a single Poisson parameter μ is often insufficient to describe the population. In fact, in many cases it can be suspected that *population heterogeneity* which has not been accounted for is causing this overdispersion. This population heterogeneity is unobserved, in other words, the population consists of several subpopulations, in this case of the Poisson type, but the subpopulation membership is *not* observed in the sample. One possibility to cope with the problem is to assume that the heterogeneity involved in the data can be adequately described by some density $\Pi(\mu)$ defined on the population of possible Poisson parameters μ. Since this heterogeneity cannot be observed directly, it is also called *latent*. We can only observe counts coming from the *marginal* or *mixture* density

$$\int_0^\infty Po(x,\mu)\,\Pi(\mu)\,(d\mu). \qquad (1)$$

Two approaches can be distinguished. One, the traditional approach, is to follow a fully *parametric* model for the mixing density Π. An example of this nature is the Gamma distribution for Π, for which the marginal density becomes the negative binomial. The second, *nonparametric* approach does not specify any parametric density for Π. Here, the nonparametric maximum likelihood estimator (NPMLE) is always finite giving weights π_j to the *latent* classes or subpopulations μ_g, $g = 1,\ldots,G$ (Simar, 1974; Böhning 1982, 1995; Lindsay, 1983). This nonparametric approach is attractive, since it is not only easy to interpret but also requires *no* specification of the number of latent classes G. Both approaches are connected with the *empirical Bayes* methodology (Maritz and Lwin, 1989), since an estimate of the distribution $\Pi(\mu)$ can be viewed as an *empirical Bayes* estimator, as an estimator for the prior distribution in the Bayes theorem. Thus, mixture models provide the tool to classify observations via the maximum posterior probability into the components or classes of the mixture model.

In this contribution we study a special form of nonparametric heterogeneity density Π, viz. a two mass distribution giving mass $(1-\pi)$ to count 0 and mass π to the second class with

mean μ. In other words, we consider a data situation in which a number of extra-zeros occur. There are a variety of applications in which extra-zeros occur. We will mention some of those in the sequel.

2. The DMF - Index in Dental Epidemiology

In dental epidemiology the DMF-index is an important and well-known indicator and overall measure for the dental status of a person. It is a count number standing for the number of DECAY, MISSING, and FILLED Teeth (in which case it is called DMFT-Index) or Tooth-Surfaces (in which case it is called DMFS-Index). As an application, we consider here data coming from a prospective study of school children from an urban area of Belo Horizonte (Brazil). The interested reader is pointed to Mendonça and Böhning (1994) and Mendonça (1995). Figure 1 shows the DMFT distribution at the beginning of the study. There is a clear spike of extra zeros representing the caries-free children. If one considers the distribution in general, the large number of children with zero-value DMFT-index is remarkable and seems a phenomenon not untypical for DMFT-distributions. Nevertheless, the line of argument followed in dental epidemiology uses the fact that the DMFT-index is a count variable, and argues that typically Poisson distributions are used for count data, finally leading to log-linear modelling to include covariates. However, the Poisson distribution does not fit well at all, in this case. If the Poisson assumption would be true, expected value and variance should coincide.

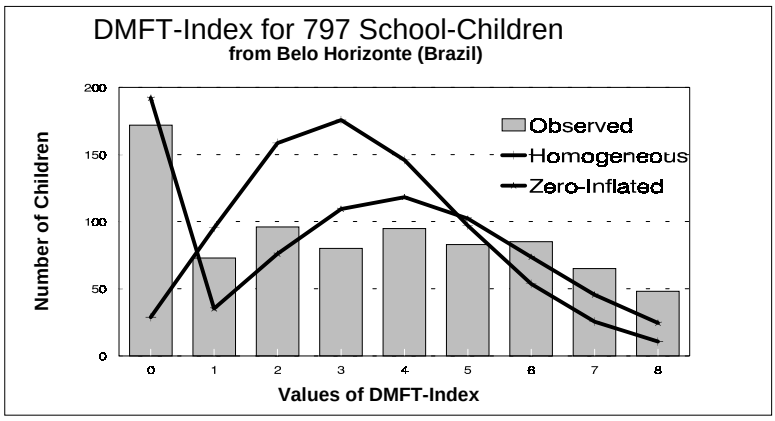

Figure 1: DMFT distribution at begin of study

Since these can be estimated by the sample mean and sample variance, it is natural to compare $s^2 = \{(x_1 - \bar{x})^2 + \ldots + (x_R - \bar{x})^2\}/(n-1)$ and $\bar{x}$, leading to the overdispersion test

$$O = \sqrt{\frac{n-1}{2}}\left(s^2 - \bar{x}\right)\Big/\bar{x} \tag{2}$$

which takes on the value 21.06 ($s^2 = 6.639, \bar{x} = 3.23$, $P - value < 0.0001$). Here the fact is used that O is approximately normal, but also equivalent to $(n-1)s^2/\bar{x}$ which is χ^2 with $(n-1)$ degrees of freedom (Böhning, 1994).

Coming back to the caries prevention study, a second measure is of interest which describes the effect of the *prevention* measure. After two years the children were measured a

second time and their dental status evaluated, including the computation of the DMFT-index. This leads to the effect measure ΔDMFT = DMFT1-DMFT2, where DMFT1 denotes the dental status at the beginning of the study and DMFT2 the dental status two years later. Only if ΔDMFT > 0 has there been an improvement of the child's dental status, otherwise it has not changed or become worse. In the BELCAP study, typical only for a few children, the dental status changed negatively. This might be completely different in other studies or in studies on older populations in which only a negative development can be expected which then would justify a definition of Δ´DMFT = DMFT2-DMFT1.

Figure 2 shows the situation for ΔDMFT, clearly pointing out that a simple Poisson would not give an adequate fit to the data. This can also be seen by the overdispersion test mentioned above $o = 21.65\,(s^2 = 3.42, \bar{x} = 1.64, P - value < 0.0001)$, indicating strong overdispersion , which will, as we will see in the following section, be explained by a quite simple model.

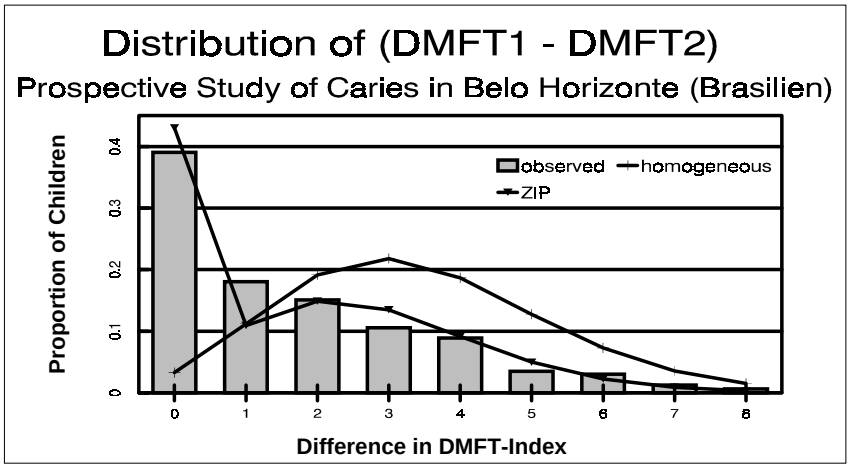

Figure 2: ΔDMFT distribution

3. The Zero-Inflated Poisson Model

A simple and frequently applied statistical model for a count distribution is the Poisson model in which we assume that X follows a Poisson density

$$p(X = x) = e^{-\mu}\,\mu^x / x! = Po(x, \mu). \tag{3}$$

As we have seen in section 2, model (3) does not fit the DMFT-data. Instead, the data show strong overdispersion $s^2 > \bar{x}$, due to the fact that a large frequency of extra-zeros have occurred in the distribution. A simple way to model this *zero-inflation* is to include a proportion $(1-\pi)$ of extra-zeros and a proportion $\pi \times e^{-\mu}$ coming from the Poisson distribution (Johnson, Kotz, and Kemp, 1992, p. 314; Lambert, 1992). We can write this *zero-inflated* Poisson density f as

$$f(x; \pi, \mu) = \begin{cases} (1-\pi) + \pi\, e^{-\mu}, & \text{if } x = 0 \\ \pi\, Po(x, \mu), & \text{if } x > 0 \end{cases} \tag{4}$$

or

$$f(x;\pi,\mu) = (1-\pi)\,\mathrm{Po}(x,0) + \pi\,\mathrm{Po}(x,\mu) \tag{5}$$

since it is one of the nice properties of the Poisson distribution that $\mathrm{Po}(x,0) = 0$ for all $x > 0$, and $\mathrm{Po}(0,0) = 1$. Thus, $\mathrm{Po}(x,0)$ is the one point distribution putting all its mass at zero. As a sideline, we note that this property is not shared by many distributions. For example, the normal does not have this property, whereas the binomial does.

Interpretation. The representation (5) points out that the ZIP-model is a special mixture model having two classes, where the first class has a *fixed* value at 0. This class can be interpreted according to the type of application, and usually rather simple interpretations exist. In the case of the DMFT-index, this class consists of children with *no* caries at all. In the case of the ΔDMFT, this zero-class corresponds to those children showing no improvement.

For the ZIP-model with zero-inflation we find

$$\mathrm{Var}(X) = \mathrm{E}(X) + \mathrm{E}(X)\left(\mu - \mathrm{E}(X)\right) \quad \text{and} \quad \mathrm{E}(X) = \pi\mu\,. \tag{6}$$

To demonstrate these statements we return to the biometric application described in section 2. For the ΔDMFT-data of Figure 2 we find an overdispersion $s^2 - \bar{x} = 1.77$. The maximum likelihood estimators for the ZIP-model turn out to be $\hat{\pi} = 0.67$ and $\hat{\mu} = 2.48$, leading to a fitted overdispersion (under (6)) $\mathrm{E}\hat{X}(\hat{\mu} - \mathrm{E}\hat{X}) = 1.36$, leading to an explained overdispersion of $\mathrm{E}\hat{X}(\hat{\mu} - \mathrm{E}\hat{X}) / (s^2 - \bar{x}) = 0.77$, with $\mathrm{E}\hat{X} = \hat{\mu}\hat{\pi}$. Thus 77% of the overdispersion would be *explained* by the ZIP-model. In the Appendix moment and maximum likelihood estimators are discussed for the ZIP-model. In the sequel, let $\hat{\mu}$ and $\hat{\pi}$ denote the maximum likelihood estimates of μ and π, respectively.

4. Further Examples

Traffic Accident Research. Kuan et al. (1991) discuss data coming from the California Department of Motor Vehicles master driver license file. Here the variable of interest is the number of accidents per driver. A possible motivation can be seen in the possibility of finding risk factors involved in the accidents.

Some statistics: $\bar{x} = 0.2032,\ s^2 = 0.2365$, the overdispersion test delivers $O = \sqrt{\frac{n-1}{2}}$ $(s^2 - \bar{x}) / \bar{x} = 8.5319$ $(\mathrm{P-value} < 0.0001)$, indicating a strong overdispersion. Figure 3 shows the distribution, Figure 4 the Pearson residuals for the simple and ZI-Poisson model. As can be seen, the ZIP model leads to an adequate fit.

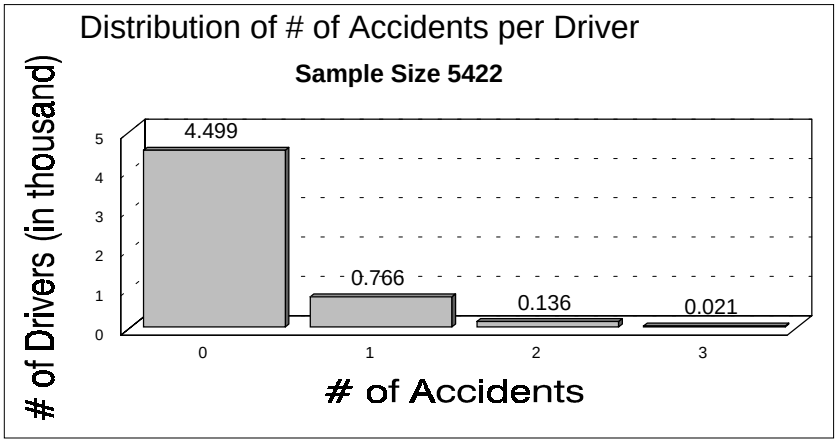

Figure 3: Distribution of **n**umber of accidents

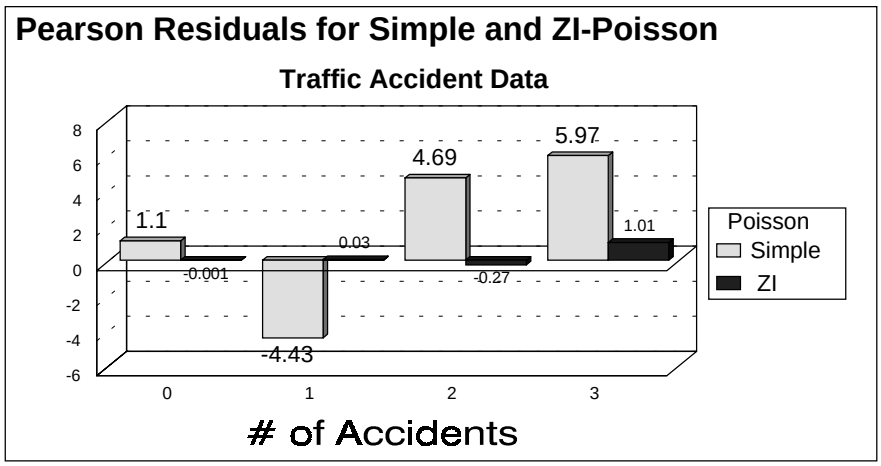

Figure 4: Pearson residuals of the two models

Crime Sociology. In a study on deviating behaviour, Dieckmann (1981) provides an analysis of a data set on 4039 persons with criminal behaviour. The variable of interest is here the number of criminal acts per person. The motivation for such a study might be to find factors leading to deviating behaviour. Figure 5 shows the corresponding distribution. The associated statistics are $\bar{x} = 0.0776, s^2 = 0.1209$ with a highly significant overdispersion test (P–value < 0.0001). Again, the ZIP-model provides an acceptable fit in this case.

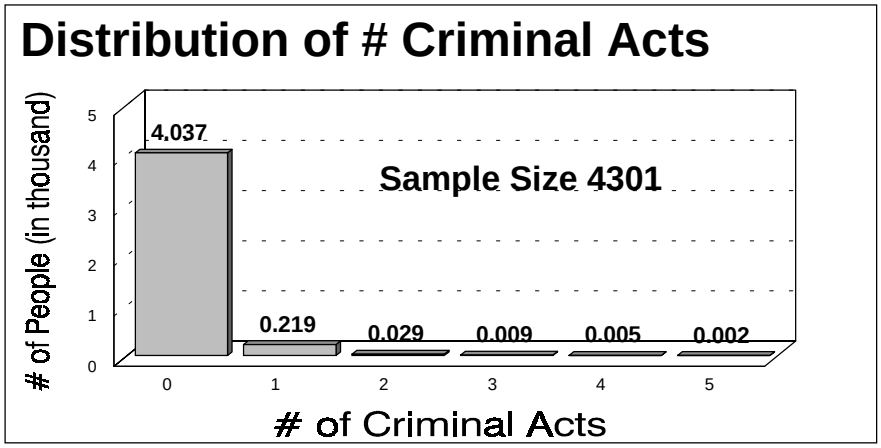

Figure 5: Distribution of number of criminal acts

Geographic Epidemiology (Disease mapping) Disease mapping is a common method of displaying the geographical (spatial) of disease occurrence. In infectious epidemiology its main value is not only the identification of areas of disease but also the mode and transmission. In non-infectious epidemiology, disease maps are used to detect areas of high risk, which then in turn can be used to detect unknown risk factors by considering their associated spatial maps. As an example we study the distribution of the Sudden Infant Death Syndrome - Rate for North-Carolina (USA). Symons et al. (1983) provide an analysis of this data set demonstrating a strong spatial gradient in the SIDS-occurrence. The motivation behind these studies can be seen in finding risk factors (not yet known) involved in SIDS which then can be further investigated in ecological studies (Schlattmann and Böhning, 1993). The analysis is further refined in Böhning et al. (1992) showing a clear spike at those counties with zero death cases. We mention this example here because in some cases the ZIP-model needs a special form, when the data are given as *rates*: x_i sudden infant deaths in county i with N_i at risk:

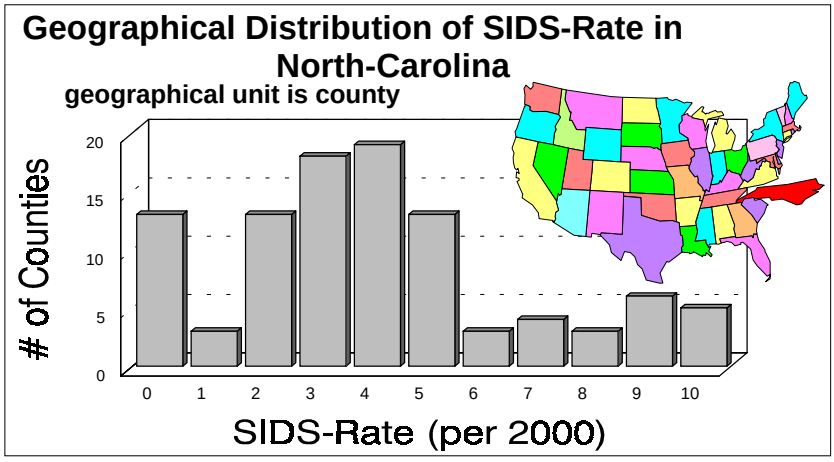

Figure 6: Geographical distribution of SIDS-rate in north-carolina

$$(1-\pi)\,\mathrm{Po}(x_i,0)+\pi\,\mathrm{Po}(x_i,\mu N_i), \tag{7}$$

where μ is the population rate, or as *ratios*, such as the Standardized Mortality Ratio: $\mathrm{SMR}=\dfrac{\text{Observed}}{\text{Expected}}=x/E$:

$$(1-\pi)\,\mathrm{Po}(x_i,0)+\pi\,\mathrm{Po}(x_i,\mu\times E_i), \tag{8}$$

where μ is the population SMR. Note that E_i denotes here the expected number of death cases in the i-th county which are computed from an external reference population (often the national standard population or world population).

5. A Graphical Representation of the ZIP-Model

In this section we provide a graphical display which presents a summary information of the various parts of the ZIP-model. Recall that the ZIP-model is a mixture model with two components, one fixed at zero with weight $(1-\pi)$, the other putting weight π on the second component mean μ. Figure 7 contains a rectangle with baseline of length μ, so that the two end points of this baseline represent the two component means of the mixture model. The height of the rectangle is π, thus showing the *distribution of the mixing distribution*. Now the area of the rectangle is $\pi\times\mu=E(X)$! Note that for the sample replacements this equation becomes $\overline{X}=\hat{\mu}\times\hat{\pi}$, because of the estimating equations given in the Appendix .

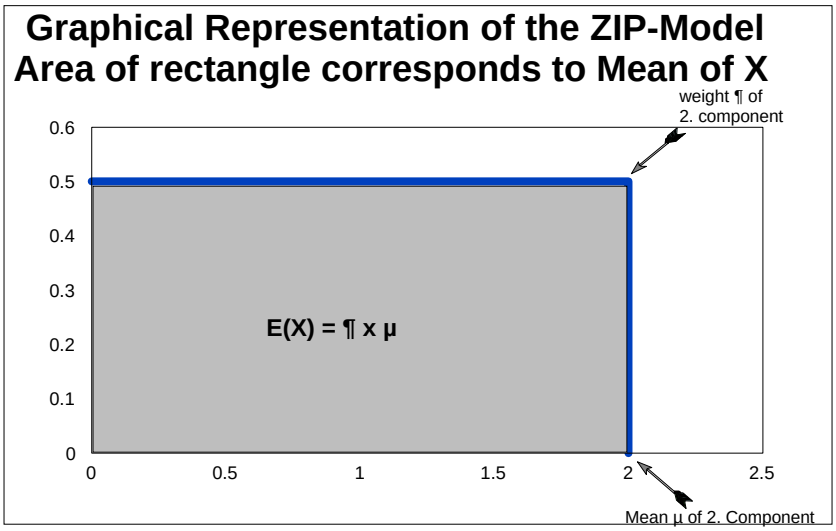

Figure 7: A graphical device for the ZIP-model

The graphical display is helpful in comparing the various parameters between groups. For example, one could define group A to be *strongly* better than group B if and only if $\pi^A\geq\pi^B$ and $\mu^A\geq\mu^B$ (with at least one inequality being strict). Graphically this means that the rectangle of group A contains the one of group B. Group A could be defined to be *weakly* better than group B if $E(X^A)>E(X^B)$, the latter being the more traditional criterion of comparison. Again, graphically this means that the area of the rectangle associated with group A is larger than the one associated with group B. Note that „strongly better“ implies „weakly better“. Let us come back to the data of the BELCAP-study of section 2. The aim

here was the evaluation of several intervention measures for the reduction of caries prevalence, measured in the ΔDMFT-variable. There were six different schools, each representing a specific prevention strategy. School 3 served as control school. As can be seen from Figure 8, School 1 is *weakly* (in the mean ΔDMFT) better than all the other schools. It is also *strongly* (in both μ and π) better than the other schools, except School 2, for which the mean of the second component is larger. This implies that in this school, there was a smaller percentage of children showing any improvement, but for those who did show improvement, it was larger than for School 1.

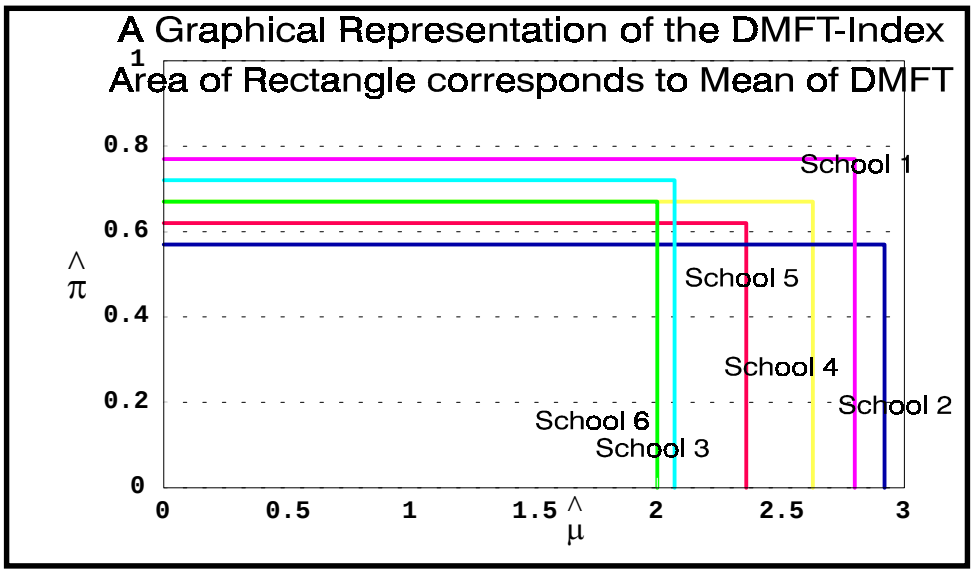

Figure 8: Six schools in the BELCAP-study in a graphical comparison

6. Including Covariates

Frequently, a variety of further variables are considered in a study either as explanatory factors or as confounders such as sex, age, and exposure covariates. The data could be arranged to form strata and strata-specific μ's and π's could be estimated as described in section 4. However, the stratified approach has its limitations when the number of covariates is increasing. Alternatively, one might try to combine the specific error structure of the ZIP-model with the framework of generalized linear models, in particular with Poisson regression. The conventional log-linear model

$$\mathrm{E}(\mathrm{X}) = \exp\left(\alpha + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}\right), \tag{9}$$

where **y** is the vector of covariates, α is an unknown intercept parameter and **β** an unknown vector of regression coefficients, can be easily generalized to the ZIP-regression, in which the distribution of our count variable X is assumed to be

$$(1-\pi)\,\mathrm{Po}(\mathrm{x},0) + \pi\,\mathrm{Po}(\mathrm{x},\mu) = (1-\pi)\,\mathrm{Po}(\mathrm{x},0) + \pi\,\mathrm{Po}\left(\mathrm{x}, \exp\left(\alpha + \boldsymbol{\beta}^{\mathrm{T}}\mathbf{y}\right)\right) \tag{10}$$

conditional on the values of the covariates. In fact, we have a model which can be placed into a class of generalized linear models for heterogeneity, considered by Dietz (1992) recently.

In the contribution of Dietz (1992), methods for finding the maximum likelihood estimates via the EM algorithm are discussed.

Let us come back to the data of the BELCAP-study of section 2. The aim here was the evaluation of several intervention measures for the reduction of caries prevalence, measured in the ΔDMFT-variable. There were seven different schools, each representing a specific prevention strategy. As additional, potential confounders the following variables were considered: SEX (binary covariate), and COLOUR (ethnic group:covariate with 3 categories). In Table 1 the results are given not only for the conventional Poisson regression and its zero-inflated generalization, but the more general mixture model is also considered which allows mixing on the intercept. It becomes clear from Table 1 that the major gain in the likelihood increase is from the non-inflated to the ZIP-model.

log-linear model: $\mathrm{Po}(x, \exp(\alpha + \boldsymbol{\beta}^T \mathbf{y}))$		*zero-inflated log-linear model* $(1-\pi)\,\mathrm{Po}(x,0) + \pi\,\mathrm{Po}(x, \exp(\alpha + \boldsymbol{\beta}^T \mathbf{y}))$	*log-linear model with two-component mixing on intercept* α $(1-\pi)\,\mathrm{Po}(x, \exp(\alpha_1 + \boldsymbol{\beta}^T \mathbf{y})) + \pi\,\mathrm{Po}(x, \exp(\alpha_2 + \boldsymbol{\beta}^T \mathbf{y}))$
class size π	1.0	$\hat{\pi} = 0.67$	$\hat{\pi} = 0.54$
log-likelihood	-1528.79	-1380.79	-1369.44
R^2 (based on deviance)	0.0202	0.1878	0.2007
deviance	1729.74	1433.74	1411.04

Table 1: Poisson and ZIP regression on the ΔDMFT-index for the covariates SCHOOL, SEX, COLOUR

Table 2 provides effect estimates with asscociated standard errors and Z-values (Z-value = estimate of effect divided by its standard error). As can be seen neither SEX nor COLOUR play a relevant role in any of the three models, whereas School 1 is most relevant in all three distributional models. One point needs to be noted. The effects of School 2 and School 5 are non-significant under the bad fitting log-linear model, whereas both effects become significant in the ZIP-model. This underlines the potential of effect bias if a wrong distributional model is chosen.

log-linear model: $\mathrm{Po}(x, \exp(\alpha + \boldsymbol{\beta}^T \mathbf{y}))$			*zero-inflated Poisson model* $(1-\pi)\,\mathrm{Po}(x,0) + \pi\,\mathrm{Po}(x, \exp(\alpha + \boldsymbol{\beta}^T \mathbf{y}))$			*log-linear model with two-component mixing on intercept* α *Mixture*: $(1-\pi)\mathrm{Po}(x, \exp(\alpha_1 + \boldsymbol{\beta}^T \mathbf{y})) + \pi\,\mathrm{Po}(x, \exp(\alpha_2 + \boldsymbol{\beta}^T \mathbf{y}))$		
estimate	Z-value	parameter	estimate	Z-value	parameter	estimate	Z-val	parameter
0.07682	1.3823	SEX	0.04161	0.6281	SEX	0.05265	0.6622	SEX
0.05248	0.8537	COLOUR(2)	0.09610	1.3298	COLOUR(2)	0.08134	0.9375	COLOUR(2)
-0.06702	-0.7563	COLOUR(3)	0.01594	0.1511	COLOUR(3)	0.05140	0.4073	COLOUR(3)
0.3643	3.8768	SCHOOL 1	0.2868	2.5859	SCHOOL 1	0.35390	2.6638	SCHOOL 1
0.1105	**1.1277**	SCHOOL 2	0.2912	**2.5070**	SCHOOL 2	0.2946	**2.1115**	SCHOOL 2
-0.02659	-0.2645	SCHOOL 4	0.07693	0.6470	SCHOOL 4	0.06141	0.4300	SCHOOL 4
0.1590	**1.7107**	SCHOOL 5	0.2188	**1.9840**	SCHOOL 5	0.2825	**2.1331**	SCHOOL 5
-0.1366	-1.2860	SCHOOL 6	-0.08608	-0.6883	SCHOOL 6	-0.04073	-0.2711	SCHOOL 6

Table 2: Effect estimates with standard errors for Poisson and ZIP regression on the ΔDMFT-index for the covariates SCHOOL, SEX, COLOUR

Table 3 compares the gain in the log-likelihood for the three distributional models. Note that the log-linear model is a special case of the ZIP-model which is a special case of the 2-component mixture model. This makes the log-likelihoods comparable vertically. Horizontally, the covariates have been included additively.

	const.	*+SEX*	*+COLOUR*	*+School*
log-linear	-1546.58	-1546.01	-1545.19	-1528.78
ZIP	-1391.94	-1391.94	-1390.88	-1380.79
Mixture	-1379.90	-1379.88	-1379.32	-1369.44

Table 3: Log-likelihood for the three distributional models and the various covariates

The point of Table 3 is as follows. One could argue that the simple log-linear model is still a valid approach, since the distributional heterogeneity is explained by the covariates. This is not the case, however, as the gain in the log-likelihood in the first row is from -1546.58 to -1528.78, whereas in the first column from -1546.58 to -1391.94. This implies that the explained heterogeneity through the covariates is dramatically smaller than through the distributional ZIP-model. Note that the same argument is *no longer valid* when the ZIP-model is compared with the 2-component mixture. If in the ZIP-model the covariates are included, a gain from -1391.94 to -1380.79 can be observed, the latter value comparing favourable to -1379.90, which is the value of the 2-component mixture with no covariates included. Thus it could be argued that the residual heterogeneity in the ZIP-model is explained through the covariates.

7. Discussion

We have seen that ZIP-models are very special *mixture* or *latent class* models which can be used in a variety of applications. It has been demonstrated that ways exist which provide maximum likelihood estimators in a reliable fashion. Since ZIP models are special (Poisson-)mixtures, software for mixture modelling such as C.A.MAN (Böhning, Schlattmann, and Lindsay 1992, Böhning 1995) might be used for fitting ZIP-models. In addition, ways to include covariates have been discussed. We assume that most of the future work will be devoted to this area of covariate inclusion in zero-inflated models.

Appendix: Estimating the Model Parameters

Moment Estimation. From (6) we have the moment equations $\mathrm{E}(\mathrm{X}) = \overline{\mathrm{X}}$ and $\mathrm{S}^2 = \mathrm{E}(\mathrm{X})(1 + \mu - \mathrm{E}(\mathrm{X}))$ which are readily solved by $\hat{\mu}_{\mathrm{MO}} = \mathrm{S}^2 / \overline{\mathrm{X}} - 1 + \overline{\mathrm{X}}$ and $\hat{\pi}_{\mathrm{MO}} = \overline{\mathrm{X}} / \hat{\mu}_{\mathrm{MO}}$.

Maximum Likelihood Estimation. Let n_i be the number of i's in the sample; in particular, n_0 is the number of zeros in the sample. Then the log-likelihood function is given as

$$\mathrm{L}(\pi, \mu) = n_0 \log\left[(1-\pi) + \pi\, e^{-\mu}\right] + \sum_{x=1}^{m} n_x \log\left[\pi\, \mathrm{Po}(x, \mu)\right] \qquad (A.1)$$

and the score vector

$$\left(n_0 \frac{e^{-\mu}-1}{1-\pi+\pi e^{-\mu}} + (n-n_0)/\pi, n_0 \frac{-\pi e^{-\mu}}{1-\pi+\pi e^{-\mu}} - (n-n_0) + n\bar{x}/\mu \right)^T$$

leading to the *score equations*

$$\pi = \frac{1-n_0/n}{1-e^{-\mu}}$$

$$\mu = \bar{x}/\pi ,$$

which can be written in one equation $\mu = \bar{x} / \frac{1-n_0/n}{1-e^{-\mu}} =: G(\mu)$, see Figure 9. Because $\frac{d}{d\mu} G(\mu) = \frac{\bar{x}}{1-n_0/n} e^{-\mu} > 0, \mu_{j+1} = G(\mu_j)$ converges for any initial value μ_0 to the MLE $\hat{\mu}_{MLE}$ satisfying the fixed point equation $\mu = G(\mu)$. $\mu_0 = \hat{\mu}_{MO}$ might be chosen as the initial value for iteration. The convergence of this algorithm is usually linear and ways of acceleration do exist (Böhning, 1993).

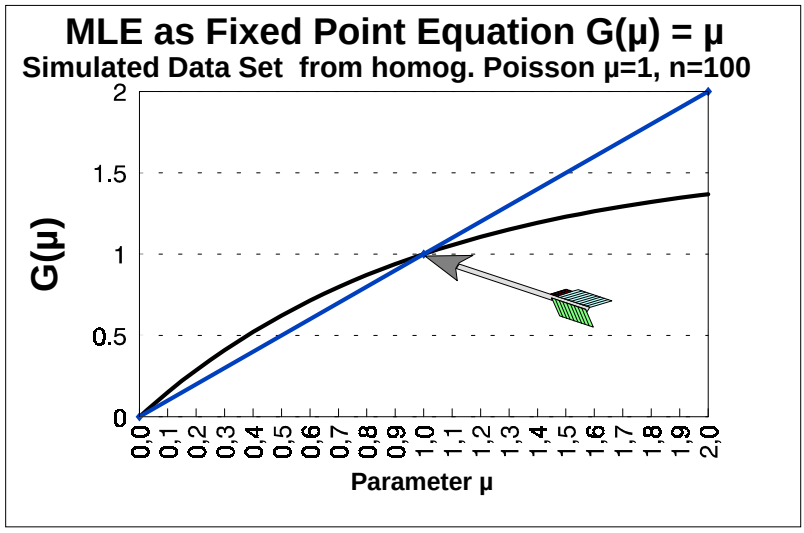

Figure 9: MLE as fixed point of G

References

Böhning, D. (1982). Convergence of Simar´s algorithm for finding the maximum likelihood estimate of a compound Poisson process. *Ann. Statist. 10,* 1006 - 1008.

Böhning, D. (1995). A review of reliable maximum likelihood algorithms for the semi-parametric mixture maximum likelihood estimator. *Journal of Statistical Planning and Inference 47,* 5-28.

Böhning, D., Schlattmann, P. and Lindsay, B.G. (1992). Computer Assisted Analysis of Mixtures (C.A.MAN): Statistical Algorithms. *Biometrics 48,* 283-303.

Böhning, D. (1993). Acceleration techniques in fixed-point methods for finding percentage points. *Statistics and Computing 3*, 1-5.

Böhning, D. (1994). A note on test for Poisson overdispersion. *Biometrika 81*, 418-419.

Dieckmann, A. (1981). Ein einfaches stochastisches Modell zur Analyse von Häufigkeitsverteilungen abweichenden Verhaltens. *Zeitschrift für Soziologie 10*, 319-325.

Dietz, E. (1992). Estimation of Heterogeneity – A GLM-Approach. In: Fahrmeir, L., Francis F., Gilchrist, R., Tutz, G. (Eds.), *„Advances in GLIM and Statistical Modelling“*, Lecture Notes in Statistics, Springer Verlag Berlin, p. 66-72.

Johnson, N., Kotz, S. and Kemp, A.W. (1992). *Univariate Discrete Distributions*, 2nd Edition, Wiley: New York.

Kuan, J., Peck, R.C. and Janke, M.K. (1991). Statistical Methods for Traffic Accident Research, in: *Proceedings of the 1990 Taipei Symposium in Statistics*, June 28-30, 1990, ed. by Min-Te Chao and Philip E. Cheng. Taipei, Institute of Statistical Science.

Lindsay, B.G. (1983). The geometry of mixture likelihoods: a general theory, *Ann. Statist. 11*, 86-94.

Lambert, D. (1992). Zero-inflated Poisson regression, with application to defects in manufacturing. *Technometrics 34*, 1-14.

Maritz, J.S. and Lwin, T. (1989). *Empirical Bayes Methods*, 2nd Edition. Chapman & Hall: London.

Mendonça, L. (1995). *Longitudinalstudie zu kariespräventiven Methoden, durchgeführt bei 7- bis 10-jährigen urbanen Kindern in Belo Horizonte (Brasilien).* Inaugural-Disseration zur Erlangung der zahnmedizinischen Doktorwürde am Fachbereich Zahn-, Mund- und Kieferheilkunde der Freien Universität Berlin.

Mendonça, L. und Böhning, D. (1994). *Die Auswirkung von Gesundheitsunterricht und Mundspülung mit Na-Fluorid auf die Prävention von Zahnkaries:* Eine Kohortenstudie mit urbanen Kindern in Brasilien. Beitrag zur 39. Jahrestagung der GMDS vom 18.- 25. September 1994 in Dresden.

Schlattmann, P. and Böhning, D. (1993). Mixture models and disease mapping, *Statistics in Medicine 12*, 1943-50.

Simar, L. (1976). Maximum likelihood estimation of a compound Poisson process. *Ann. Statist. 4*, 1200-1209.

Symons, M.J., Grimson, R.C. and Yuan, Y.C. (1983). Clustering of rare events. *Biometrics 39***,** 193-205.