Some General Points for Inflation Models

Dankmar Böhning[a], Sunisa Junnumtuam[b]

[a] *Southampton Statistical Sciences Research Institute and Mathematical Sciences, University of Southampton, Southampton, SO17 1BJ, United Kingdom*
[b] *Research Center for Academic Excellence in Mathematics, Department of Mathematics, Faculty of Science, Naresuan University, Phitsanulok, 65000, Thailand*

Abstract

This paper presents a general approach to inflation models with an arbitrary number of inflation points. Using the convex theory of discrete mixture models it achieves some general results including closed form expressions for the estimates of the inflation weights and separability of the likelihood.
Keywords: Likelihood separation, mixture modelling, several inflation points

1. Introduction

Inflation modelling for count data is most developed for zero-inflation modelling. An introduction, over- and review can be found in Young *et al.* [10]. One of the most widely used is a zero-inflated Poisson (ZIP) distribution, which was introduced by Lambert [5]. However, inflation modelling need not be restricted to zero-inflation. Recently, in the setting of marginal capture-recapture modelling, one-inflation has attracted considerable attention. A review on one-inflation modelling (with zero-truncation) is given by Böhning and Friedl [4]. However, work that has focus on more than one inflation point is rare. In some other situations, not only zeros but also another inflated count ($k > 0$) happened concurrently, which is known as zero and k-inflation; this resulted in the development of a new model. For example, Aurora et al. (2021) used a zero and k-inflated Conway-Maxwell-Poisson distribution to model the data. When a large number of zeros and ones occur simultaneously, which is called zero-one inflation. An exception is Junnumtuam *et al.* [9] who looks at zero-one-inflation in context of the Covid-19 pandemic. Furthermore, in continuous datasets, only one of the extremes, such as zero or one, may exist; thus, Ospina and Ferrari (2012) investigated a broad class of zero-or-one inflated beta regression models.

This work extends these approaches in a more general way.

- We allow a general number of inflation points.
- We find a general and closed form solution for the maximum likelihood estimates of the inflation weights. Hence there is no need to use any algorithmic approach for solving the score equations.
- We show that the fitted values agree with the empirical frequencies.
- We show that the likelihood can be separated into two parts: one part depends only on the inflation weights and does not involved the parametric density; a second part consists of the parametric model in the form of a truncated density which is truncated by all inflation points.

The inflation and the truncation model. Suppose there are $m+1$ integer points $x_0, x_1, \ldots, x_m \in \{0, 1, 2, \ldots\}$ receiving respective weights $\alpha_0, \alpha_1, ..., \alpha_m$ where $\alpha_j \geq 0$ for $j = 0, 1 \ldots, m$ and $\sum_j \alpha_j \leq 1$. We assume that there is a general base distribution $p_x = P(X = x)$ where $x = 0, 1, 2, ...$ and that this distribution is potentially inflated at the points $x_0, x_1, \ldots, x_m$ which we call the *inflation points* carrying the *inflation weights* $\alpha_0, \alpha_1, ..., \alpha_m$. The base distribution might depend on some unknown parameter θ, but we will not consider this for the time being and concentrate on the inflation part. Given this setting count data will arise from the probability mass function of some count variable X' given by (1):

$$P(X' = x) = \begin{cases} \alpha_j + \bar{\alpha} p_{x_j}; & \text{if } \ x = x_j \in \{x_0, x_1, \dots, x_m\} \\ \bar{\alpha} p_x; & \text{otherwise,} \end{cases}, \tag{1}$$

where $\bar{\alpha} = 1 - \sum_{j=0}^{m} \alpha_j$. Alternatively, we can say that the vector $(\alpha_0, \alpha_1, ..., \alpha_m, \bar{\alpha})^T$ is in the probability simplex $S = \{(\alpha_0, \alpha_1, ..., \alpha_{m+1})^T | \alpha_j \geq 0$ for $j = 0, 1, \dots, m, (m+1)$ and $\sum_{j=0}^{m+1} \alpha_j = 1\}$. The case most frequently considered is where $m = 0$ and $x_0 = 0$ leading to zero-inflation models. More recently, interest developed in the case $m = 0$ and $x_0 = 1$ corresponding to one-inflation models. We will further down consider a case where $m = 1$ and $x_0 = 0$ and $x_1 = 1$, the zero-one-inflation model.

We will also need the probability mass function of the count variable X^* which occurs if X is truncated by the inflation points $\{x_0, x_1, \dots, x_m\}$. This is given in (2):

$$P(X^* = x) = \frac{p_x}{1 - \sum_{j=0}^{m} p_{x_j}}; \text{ if } \ x \notin \{x_0, x_1, \dots, x_m\}. \tag{2}$$

We will see in section 5 that both models are strongly related.

Data and likelihood. Suppose in the sample of size N, n different counts have been observed and let $f_0, f_1, ..., f_m$ be the frequencies of the $m+1$ inflation points $x_0, x_1, \dots, x_m$ and $f_{m+1}, ..., f_n$ the frequencies of the $n - m$ non-inflated observation counts. The likelihood function L is given by

$$L = \prod_{j=0}^{m} (\alpha_j + \bar{\alpha} p_{x_j})^{f_j} \prod_{j=m+1}^{n} (\bar{\alpha} p_{x_j})^{f_j}, \tag{3}$$

from where the log-likelihood function of follows as

$$\log L = \sum_{j=0}^{m} f_j \log(\alpha_j + \bar{\alpha} p_{x_j}) + N_0 \log \bar{\alpha} + c. \tag{4}$$

The constant c in (4) does not depend on $(\alpha_0, \alpha_1, ..., \alpha_m, \bar{\alpha})^T$ and $N_0 = \sum_{j=m+1}^{n} f_j$ is the size of the sub-sample of the non-inflation counts. So, $N_1 = N - N_0 = \sum_{j=0}^{m} f_j$ is the size of the sub-sample of the inflation counts. Differentiation of $\log L$ with respect to α_j and $\bar{\alpha}$ provides

$$\frac{\partial \log L}{\partial \alpha_j} = \frac{f_j}{\alpha_j + \bar{\alpha} p_{x_j}} \text{ for } j = 0, 1, ..., m \tag{5}$$

and

$$\frac{\partial \log L}{\partial \bar{\alpha}} = \frac{N_0}{\bar{\alpha}} + \sum_{j=0}^{m} \frac{f_j p_{x_j}}{\alpha_j + \bar{\alpha} p_{x_j}}. \tag{6}$$

Note that the gradient, the vector of partial derivatives given in (5) and (6), is positive and defined as long as $(\alpha_0, \alpha_1, ..., \alpha_m, \bar{\alpha})^T$ is an interior point of the probability simplex S. Also, a direct argument using the concavity of the log-function shows that the log-likelihood is a *concave* function of the parameter vector $(\alpha_0, \alpha_1, ..., \alpha_m, \bar{\alpha})^T$. We build on this important fact in the next section. Also, for simplicity, from now on we write p_j for p_{x_j}.

2. Estimation and separation of likelihoods

Maximum likelihood estimation. As the inflation parameters model the inflation points we expect that the fitted values correspond to the relative frequencies. In other words, we assume that

$$\hat{\alpha}_j + \hat{\bar{\alpha}} p_j = \frac{f_j}{N}, \text{ for } j = 0, 1, ..., m \tag{7}$$

where $\hat{\alpha}_j$ and $\hat{\bar{\alpha}}$ are the maximum likelihood estimators for parameters α_j and $\bar{\alpha}$ for $j = 0, \ldots, m$. Now, summing up the $m + 1$ equations in (7) we achieve

$$\sum_{j=0}^{m} \hat{\alpha}_j + \hat{\bar{\alpha}} \sum_{j=0}^{m} p_j = \frac{N_1}{N}. \tag{8}$$

Using that $\sum_{j=0}^{m} \hat{\alpha}_j = 1 - \hat{\bar{\alpha}}$ we arrive at

$$\begin{aligned} 1 - \hat{\bar{\alpha}} + \hat{\bar{\alpha}} \sum_{j=0}^{m} p_j &= \frac{N_1}{N} \\ \hat{\bar{\alpha}} &= \frac{1 - N_1/N}{1 - \sum_{j=0}^{m} p_j}. \end{aligned} \tag{9}$$

This in turn delivers closed form solutions for the parameter estimates of the inflation weights as given in (10):

$$\hat{\alpha}_j = \frac{f_j}{N} - \hat{\bar{\alpha}} p_j. \tag{10}$$

We have derived intuitive estimates for the parameter $(\alpha_0, \alpha_1, ..., \alpha_m, \bar{\alpha})^T$, but now need to show that these are indeed the maximum likelihood estimators.

Theorem 2.1 (MLE)**.** *The estimators given in* (9) *and* (10) *are maximum likelihood estimators given they are all positive.*

Proof: According to the theorem for concave functionals on the probability simplex (see the appendix) we have to show that the partial derivatives at $(\hat{\alpha}_0, \hat{\alpha}_1, ..., \hat{\alpha}_m, \hat{\bar{\alpha}})^T$ are identical. We start by plugging in the estimates in the partial derivatives (5) and (6) and yield that

$$\frac{\partial \log L}{\partial \alpha_j} = \frac{f_j}{\frac{f_j}{N} - \hat{\bar{\alpha}} p_j + \hat{\bar{\alpha}} p_j} = N, \text{ for } j = 0, 1, ..., m \tag{11}$$

and

$$\frac{\partial \log L}{\partial \bar{\alpha}} = \frac{N_0}{\hat{\bar{\alpha}}} + \sum_{j=0}^{m} \frac{f_j p_j}{\hat{\alpha}_j + \hat{\bar{\alpha}} p_j} = \frac{N_0}{\frac{1-N_1/N}{1-\sum_{j=0}^{m} p_j}} + \sum_{j=0}^{m} \frac{f_j p_j}{\frac{f_j}{N} - \hat{\bar{\alpha}} p_j + \hat{\bar{\alpha}} p_j}$$

$$= \frac{N_0 - N_0 \sum_{j=0}^{m} p_j}{1 - N_1/N} + N \sum_{j=0}^{m} p_j = \frac{N_0(1 - \sum_{j=0}^{m} p_j)}{N_0/N} + N \sum_{j=0}^{m} p_j$$

$$= N(1 - \sum_{j=0}^{m} p_j + \sum_{j=0}^{m} p_j) = N \tag{12}$$

Hence all partial derivatives agree which ends the proof.

The following theorem states that the fitted values agree with the empirical relative frequencies at the inflation points.

Theorem 2.2 (fitted values)**.** *Let* $(\hat{\alpha}_0, \hat{\alpha}_1, ..., \hat{\alpha}_m, \hat{\bar{\alpha}})^T > 0$ *be the maximum likelihood estimators provided in* (9) *and* (10) *. Then*

$$\hat{\alpha}_j + \hat{\bar{\alpha}} p_j = \frac{f_j}{N}, \text{ for } j = 0, \ldots, m.$$

We omit the proof as it is obvious.

Separation of likelihoods. The total sample log-likelihood function can be expressed as

$$\log L = \sum_{j=0}^{m} f_j \log(\alpha_j + \bar{\alpha} p_j) + N_0 \log(\bar{\alpha}) + \sum_{j=m+1}^{n} f_j \log p_j. \tag{13}$$

Then we plug-in $\hat{\alpha}_j$ and $\hat{\bar{\alpha}}$ by using

$$\hat{\alpha}_j + \hat{\bar{\alpha}} p_j = \frac{f_j}{N}.$$

We see that the total log-likelihood function can be written as

$$\begin{aligned}
\log L &= \sum_{j=0}^{m} f_j \log f_j - N_1 \log N + N_0 \log\left(\frac{1 - N_1/N}{1 - \sum_{j=0}^{m} p_j}\right) + \sum_{j=m+1}^{n} f_j \log p_j \\
&= \sum_{j=0}^{m} f_j \log f_j - N_1 \log N + N_0 \log\left(1 - \frac{N_1}{N}\right) + \sum_{j=m+1}^{n} f_j \log\left(\frac{p_j}{1 - \sum_{j=0}^{m} p_j}\right) \\
&= \sum_{j=0}^{m} f_j \log\left(\frac{f_j}{N}\right) + N_0 \log\left(\frac{N_0}{N}\right) + \sum_{j=m+1}^{n} f_j \log\left(\frac{p_j}{1 - \sum_{j=0}^{m} p_j}\right).
\end{aligned} \tag{14}$$

Hence, we have a partition of the total log-likelihood function into two parts as follows:

$$\begin{aligned}
\log L &= \sum_{j=0}^{m} f_j \log\left(\frac{f_j}{N}\right) + N_0 \log\left(\frac{N_0}{N}\right) + \sum_{j=m+1}^{n} f_j \log\left(\frac{p_j}{1 - \sum_{j=0}^{m} p_j}\right) \\
&= \log L_1 + \log L_0,
\end{aligned} \tag{15}$$

where

$$\log L_1 = \sum_{j=0}^{m} f_j \log\left(\frac{f_j}{N}\right) + N_0 \log\left(\frac{N_0}{N}\right) \tag{16}$$

and

$$\log L_0 = \sum_{j=m+1}^{n} f_j \log\left(\frac{p_j}{1 - \sum_{j=0}^{m} p_j}\right). \tag{17}$$

Thus, the total sample log-likelihood function can be written as a sum of two independent log-likelihoods: $\log L = \log L_1 + \log L_0$, where L_1 and L_0 are given in (16) and (17), respectively. Note that $\log L_1$ arises from the inflation part but no longer involves neither inflation parameters nor any potential parameters arising from the base distribution. The second part $\log L_0$ involves only the base distribution but no inflation parameters – it is inflation parameter independent. Note also that it is a truncated log-likelihood, truncated for the $m + 1$ inflation points. This result contains as a special case the result for one inflation point provided in Böhning and van der Heijden [3].

The consequences of this separability have impact as it allows the focus of fitting to be placed on the base distribution.

Uncertainty assessment. So far we have discussed estimation of parameters using parts of the mixture maximum likelihood theory. When it comes to uncertainty assessment, we suggest using the nonparametric bootstrap. This can be accomplished as follows. From the observed data of size N, we sample with replacement a bootstrap sample of size the same size N. All inflation points are truncated and the truncated likelihood of the base distribution is maximized. Then the inflation weights are determined by (9) and (10). This is repeated B times, where B is typically large. From the bootstrap distribution of estimates we can determine standard errors and confidence intervals, for the latter preferable percentile confidence intervals.

3. A case study

We now apply these ideas to count data representing the number of deaths per day due to COVID-19 during 24 February 2020 and 31 December 2020 in Luxembourg, which are reported here in Table 1 (Data source: European Centre for Disease Prevention and Control (ECDC); `https://shorturl.at/bdMaE`). In Table 2 we see the log-likelihood and BIC

Table 1: The number x of COVID-19 daily new deaths in Luxembourg in 2020.

x	0	1	2	3	4	5	6	7	8	9	10	11
f_x	167	47	17	20	15	9	13	15	4	3	1	1

for the various inflation point models, starting from no inflation and then adding inflation points 1, 2 and 3. There is strong evidence for 0-1-inflation. Table 3 shows the associated inflation weights. For the 0-1 inflation model we find a large inflation weight for count 0 and a moderate inflation weight for count 1. For the 0-1-2- and 0-1-2-3-inflation models the inflation weights for 2 or 2-3, respectively, are rather small. We have also looked at other

Table 2: Some inflation models and their log-likelihoods with associated Bayesian Information Criterion (BIC) for $\log L$; considered baseline models are the Poisson, the negative-binomial and the Poisson-Lindley distribution

					BIC	
inflation points	$\log L$	$\log L_1$	$\log L_0$	Poisson	NB	Poisson-Lindley
none	-725.09	0	-725.09	1455.92	1058.25	1117.79
0	-545.88	-215.49	-330.39	1103.24	1055.55	**1048.86**
0, 1	-511.97	-306.83	-205.14	**1041.17**	**1044.31**	1049.72
0, 1, 2	-510.42	-352.04	-158.38	1043.81	1049.24	1051.43
0, 1, 2, 3	-508.79	-397.31	-111.48	1046.30	1052.04	1056.18
				no parametric baseline distribution		
0, ..., n	-504.87	-504.87			1072.92	

Table 3: Inflation weight estimates for various inflation models with a Poisson distribution as baseline; $\hat{\lambda}$ is the estimated Poisson parameter

inflation points	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$\hat{\lambda}$
0	0.5190	-	-	-	3.39
0, 1	0.5314	0.1335	-	-	4.47
0, 1, 2	0.5326	0.1380	0.0246	-	4.74
0, 1, 2, 3	0.5335	0.1427	0.0341	0.0293	5.12

Table 4: Uncertainty assessment of parameters in the zero-one inflated Poisson model

parameters	estimate	s.e.	95% percentile CI
α_0	0.5314	0.0278	0.4778 – 0.5860
α_1	0.1335	0.0205	0.0942 – 0.1747
λ	4.47	0.2564	3.9707 – 4.9687

baseline distributions such as the negative-binomial or the Poisson-Lindley distribution, but none provided a better fit as Table 2 shows. Note that the BIC-values can be compared vertically (across different number of inflation points) and horizontally (across different baseline distributions.) We have also included as a further benchmark the situation that every data point becomes an inflation point, in other words, the empirical, relative frequencies. See the last row in Table 2. It shows that the zero-one-inflated Poisson model performs by far better than the latter. Uncertainty assessment for the best-fitting model, the zero-one-inflated Poisson, has also been performed using the nonparametric bootstrap as outlined in the previous section with bootstrap replication size of $B = 10,000$.. Table 4 shows the results. The associated *R*-code is available in the supplement. Summing up, the analysis shows that a Poisson model with zero- and one-inflation is the most suitable model.

The question arises if inflation, including the number of inflation points, can be diagnosed in an easy way. One way to approach this question would be by analysis of residuals. In Figure 1 we show the Pearson residuals for the data set at hand. The standard residuals $(f_j - \hat{f}_j)/\sqrt{\hat{f}_j}$ are typically not very useful in detecting inflation as they are dominated by the large number of observations. Only if we start to consider residuals where the

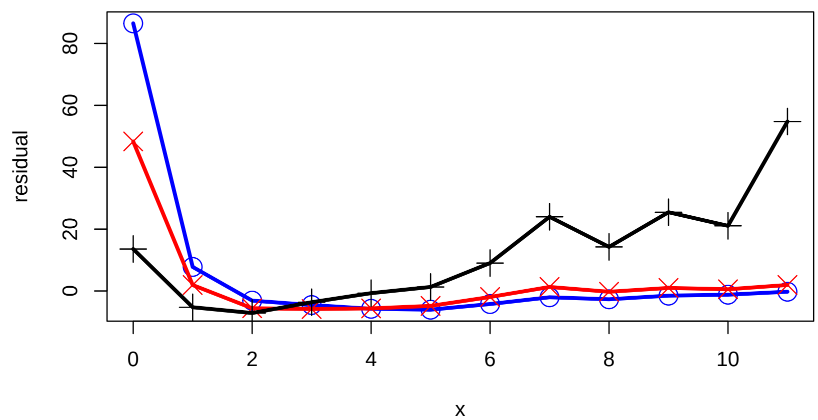

Figure 1: Residual diagnostics based on Pearson residuals for the Poisson baseline distribution; "plus" represents the standard residual, "cross" is leaving out zeros and "circle" is leaving out zeros and ones.

fit has been achieved without zeros or without zeros and ones (jackknifing), inflation becomes apparent. However, this analysis will also depend on the baseline distribution (for example, for the data set used here, in the case of a Poisson-Lindley baseline, the zero-inflated model is best), and it seems most appropriate to include this choice in the analysis as we have done using the information criteria approach above.

4. Discussion

We have presented a general theory of inflation models which bridges between fully non-parametric and fully parametric models. In fact, if the number of inflation points $m+1$ is equal to the number of different observed counts n, we obtain the empirical distribution function as Theorem 2.2 says.

If there are no inflation points, then inference is based solely on the parametric part. The choice of $m \in \{0, 1, 2, \ldots, n\}$ offers considerably flexibility. The results presented here reduce the computational burden considerably as there are cloased form solutions for the inflation weights available. Furthermore, inflation models connect to robust statistics. Any added inflation point removes this point from the inference in the non-inflated data part as (15) implies. Of course, on the parametric part a diversity of models is possible and we have mentioned here only a few, as our focus is on the generic results, which are independent of any baseline distribution. Clearly, it is possible to allow more models of semi-parametric nature for the baseline distribution, although it seems reasonable to stay with simple parametric models for the baseline distribution. Mixture models have been around for some time but it has been the fundamental work of Lindsay [6] who put discrete mixture models into the context of convex theory. This note adds a further piece into that elegant theory.

5. Appendix: An optimality result for concave functionals on the probability simplex

Let $\phi(\alpha)$ be a concave and differentiable functional defined on the probability simplex $S = \{\alpha = (\alpha_0, \alpha_1, ..., \alpha_m)^T | \alpha_j \geq 0$ for $j = 0, 1, \ldots, m$ and $\sum_{j=0}^{m} \alpha_j = 1\}$.

Theorem 5.1 (General equivalence theorem for mixtures of Lindsay[6]). *Let $\hat{\alpha} \in S$ and $\hat{\alpha} > 0$. Then the following conditions are equivalent:*

1.

$$\phi(\hat{\alpha}) \geq \phi(\alpha), \text{ for all } \alpha \in S$$

2.

$$\frac{\partial \phi}{\partial \alpha_j}(\hat{\alpha}) = \nabla \phi(\hat{\alpha})^T \hat{\alpha}, \ \textit{for } j = 0, \dots, m$$

3. *There is a constant* $c \neq 0$ *such that*

$$\frac{\partial \phi}{\partial \alpha_j}(\hat{\alpha}) = c, \ \textit{for } j = 0, \dots, m.$$

Here $\nabla \phi(\alpha)$ denotes the gradient, the vector of partial derivatives at α.

Acknowledgements

Both authors are grateful to the Editor and two anonymous reviewers for their helpful comments.

References

[1] Arora, M., Chaganty, N. R., and Sellers, K. F. (2021). A flexible regression model for zero- and k-inflated count data. *Journal of Statistical Computation and Simulation*, **91**(9), 1815 – 1845.

[2] Böhning, D. (1989). Likelihood inference for mixtures: geometrical and other constructions of monotone step-length algorithms. *Biometrika* **76** 375–383.

[3] Böhning, D. and van der Heijden, P. (2019) The identity of the zero-truncated, one-inflated likelihood and the zero-one-truncated likelihood for general count densities with an application to drink-driving in Britain. *Annals of Applied Statistics* **13**, 1109 –1211.

[4] Böhning, D. and Friedl, H. (2024). One-inflation and zero-truncation count data modelling revisited with a view on

Horvitz–Thompson estimation of population size. *International Statistical Review* doi: 10.1111/insr.12570.

[5] LAMBERT, D. (1992). Zero-Inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics*, **34**(1), 1.

[6] LINDSAY, B. (1995). Mixture Models: Theory, Geometry and Applications. In *NSF-CBMS Regional Conference Series in Probability and Statistics.* Hayward: Institute of Mathematical Statistics.

[7] OSPINA, R. AND FERRARI,S. L. (2012). A general class of zero-or-one inflated beta regression models. *Computational Statistics and Data Analysis*, **56**(6), 1609 – 1623.

[8] SANKARAN, M. (1970). The discrete Poisson-Lindley distribution. *Biometrics* **26**, 145 – 149.

[9] JUNNUMTUAM, S., NIWITPONG, S., NIWITPONG, S. (2022). A zero-and-one-inflated cosine geometric distribution and its applications. *Mathematics* **10** 4012.

[10] YOUNG, D.S., ROEMMELE, E.S., AND YEH, P. (2022). Zero-inflated modeling Part I: Traditional zero-inflated count regression models, their applications, and computational tools. *WIREs Computat. Stat.* **14** e1541.