

# Power series mixtures and the ratio plot with applications to zero-truncated count distribution modelling

Dankmar Böhning<sup>1</sup>

Received: 20 October 2014 / Accepted: 22 July 2015 / Published online: 30 July 2015  
© Sapienza Università di Roma 2015

**Abstract** The purpose of this note is to contribute some general points on how mixtures of power series distributions relate to their ratios of neighboring probabilities and how the associated graph, the ratio plot, can be used as diagnostic device as suggested in Böhning (J Comput Graph Stat 22:133–155, 2013). This work is continued here and extensively used to explore the aptness of the negative-binomial and beta-binomial model as capture-recapture zero-truncated count models. It is concluded that these models are less suitable for capture-recapture modelling as frequently readily assumed. This is mainly due to an inherent boundary problem that is elaborated here and illustrated at hand of some case studies.

**Keywords** Power series distribution · Mixtures · Negative-binomial · Beta-binomial

## 1 Introduction

We consider discrete distributions of the power series family with density

$$p_x(\theta) = a_x \theta^x / \eta(\theta), \quad (1)$$

where  $a_x$  is a known, nonnegative coefficient,  $\theta$  a positive parameter and  $x = 0, 1, \dots$  ranges over the set of nonnegative integers. Also,  $\eta(\theta) = \sum_{x=0}^{\infty} a_x \theta^x$  is the normalizing constant. The power series distributional family is very general; it contains the Poisson, the binomial, the geometric, the negative-binomial with known shape parameter, the log-series and others. In fact, it is equivalent to the one-parameter discrete exponential family. Note further that the coefficient  $a_x$  defines the specific member of the power series, for example  $a_x = 1/x!$  defines the Poisson,  $a_x = \binom{T}{x}$  for  $x = 0, \dots, T$  with positive integer  $T$  defines the binomial ( $a_x = 0$

---

The author is grateful to the Editors of the Special Issue for the invitation to submit this paper.

✉ Dankmar Böhning  
d.a.bohning@soton.ac.uk

<sup>1</sup> Southampton Statistical Sciences Research Institute and Mathematical Sciences, University of Southampton, Southampton SO17 1BJ, UK

for  $x > T$ ) or  $a_x = 1$  gives the geometric. For known value of the shape parameter  $k > 0$ , the negative-binomial  $p_x = \frac{\Gamma(x+k)}{\Gamma(x+1)\Gamma(k)}\theta^x(1-\theta)^k$  is also part of the power series family (here  $\theta \in (0, 1)$  is the event parameter). Note that the coefficient  $a_x$  is given by  $a_x = \frac{\Gamma(x+k)}{\Gamma(x+1)\Gamma(k)}$ . For  $k = 1$  the negative-binomial becomes the geometric distribution and for  $k \rightarrow \infty$  the negative-binomial approaches the Poisson distribution.

Assume that the target population of interest is heterogeneous and this heterogeneity is unobserved and, hence, can only be described by a latent variable  $Z$ . Furthermore, the joint distribution of the quantity of interest, the count  $X$ , and the latent variable  $Z$  is given by

$$f(x, z) = f(x|z)f(z),$$

where  $f(x|z) = f(x, z)/f(z)$  and  $f(z) = \int_x f(x, z)dz$ . Since the state of the latent variable  $z$  is unobserved we consider the marginal density

$$\int_z f(x|z)f(z)dz.$$

If the conditional density  $f(x|z)$  can be described by a power series density  $p_x(\theta)$  where the state of the latent variable is identified by the parameter  $\theta$  of the power series, we arrive at the general mixture model for the power series family

$$m_x = \int_\theta p_x(\theta) f(\theta)d\theta. \tag{2}$$

Whereas the modelling capacity of the power series distribution is limited, mixtures of power series distributions experience enhanced flexibility in model fitting. The mixture (2) has two parts the mixture kernel  $p_x(\theta)$  and the mixing distribution  $f(\theta)$ . If we leave the mixing distribution unspecified, the nonparametric estimate is discrete ([17]) and connects to clustering. If we choose the mixing distribution parametric and continuous the associated modelling leads latent trait approaches ([23]).

We are interested in using the above modelling with continuous, parametric mixing distributions in the context of zero-truncated count distributional modelling which arises naturally in capture-recapture experiments or studies. The size  $N$  of a target population needs to be determined. For this purpose a trapping experiment or study is done where members of the target population are identified at  $T$  occasions where  $T$  might be known or not. For each member  $i$  the count of identifications  $X_i$  is returned where  $X_i$  takes values in  $\{0, 1, 2, \dots\}$  for  $i = 1, \dots, N$ . However, zero-identifications are not observed, they remain hidden in the experiment. Hence, a zero-truncated sample  $X_1, \dots, X_n$  is observed, where we have assumed w.l.o.g. that  $X_{n+1} = \dots = X_N = 0$ . The associated zero-truncated densities will be denoted as  $p_x^+(\theta) = p_x(\theta)/[1 - p_0(\theta)]$  and  $m_x^+(\theta) = m_x(\theta)/[1 - m_0(\theta)]$  for the zero-truncated power series and the zero-truncated mixture of power series distributions, respectively.

## 2 Mixtures of power series distributions and the monotonicity of the probability Ratio

The power series (1) has an important property. If we consider ratios of neighboring probabilities multiplied by the inverse ratios of their coefficients then

$$r_x = \frac{a_x}{a_{x+1}} \frac{p_{x+1}}{p_x} = \theta, \tag{3}$$

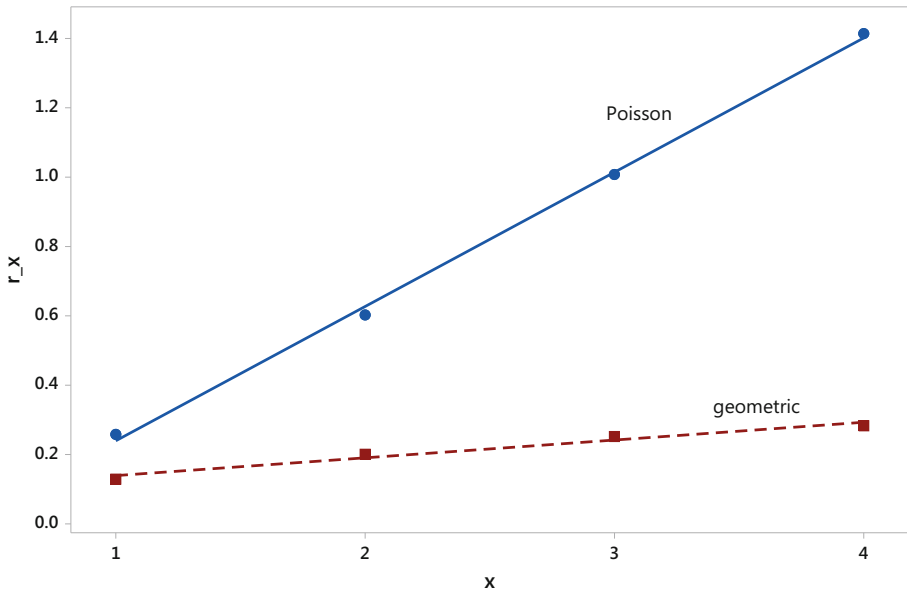
in other words,  $r_x$  is constant over the range of  $x$  with value equal to the unknown parameter  $\theta$ . Note that  $r_x$  is also identical to  $\frac{a_x}{a_{x+1}} \frac{p_{x+1}^+}{p_x^+}$ . A nonparametric estimate of  $r_x$  is readily available with  $\hat{r}_x = \frac{a_x}{a_{x+1}} \frac{f_{x+1}}{f_x}$  where  $f_x$  is the frequency of observations with count value  $x$ . The graph  $x \rightarrow \hat{r}_x$  is called ratio plot and was developed in [6] as a diagnostic device providing evidence for the aptness of a distribution.

We now show how the ratio plot might be used for a given data set. In [24] a study was reported on domestic violence in the Netherlands for the year 2009. Here the perpetrator study is reported with the data given in Table 1. To understand the table, there were 15,169 perpetrators identified being involved in a domestic violence incident exactly once, 1957 exactly twice, and so forth. In total, there were 17,662 different perpetrators identified in the Netherlands for 2009. The data represent the Netherlands except the police region for The Hague.

In Fig. 1 we see the two ratio plots, one using  $\hat{r}_x = (x + 1)f_{x+1}/f_x$  for the diagnosis of a Poisson, the other using  $\hat{r}_x = f_{x+1}/f_x$  for the diagnosis of a geometric. Clearly, the ratio plot for the Poisson does not show a horizontal line pattern and, hence, the Poisson does not appear to be an appropriate distribution for these data. The ratio plot for the geometric is much closer to a horizontal line, although some positive slope appears to be present. One could easily construct a test for linear trend based on the estimated slope parameter, but we

**Table 1** Frequencies of the number of times perpetrators have been identified in a domestic violence incident in the Netherlands in the year 2009

Year	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_{6+}$	$n$
2009	15,169	1957	393	99	28	16	17,662



**Fig. 1** Ratio plot for perpetrator domestic violence identifications in the Netherlands 2009 using the Poisson and the geometric

do not follow up on this here as there are various statistical tests for homogeneity available ([3]).

We have more interest in connecting the presence of unobserved heterogeneity (which could be described by a latent variable) with the concept of the ratio plot. We have seen in (2) that the occurrence of unobserved heterogeneity leads to the mixture of power series distributions. We can likewise consider the ratio plot for mixtures

$$r_x = \frac{a_x}{a_{x+1}} \frac{m_{x+1}}{m_x}, \tag{4}$$

where we use the coefficients  $a_x$  associated with the mixture kernel, for example, in the case of a Poisson kernel  $a_x = 1/x!$  or the case of a geometric kernel  $a_x = 1$ . The estimate of  $r_x$  will not change, however, the interpretation of the observed pattern in the ratio plot will. This is mainly due to the following result [5]:

**Theorem 1** *Let  $m_x = \int_{\theta} p_x(\theta) f(\theta) d\theta$  where  $p_x(\theta)$  is a member of the power series family and  $f(\theta)$  an arbitrary density. Then, for  $r_x = \frac{a_x}{a_{x+1}} \frac{m_{x+1}}{m_x}$  we have the following monotonicity:*

$$r_x \leq r_{x+1}$$

for all  $x = 0, 1, \dots$

A proof of this result—using the Cauchy-Schwarz inequality—is available in [5]. The ratio  $r_x = \frac{a_x}{a_{x+1}} \frac{m_{x+1}}{m_x}$  has an interesting connection to Bayesian inference. In fact,

$$\begin{aligned} r_x &= \frac{a_x}{a_{x+1}} \frac{m_{x+1}}{m_x} = \frac{a_x}{a_{x+1}} \frac{\int_{\theta} a_{x+1} \theta^x / \eta(\theta) \times \theta f(\theta) d\theta}{\int_{\theta} a_x \theta^x / \eta(\theta) f(\theta) d\theta} \\ &= \frac{\int_{\theta} a_x \theta^x / \eta(\theta) \times \theta f(\theta) d\theta}{\int_{\theta} a_x \theta^x / \eta(\theta) f(\theta) d\theta} = \int_{\theta} \theta f(\theta|x) d\theta \end{aligned}$$

is the posterior mean for prior distribution  $f(\theta)$  on  $\theta$ . Here  $f(\theta|x) = \frac{a_x \theta^x / \eta(\theta) f(\theta)}{\int_{\theta} a_x \theta^x / \eta(\theta) f(\theta) d\theta}$  is the posterior distribution. Hence  $\hat{r}_x = \frac{a_x}{a_{x+1}} \frac{f_{x+1}}{f_x}$  provides an estimate of the posterior mean without implying any knowledge on the prior distribution nor is there any requirement for estimating the prior distribution. This idea goes back to [20] and is considered the manger of empirical Bayes. For more details see [9].

The major use of the monotonicity result of Theorem 1 is the following. If a monotone pattern occurs in the ratio plot this can be taken as indicative for the presence of heterogeneity. Coming back to Fig. 1, there is a clear monotonicity present in both ratio plots. However, the linear trend is stronger for the ratio plot of the Poisson whereas the trend is less strong for the geometric. This may be interpreted in a way that part of the heterogeneity has been adjusted for in the geometric model as it is itself a Poisson mixture with a mixing exponential density:

$$(1 - p)p^x = \int \exp(-\theta)\theta^x/x! \times \frac{1}{\lambda} \exp(-\theta/\lambda)d\theta,$$

where  $p = \lambda/(1 + \lambda)$ . The slight positive trend in the ratio plot for the geometric may indicate some residual heterogeneity not adjusted for in the Poisson-exponential mixture. If the geometric distribution is appropriate a simple estimate of  $p$  can be achieved as  $[\sum_{x=2}^T f_x]/[\sum_{x=1}^{T-1} f_x]$  where  $T$  is the largest observed count. This estimate is asymptotically unbiased and feasible if  $f_1 > f_T$  which is usually the case. We mention this estimate, not necessarily as a better alternative to the maximum likelihood estimate, since it is ratio plot

based in the sense that it uses a weighted average of the ratios  $f_{x+1}/f_x, x = 1, 2, \dots, T$ . In practice, working with the geometric is unproblematic. This is in contrast to its generalization, the negative binomial distribution, which we discuss now.

### 3 The failure of the negative-binomial

It is natural to consider the negative-binomial distribution as a generalization of the geometric. The negative binomial distribution arises as a mixture of a Poisson distribution with a Gamma distribution, hence it is sometimes also called the Poisson-Gamma model. Due to its enhanced flexibility in fitting count data [8, 15, 18, 25], in comparison to the Poisson, it has been suggested as a more flexible approach in zero-truncated count data modelling ([2]; Chapter 4, [12]). We will show that using the negative-binomial frequently encounters fitting problems. These problems do also arise in untruncated count data, but are more pronounced in zero-truncated count data. The negative-binomial distribution is given as

$$p_x = \frac{\Gamma(x + k)}{\Gamma(x + 1)\Gamma(k)} p^x (1 - p)^k, \tag{5}$$

where  $p \in (0, 1)$  is the event parameter,  $k > 0$  is the shape parameter, and  $x = 0, 1, \dots$ . The negative-binomial contains as special cases the geometric, for  $k = 1$ , and the Poisson for  $k \rightarrow \infty$ . The ratio of neighboring probabilities for the negative-binomial turns out to be

$$r_x = (x + 1) \frac{p_{x+1}}{p_x} = (x + k)p = \alpha + \beta x, \tag{6}$$

where we have used the reparameterization  $\alpha = kp$  and  $\beta = p$ . This shows that, for a negative-binomial distribution, these ratios are a linear function of  $x$ . This property could also be used to develop an estimator for  $f_0$  as proposed in [21]. Here, instead, we would like to follow a suggestion in [6] to take a straight line pattern in the graph  $x \rightarrow \hat{r}_x = (x + 1)f_{x+1}/f_x$  as indicative for the aptness of the negative-binomial as a distributional model for the zero-truncated count data. To illustrate we look again at the ratio plot given in Fig. 2. There is a clear evidence that the ratios follow a straight line pattern.

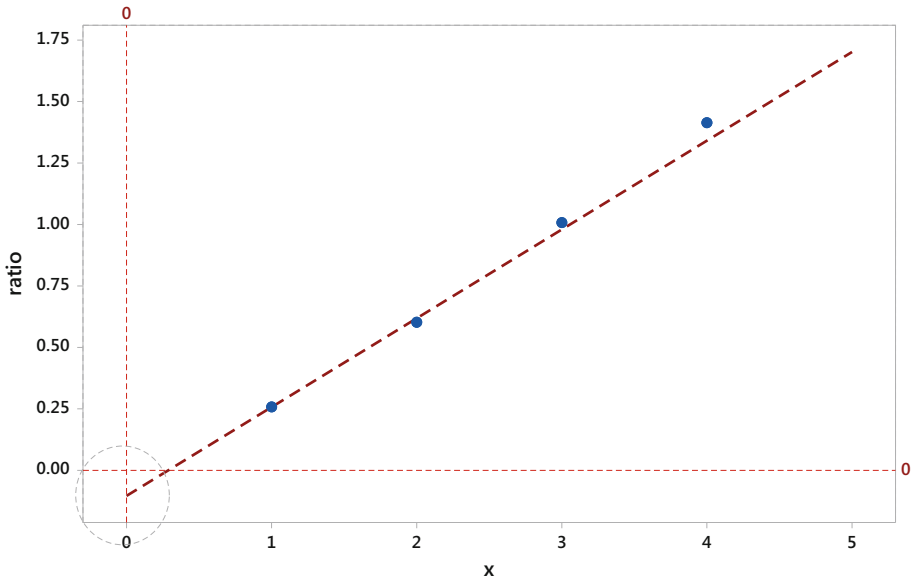
However, despite this apparent fact, the negative-binomial is not an appropriate model for these domestic violence data. As the best fitting line in Fig. 2 indicates, the estimate for the intercept is negative and hence violates the restriction  $\alpha > 0$ . This not only means that the distribution is no longer valid, it also implies that the predicted ratio  $\hat{r}_1^p = f_1/\hat{f}_0$  becomes negative, leading to a useless prediction value of  $f_0$ .

Furthermore, taking logarithms on both sides of the first equation in (6), leading to

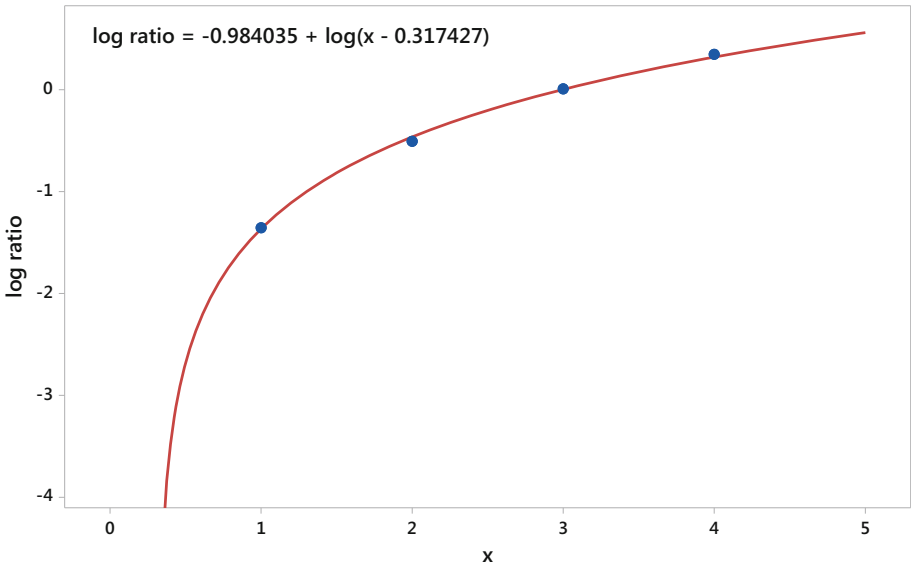
$$\log r_x = \log(1 - p) + \log(x + k),$$

does not help, as the graph  $x \rightarrow \log \hat{r}_x$  in Fig. 3 with embedded fitted nonlinear model  $\log(1 - p) + \log(x + \hat{k})$  shows. Implementing a boundary condition  $k > 0$  does not help either as the parameter estimate of  $k$  will lie on the boundary. The ratio plot does not offer a solution for the problem, but it clearly points out the reason why the negative-binomial is not appropriate. This is in contrast to standard packages such as R or STATA where only error information is returned such as parameter estimates have not converged or log-likelihood is not concave which adds more to the confusion than to its enlightenment.

The question arises to which extent the boundary problem occurs if it is known that the data come from a negative-binomial distribution. In a simulation, 100 counts were sampled from



**Fig. 2** Ratio plot for perpetrator domestic violence identifications in the Netherlands 2009; *dashed line* corresponds to the weighted least-squares regression line



**Fig. 3** Log-ratio plot for perpetrator domestic violence identifications in the Netherlands 2009; *solid line* corresponds to the weighted least-squares nonlinear model

a negative-binomial with event parameter  $p = 0.5$  and shape parameter  $k = 2$ . Zero-counts were truncated and a regression line estimated. In 100 replications, 24 % of the intercept estimates were negative indicating a boundary problem. If  $k = 2$  (doubling the variance) this percentage goes down to 12 %. If  $k = 1$  and  $p = 0.3$  (decreasing the variance) the percentage

of boundary problem occurrence goes up to 35 %. These simple simulations indicate that the boundary problem is by far not negligible. As mentioned previously, standard software is unable to deliver estimates in these situations whereas the ratio plot is able to provide some valuable insights on the occurrence of a potential boundary problem.

### 4 The failure of the beta-binomial

We now consider the situation that the number of trapping/identification occasions  $T$  is known. Then it is natural to consider the binomial

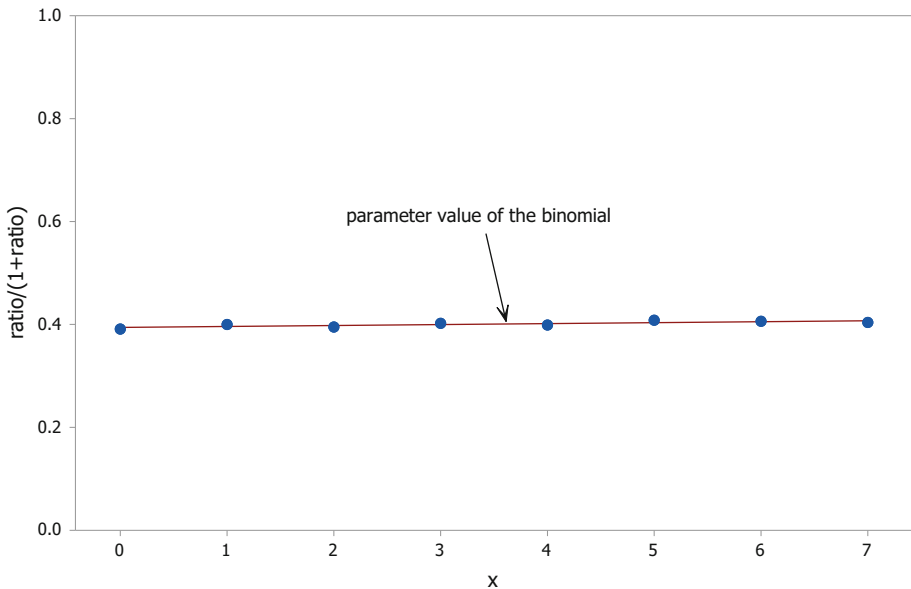
$$p_x = \binom{T}{x} \theta^x (1 - \theta)^{T-x} \tag{7}$$

as potential distribution for the count  $X$  of identifications per unit out of  $T$ . The associated ratio plot uses  $a_x = \binom{T}{x}$ , so that

$$r_x = \frac{x + 1}{T - x} p_{x+1} / p_x = \theta / (1 - \theta)$$

becomes a straight line. Note that  $r_x / (1 + r_x) = \theta$  so that  $\log r_x$  represents the canonical logit link-function. In Fig. 4 we see the ratio plot using  $\hat{r}_x = \frac{x+1}{T-x} f_{x+1} / f_x$  for 100,000 replications from a binomial with size parameter 8 and event parameter  $\theta = 0.4$ ; mainly for scaling purposes we are showing the graph  $x \rightarrow \hat{r}_x / (1 + \hat{r}_x)$ . There is clear evidence of a horizontal line with intercept  $\hat{\theta} = 0.4$ .

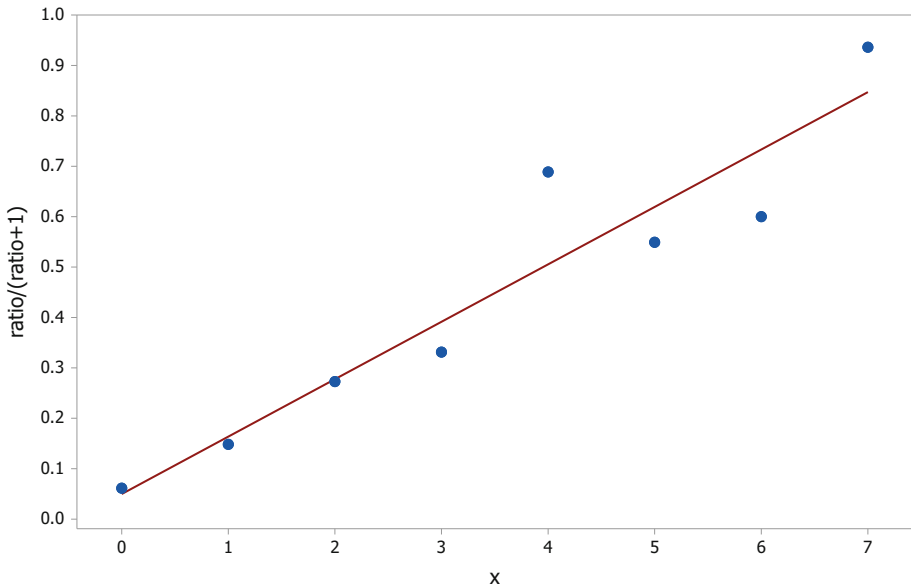
We illustrate the concept at the example of the golf tees study of St. Andrews ([7]). 250 clusters of golf tees were placed in an area of 1680 m<sup>2</sup>. The area was surveyed by eight students of the University of St. Andrews. Their task was to retrieve the golf tee clusters without prior



**Fig. 4** Ratio plot for 100,000 simulated data from a binomial with size parameter  $T = 8$  and success parameter  $\theta = 0.4$ ; solid line corresponds to the weighted least-squares line

**Table 2** Frequencies of the number of times a golf tee has been recovered by the 8 observers in the St. Andrews capture-recapture experiment

$f_0$	$f_1$	$f_2$	$f_3$	$f_4$	$f_5$	$f_6$	$f_7$	$f_8$	$n$
(88)	46	28	21	13	23	14	16	11	162



**Fig. 5** Ratio plot for golf tees identifications in the capture-recapture experiment in St. Andrews (Scotland); solid line corresponds to the weighted least-squares nonlinear model

knowledge on their placements. The results are provided with Table 2. According to this Table 46 golf tee clusters were found by exactly 1 surveyor, 28 by exactly 2 surveyors, 21 by exactly 3, etc. It should be pointed out that 88 of the 250 golf tee clusters remained undetected.

In Fig. 5, we see clear evidence that the ratios  $\hat{r}_x$  are monotone increasing and do not follow a horizontal line. This points into the direction of potential heterogeneity in golf tees detection (some golf tees might be found easier than others) but also into the direction of potential heterogeneity in golf tee detection ability among the surveyors.

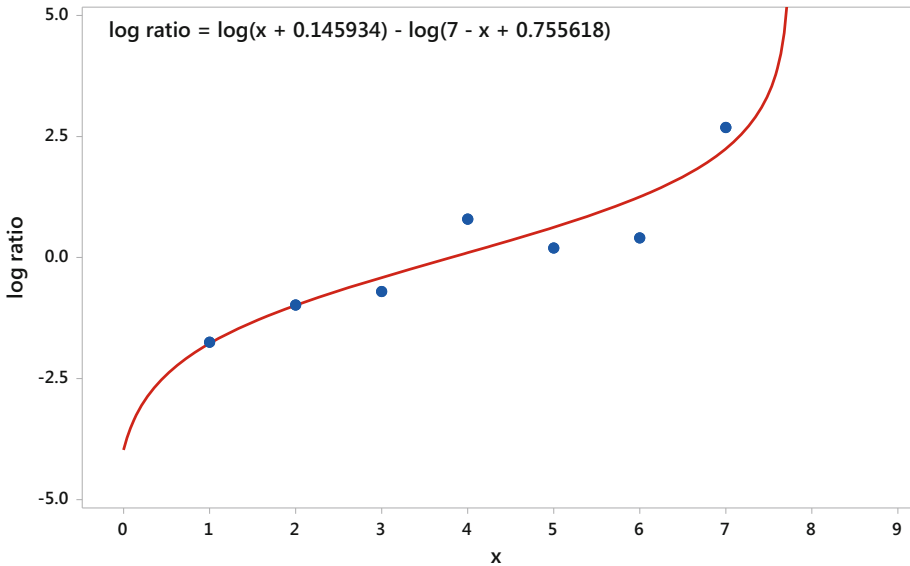
To cope with heterogeneity mixtures of binomials have been suggested, and in particular, for the setting of zero-truncated count data, the beta-binomial ([13]). The beta-binomial occurs when the binomial distribution is mixed with a beta distribution

$$m_x = \int_0^1 \binom{T}{x} \theta^x (1 - \theta)^{T-x} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1-\theta)^{\beta-1} \widehat{g(\theta)} d\theta, \tag{8}$$

where  $\alpha$  and  $\beta$  are positive parameters, and the marginal takes a simple form as the integral can be analytically solved to give

$$m_x = \binom{T}{x} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(x + \alpha)\Gamma(T - x + \beta)}{\Gamma(T + \alpha + \beta)}. \tag{9}$$





**Fig. 6** Log-ratio plot for golf tees identifications in the capture-recapture experiment in St. Andrews (Scotland); *solid line* corresponds to the weighted least-squares nonlinear model

It is easy to see that the ratio plot for the beta-binomial is provided by

$$x \rightarrow a_x \frac{m_{x+1}}{m_x} = \frac{\Gamma(x + 1 + \alpha)\Gamma(T - x - 1 + \beta)}{\Gamma(x + \alpha)\Gamma(T - x + \beta)} = \frac{x + \alpha}{T - x - 1 + \beta}, \tag{10}$$

a very simple expression on the RHS of (10). Note that  $a_x = (x + 1)/(T - x)$  in this case. In Fig. 6 we see the ratio plot on the log-scale  $x \rightarrow a_x \frac{f_{x+1}}{f_x}$  for the golf tee clusters capture-recapture study.

Also, in Fig. 6 as the solid line, the fit for the non-linear model  $\log r_x = \log(x + \alpha) - \log(7 - x + \beta)$  has been included. The odd behavior of the model at the boundaries, in particular between 0 and 1, becomes apparent. To further illustrate the erratic behavior of the beta-binomial, consider what happens if we predict  $\hat{r}_0 = \frac{1}{T} \frac{f_1}{f_0}$ . Having values  $\hat{\alpha}$  and  $\hat{\beta}$  from the model fit, we get

$$\hat{f}_0 = \frac{T - 1 + \hat{\beta}}{\hat{\alpha}} \frac{f_1}{T}, \tag{11}$$

which takes on the value 306 for the golf tees data, a substantial overestimate and misleading guess of the hidden 88 clusters. This is exactly what has been observed in simulation studies [22], but the ratio plot explains why this occurs. For the golf tees data we are in the lucky situation that the fitted parameter values for  $\alpha$  and  $\beta$  are positive, so that the beta-binomial could be fitted. But this, by no means, needs to be always the case. Even a slight perturbation of the observed value  $\log \hat{r}_x = -1.75$  to  $-2.00$  leads to the situation described in Fig. 7, where infeasible parameter estimates occur, and invoking parameter constraints would only mean that the solution lies in the boundary and no regular beta-binomial distribution fit exists (in the sense that parameter estimates lie in an open neighborhood of the feasible parameter space).

### 5 Discussion

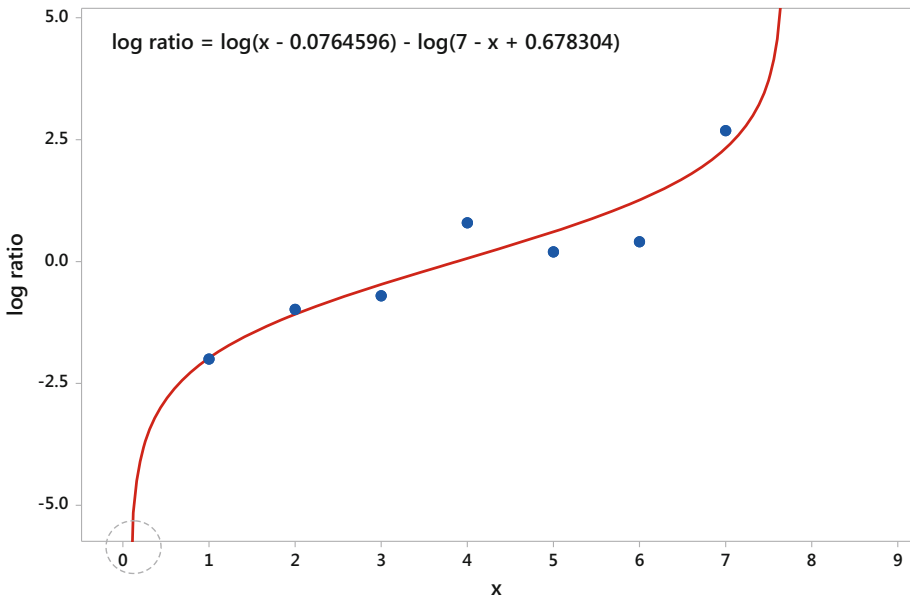
We have seen in the previous section that the beta-binomial has inherent model features that does not make this model the prime choice for predicting the missing cell frequency  $f_0$ . On the other hand, as Fig. 7 shows, the golf tees data experience a certain structure that should allow some simple form of modelling. In fact, taking the log-ratio as response, a simple straight line model seems a reasonable approximation of the observed pattern. Whereas, in generality, every regression model for  $\log r_x$  will lead to valid distribution  $p_x$ , it might be not easy, if it all, to derive at a closed form solution for this associated distribution  $p_x$ . In this case, however, when we use the model  $\log r_x = \alpha + \beta x$ , the associated distribution is the multiplicative-binomial distribution, introduced by [1], further discussed by [19] and more recently used in Freirer et al. [14] and [16]. The multiplicative-binomial has density

$$m_x = \binom{T}{x} \theta^x (1 - \theta)^{T-x} \nu^{x(T-x)} / c, \tag{12}$$

where  $c$  is the normalizing constant  $\sum_{x=0}^T \binom{T}{x} \theta^x (1 - \theta)^{T-x} \nu^{x(T-x)}$  and  $\nu > 0$  the additional parameter, indicating over- or underdispersion. If  $\nu = 1$  the multiplicative-binomial reduces to the binomial. Clearly, the ratio  $\log r_x = \log[\frac{x+1}{T-x} \frac{m_{x+1}}{m_x}]$  leads to

$$\log r_x = \log \theta / (1 - \theta) + (T - 1) \log \nu - 2x \log \nu,$$

for  $x = 0, 1, \dots, T - 1$ . This is a straight line model  $\alpha + \beta x$  with  $\alpha = \log \theta / (1 - \theta) + (T - 1) \log \nu$  and  $\beta = -2 \log \nu$ . Whereas the multiplicative-binomial involves the parameter constraints  $0 < \theta < 1$  and  $\nu > 0$ , there are no more constraints for the log ratio straight line



**Fig. 7** Log-ratio plot for golf tees identifications in the capture-recapture experiment in St. Andrews (Scotland) with the first log-ratio slightly changed to  $-2$ ; solid line corresponds to the weighted least-squares nonlinear model

model  $\alpha + \beta x$ . Note that the ratio method eliminates the unpleasant normalizing constant  $c$  as well.

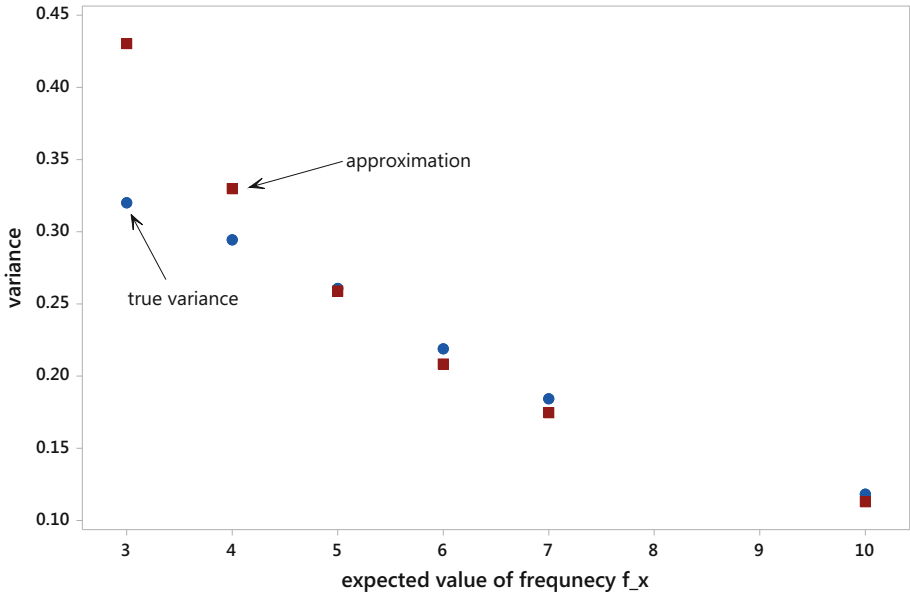
Although it is not really required for predicting the missing frequency  $f_0$  to derive the distribution associated with the regression model, it is of help interpreting the value and meaning of the regression model at hand. In closing this point, we would like to raise an open question. As the value of the additional parameter involved in the multiplicative-binomial is associated with the occurrence of over- or underdispersion, the question arises whether the multiplicative-binomial distribution can be obtained by mixing the binomial with some (yet unknown) mixing distribution to reflect unobserved heterogeneity.

We use  $\hat{r}_x = \frac{a_x}{a_{x+1}} \frac{f_{x+1}}{f_x}$  as the general estimate for  $r_x$ . This is the natural, nonparametric estimate in the count data situation and can be expected to work well with large sample sizes as it is a consistent estimator. However, if sample sizes become smaller the nonparametric estimate will be less reliable and might even fail to exist in the case of  $f_x = 0$  for some  $x$ . The latter leads to the unpleasant feature of the appearance of certain holes in the ratio plot, typically for large  $x$ . One could think of using the nonparametric mixture  $\hat{m}(x) = \int_{\theta} p_x(\theta) \hat{f}(\theta) d\theta$  with the nonparametric mixing distribution estimate  $\hat{f}(\theta)$ . However, this is prohibitive as the nonparametric mixture will impose a monotone relationship in the ratio plot which we are interested in finding out in the first place. As a simple measure a simple smoothing constant could be used, e.g. using  $f_x + c$  instead of  $f_x$  with  $c = 0.5$  as a house number. We must leave it as an open problem here how much bias this could implement in the ratio plot. To adjust for uncertainty in the ratio plot, pointwise standard error bars could be supplemented to the ratios as done before (see [6]). The variance of  $\log \hat{r}_x$  can easily be approximated, using the  $\delta$ -method, by  $1/f_{x+1} + 1/f_x$ . This approximation seems to be reasonably good for values of  $f_x$  of 5 and above as a small simulation under the Poisson assumption for  $f_x$  shows (Fig. 8). Including an error bar appears to be the best option at the moment to address the reliability of the nonparametric estimate  $\hat{r}_x$ .

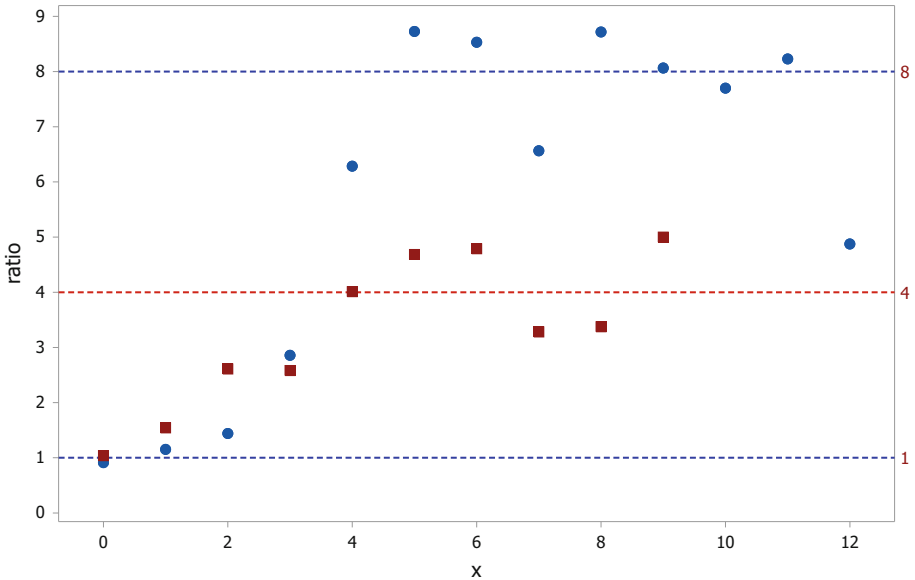
Another question arises on how a continuous mixture can be separated from a discrete mixture by means of the ratio plot. Is it possible to draw any conclusions about the nature of the mixing distribution from the appearance of the ratio plot? From the theory, in the case of discrete mixing, we would expect the ratio plot to show a step-function behavior, at least if the component distributions are well separated. Figure 9 shows a discrete mixture of two Poisson distributions (simulated with a sample size of 1000): one is well-separated with component means 1 and 8, the other less separated with component means 1 and 4. This question is also connected to the omitted covariate situation: in this case a binary covariate, which is ignored, would marginally lead to a discrete mixture. How does this affect the ratio plot?

Whereas for the first situation the discrete character of the underlying distribution is clearly recognizable (even the component means can be seen here), this is considerably more masked in the second situation. In fact, here it is even not clear whether a discrete or smooth mixing distribution operates in the background. Hence, it seems that the ratio plot can only be used to detect large contaminating components.

Another issue arises with the question to which extent the approach can be extended to the situation of multiple sources or lists. Up to now we have been implicitly assuming that the count arises from a repeated identification of the same subject within a given time period. However, a count also arises in settings where a subject is identified by several, different sources or lists. For illustration, [26] revisits an application of estimating the number of children born with Down's syndrome. Health records from five ( $T = 5$ ) different sources (hospital records, Department of Health, schools, Department of Mental Health, and obstetric



**Fig. 8** True (blue bullets) and approximated (red squares) variance of  $\log f_x$  assuming that  $f_x$  is a Poisson count with  $E(f_x) = 3, 4, 5, 6, 7, 10$  (replication size is 10,000)



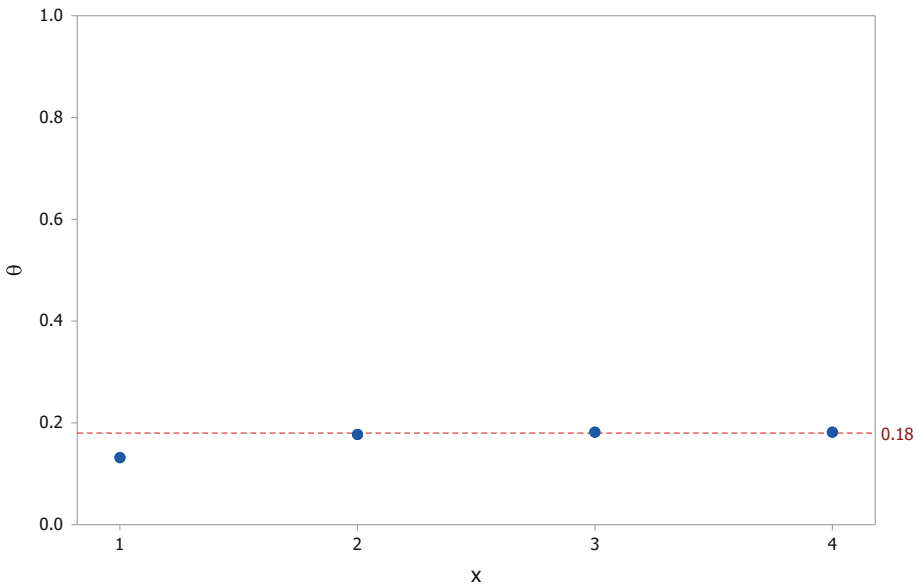
**Fig. 9** Ratio plot for two discrete Poisson mixture models: one is giving equal weights to component parameters 1 and 8 (blue bullets), the other equal weight to component parameters 1 and 4 (red squares)

records) were searched to obtain the names of children born in Massachusetts between 1955 and 1959 with the birth defect Down's syndrome. All told, 537 names were found. Of these, 248 appeared in only one source, 188 appeared in just two different sources, 81 in exactly

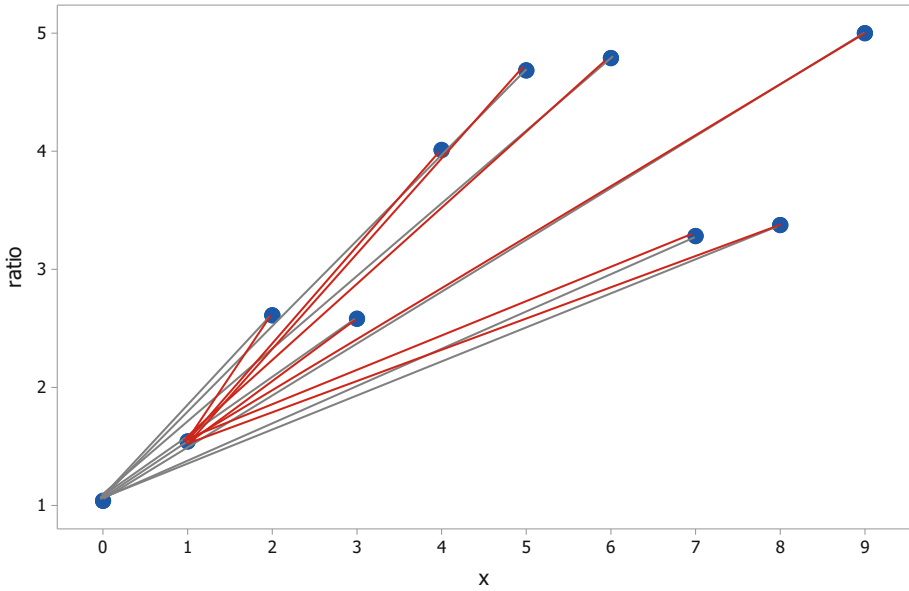
three different sources, 18 in just four and 2 in exactly 5 different sources. The problem is to estimate the total number of children born between 1955 and 1959 with Down’s syndrome.

We show in Fig. 10 the ratio plot for these data on Down’s syndrome. Evidently, the binomial distribution (using  $T = 5$ ) appears justified. It is interesting to note that [26] ignores the finite number of sources character of the data and uses a Poisson approach. In more generality, one needs to be careful using the ratio plot for marginal counts for multiple lists as multiple lists might have different identification probabilities and/or also experience complex dependency structures. A reasonable strategy seems to be to decide this on a case-by-case approach. In particular if there are few sources, the alternative log-linear modelling approach ([4]: Chapter 6) might be more appropriate. Admittedly, with a large number of lists, the marginal count ratio plot still is an interesting option worthwhile to consider.

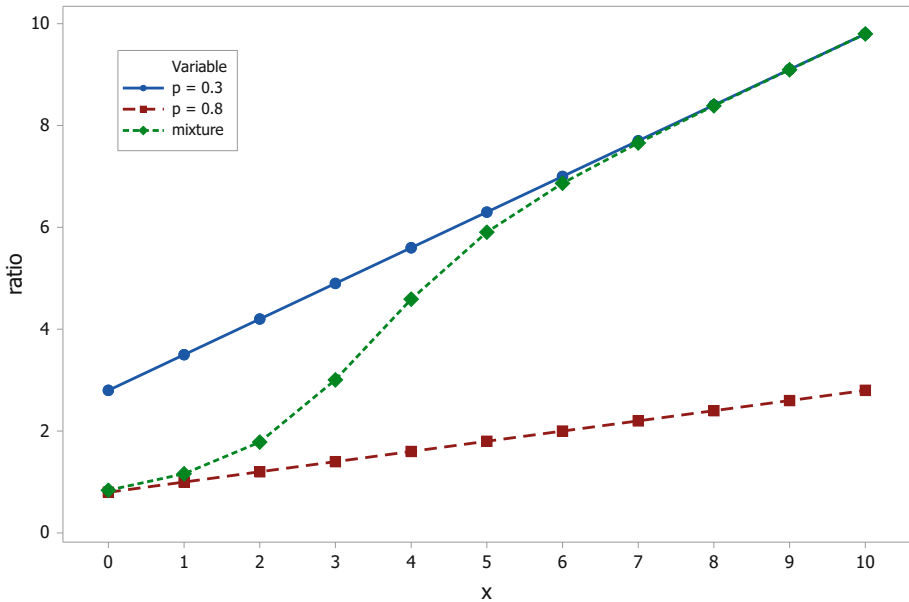
Another question is how the ratio plot connects to established estimators such as Chao’s estimator [10, 11]. Chao’s estimator of  $f_0$  has been developed as a lower bound estimator under  $m_x = \int_{\theta} p_x(\theta) f(\theta) d\theta$  where  $p_x(\theta)$  is the Poisson density and  $f(\theta)$  an arbitrary mixing distribution. The estimator takes the form  $\hat{f}_0 = f_1^2 / (2f_2)$  and has been developed using the Cauchy-Schwarz inequality. We see in Fig. 11 that  $f_1 / f_0 \leq 2f_2 / f_1$  and solving for  $f_0$  gives Chao’s estimator. Many other bounds are possible as Fig. 11 shows, for example we have that  $f_1 / f_0 \leq 4f_4 / f_3$  which leads to  $\hat{f}_0 = f_1 f_3 / (4f_4)$ . We also see that all these estimators are underestimating as they use equality (Poisson homogeneity) in their development. Clearly, Chao’s estimator has the lowest bias, but it does have bias. The ratio plot implements the idea of reducing bias. Instead of taking horizontal lines, for example  $f_1 / f_0 = 2f_2 / f_1$ , it appears more appropriate to consider lines that connect  $f_1 / f_0$  with any other ratio. These are the grey lines in Fig. 11. As  $f_0$  is not known in applications we need to use some surrogate lines. It seems natural to use  $2f_2 / f_1$  as anchor point as most of the information is frequently concentrated on ones and twos. Then this point is connected with any other point in the ratio



**Fig. 10** Ratio plot for Down’s syndrome data used in [26]: the plot shows the binomial parameter estimate  $\theta_x$  gained from the ratio  $\hat{r}_x = \frac{x+1}{T-x} f_{x+1} / f_x$

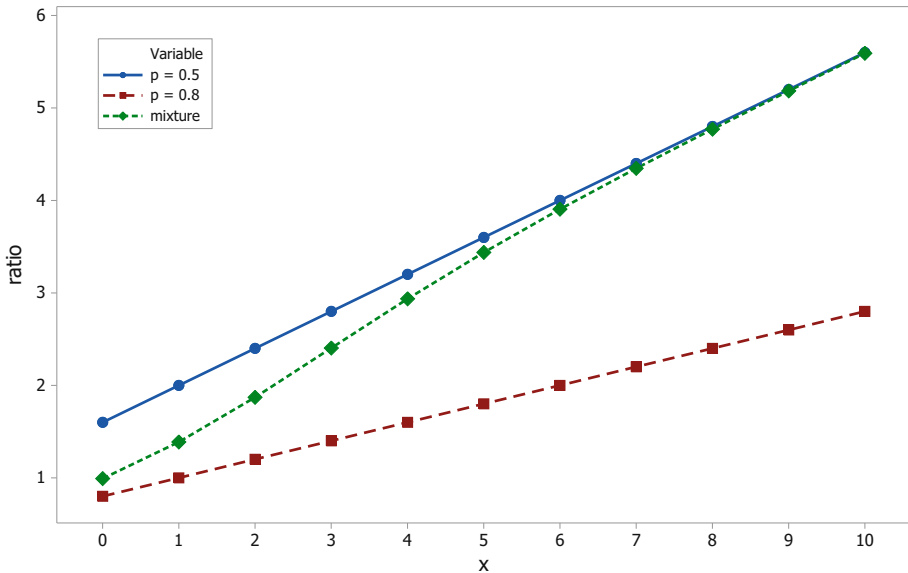


**Fig. 11** Ratio plot for simulated data arising from a two-component, equally weighted mixture of Poisson distributions with sample size 1000



**Fig. 12** Ratio plot for a two-component, equally weighted mixture of negative-binomial distributions with respective event parameter 0.3 and 0.8 and common shape parameter  $k = 4$

plot. These are the red lines. The line  $\alpha + \beta x$  that connects  $2f_2/f_1$  with  $4f_4/f_3$  leads to  $\hat{f}_0 = f_1/[3f_2/f_1 - 2f_4/f_3]$  with a value of 205 for the simulated data set of Fig. 11. This compares well with the true (but ignored value of  $f_0 = 202$ ). Chao's estimate is here 136.



**Fig. 13** Ratio plot for a two-component, equally weighted mixture of negative-binomial distributions with respective event parameter 0.5 and 0.8 and common shape parameter  $k = 4$

Of course, the choice of the line was lucky and a less lucky choice is the line that connects  $2 f_2/f_1$  with  $3 f_3/f_2$  leads to  $\hat{f}_0 = f_1/[4 f_2/f_1 - 3 f_3/f_2]$  with a value of 442, clearly far off from the true value. As some lines will lie above, some below, a proper statistical approach could develop some form of averaging. This is beyond the scope of the present paper but it shows the creative power of the ratio plot.

The ratio plot for a negative-binomial distribution is a straight line. The larger the variance of the negative-binomial, the steeper the slope. The question arises how departures from the negative-binomial could result in specific forms of the ratio plot. For example, how would the ratio plot appear for more dispersed distributions than the negative-binomial? We take as an example an equally weighted, discrete mixture of two negative-binomials with event parameters 0.3 and 0.8 (and common shape parameter  $k = 4$ ).

The associated ratio plot is provided in Fig. 12. It can be seen that the ratio plot for the mixture moves from the straight line for ratio plot of the negative-binomial with event parameter 0.3 to the ratio plot for the negative-binomial with event parameter 0.8. This clear pattern arises as the two components in the mixture are well-separated. If the two components become closer the pattern of the ratio plot changes and the mixture might be more difficult to be separated from a negative-binomial itself as Fig. 13 indicates where an equally weighted, discrete mixture of two negative-binomials with event parameters 0.5 and 0.8 (and common shape parameter  $k = 4$ ) is considered.

## References

1. Altham, P.M.E.: Two generalizations of the binomial distribution. *J. Roy. Stat. Soc. Ser. C (Appl. Stat.)* **27**, 162–167 (1978)
2. Amstrup, S.C., McDonald, T.L., Manly, B.F.J.: *Handbook of Capture–Recapture Analysis*. University Press, Princeton (2005)

3. Baksh, M.F., Böhning, D., Lerdsuwansri, R.: An extension of an over-dispersion test for count data. *Comput. Stat. Data Anal.* **55**, 466–474 (2011)
4. Bishop, Y.M., Fienberg, S.E., Holland, P.W.: *Discrete Multivariate Analysis. Theory and Practice*. Springer, New York (2007)
5. Böhning, D., Del Rio Vilas, V.: Estimating the hidden number of Scrapie affected holdings in Great Britain using a simple, truncated count model allowing for heterogeneity. *J. Agric. Biol. Environ. Stat.* **13**, 1–22 (2008)
6. Böhning, D., Baksh, M.F., Lerdsuwansri, R., Gallagher, J.: The use of the ratio-plot in capture-recapture estimation. *J. Comput. Graph. Stat.* **22**, 133–155 (2013)
7. Borchers, D.L., Buckland, S.T., Zucchini, W.: *Estimating Animal Abundance. Closed Populations*. Springer, Heidelberg (2004)
8. Cameron, A.C., Trivedi, P.K.: *Regression Analysis of Count Data*. Cambridge University Press, Cambridge (1996)
9. Carlin, B.P., Louis, T.A.: *Bayesian Methods for Data Analysis. Monographs on Statistics and Applied Probability*, 3rd edn. Chapman & Hall/CRC, London (2011)
10. Chao, A.: Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791 (1987)
11. Chao, A.: Estimating population size for sparse data in capture-recapture experiments. *Biometrics* **45**, 427–438 (1989)
12. Cruyff, M.J.L.F., van der Heijden, P.G.M.: Point and interval estimation of the population size using a zero-truncated negative binomial regression model. *Biometrical J.* **50**, 1035–1055 (2008)
13. Darazio, R.M., Royle, J.A.: Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics* **59**, 351–364 (2003)
14. Feirer, V., Friedl, H., Hirn, U.: Modelling over- and underdispersed frequencies of successful ink transmissions onto paper. *J. Appl. Stat.* **40**, 626–643 (2013)
15. Hilbe, J.M.: *Negative Binomial Regression*, 2nd edn. University Press, Cambridge (2007)
16. Leask, K., Haines, L.M.: The Altham-Poisson distribution. *Stat. Model. Int. J.* (2014) (**accepted for publication**)
17. Lindsay, B.G.: Mixture models: theory, geometry, and applications. In: *NSF-CBMS Regional Conference Series in Probability and Statistics*, vol. 5. IMS, Hayward (1995)
18. Lindsey, J.K.: *Modelling frequency and Count Data*. Clarendon Press, Oxford (1995)
19. Lovison, G.: An alternative representation of Altham's multiplicative-binomial distribution. *Stat. Probab. Lett.* **36**, 415–420 (1998)
20. Robbins, H.: An empirical Bayes approach to statistics. In: *Proceedings of 3rd Berkeley Symposium on Math Statist. and Prob.*, 1, pp. 157–164. University of California Press, Berkeley (1955)
21. Rocchetti, I., Bunge, J., Böhning, D.: Population size estimation based upon ratios of recapture probabilities. *Ann. Appl. Stat.* **5**, 1512–1533 (2011)
22. Rocchetti, I., Alfö, M., Böhning, D.: A regression estimator for mixed binomial capture-recapture data. *J. Stat. Plann. Infer.* **145**, 165–178 (2014)
23. Rost, J., Langeheine, R.: *Applications of Latent Trait and Latent Class Models in the Social Sciences*. Waxmann, Münster (1997)
24. Van der Heijden, P.G.M., Cruyff, M., Böhning, D.: Capture-recapture to estimate crime populations. In: *Bruinsma, G.J.N., Weisburd, D.L. (eds.) Encyclopedia of Criminology and Criminal Justice*, pp. 267–278. Springer, Berlin (2014)
25. Winkelmann, R.: *Econometric Analysis of Count Data*. Springer, New York (2003)
26. Zelterman, D.: Robust estimation in truncated discrete distributions with applications to capture-recapture experiments. *J. Stat. Plann. Infer.* **18**, 225–237 (1988)
27. Zelterman, D.: *Models for Discrete Data*. Oxford University Press, Oxford (2006)