Statistical Methods in Medical Research
http://smm.sagepub.com/

The covariate-adjusted frequency plot

Heinz Holling, Walailuck Böhning, Dankmar Böhning and Anton K Formann

Stat Methods Med Res published online 1 February 2013
DOI: 10.1177/0962280212473386

The online version of this article can be found at:
http://smm.sagepub.com/content/early/2013/01/30/0962280212473386

Published by:
SAGE
http://www.sagepublications.com

Original Article

SMMR
STATISTICAL METHODS IN MEDICAL RESEARCH

Statistical Methods in Medical Research
0(0) 1–15

DOI: 10.1177/0962280212473386
smm.sagepub.com
SAGE

The covariate-adjusted frequency plot*

Heinz Holling,[1] Walailuck Böhning,[1] Dankmar Böhning[2] and Anton K Formann[3,†]

Abstract

Count data arise in numerous fields of interest. Analysis of these data frequently require distributional assumptions. Although the graphical display of a fitted model is straightforward in the univariate scenario, this becomes more complex if covariate information needs to be included into the model. Stratification is one way to proceed, but has its limitations if the covariate has many levels or the number of covariates is large. The article suggests a marginal method which works even in the case that all possible covariate combinations are different (i.e. no covariate combination occurs more than once). For each covariate combination the fitted model value is computed and then summed over the entire data set. The technique is quite general and works with all count distributional models as well as with all forms of covariate modelling. The article provides illustrations of the method for various situations and also shows that the proposed estimator as well as the empirical count frequency are consistent with respect to the same parameter.

Keywords

frequency plot, adjusting for covariates, residual analysis

1 Introduction

We are interested in the question on how well a given model fits a given set of counts. Consider a random variable Y with possible count values $y \in \{0, 1, 2, \ldots\}$ and suppose further a random sample $y_1, y_2, \ldots, y_n$ is available. Also, let f_y be the frequency of counts in the sample $y_1, y_2, \ldots, y_n$ equal to y where y ranges from 0 to the largest observed count $m = \max\{y_1, \ldots, y_n\}$. Suppose there is a candidate distributional model $P(Y = y) = p_y = p_y(\theta)$ available which might depend on some

[1]Statistics and Quantitative Methods, Faculty of Psychology and Sports Science, University of Münster, Münster, Germany
[2]Southampton Statistical Sciences Research Institute, Mathematics & Medicine, Southampton, UK
[3]Faculty of Psychology, University of Vienna, Vienna, Austria

*Part of this article represents joint research with the late Anton (Toni) K Formann and was initially presented at a memorial symposium for Toni Formann at the 16th of January 2012 in Vienna.

†Anton (Toni) K Formann died 12th of July 2010 in Vienna (Austria). Details on his career and life can be found at http://en.wikipedia.org/wiki/Anton_Formann.

Corresponding author:
Dankmar Böhning, Southampton Statistical Sciences Research Institute, Mathematics & Medicine, Southampton, SO17 1BJ, UK.
Email: d.a.bohning@soton.ac.uk

unknown parameter θ and which we would like to compare to f_y. Conventional practice is to find a consistent estimate $\hat{\theta}_n$ of θ and compare $p_y(\hat{\theta}_n) \times n$ with f_y. The rationale of such a procedure is that

$$p_y(\hat{\theta}_n) \to p_y(\theta) \text{ and } f_y/n \to p_y(\theta),$$

if the model is correct, in other words, if $Y_i \sim p_y(\theta)$. Here "$\to$" refers to convergence in probability. Hence, in such a situation, we can expect that $p_y(\hat{\theta}_n) \times n$ and f_y are *close* in some sense, at least if n is becoming large.

1.1 Example: Poisson and binomial

We consider the *Poisson* distribution:

$$P(Y = y) = p_y(\theta) = Po(y|\theta) = \exp(-\theta)\theta^y/y!,$$

where θ is the Poisson parameter, usually unknown. Assume the Poisson model holds and assume further that we have a consistent estimator $\hat{\theta}_n$ of θ – an example is $\hat{\theta}_n = (y_1 + y_2 + \cdots + y_n)/n$ – then $Po(y|\hat{\theta}_n) \to Po(y|\theta)$ for every $y = 0, 1, 2, \ldots$. Hence we compare $nPo(y|\hat{\theta}_n)$ and f_y for $y = 0, 1, \ldots$, for example, by producing a graph in which $nPo(y|\hat{\theta}_n)$ and f_y are plotted against y. Pawitan[1] (p. 84) – as an example of a standard textbook illustration – exemplifies this idea for count data on accidents among factory works for the Poisson (and negative binomial) distribution. Maritz and Lwin[2] (p. 256) discuss count data on oilwell discoveries in Alberta (Canada). Here the quantity of interest is the number Y_i of oilwell discoveries of the 3rd month of the i-th half year for the 18-years period from 1953 to 1970, so $i = 1, 2, \ldots, 36$. Figure 1 shows the frequency f_y as 19, 10,

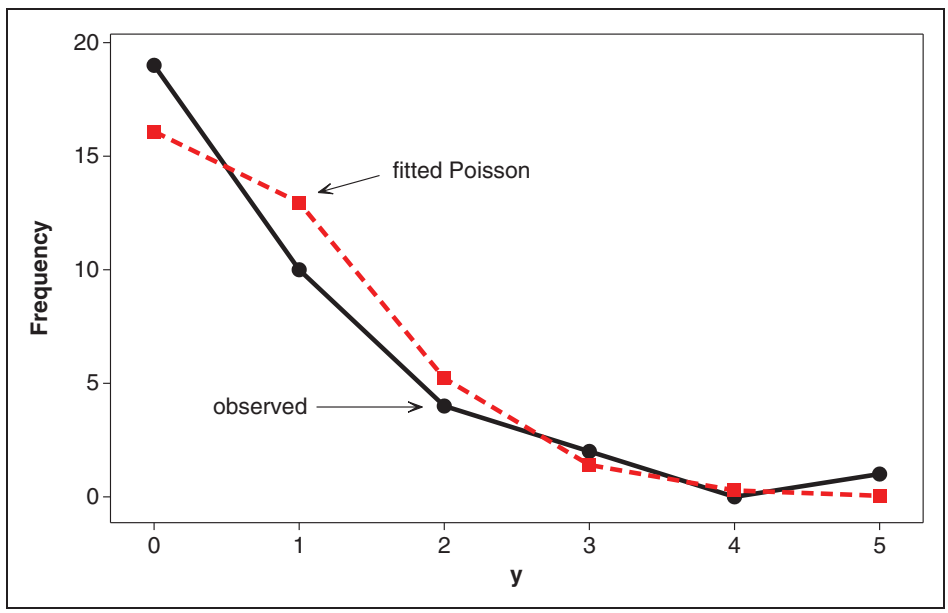

Figure 1. Empirical f_y and fitted Poisson distribution $Po(y|\hat{\theta}) \times n$ for counts Y_i of oilwell discoveries in Alberta.

4, 2, 0, 1 for each of the different observed counts $y = 0, 1, 2, 3, 4, 5$, respectively. The fitted Poisson distribution using $\hat{\theta} = \sum_{y=0}^{5} y f_y / n$ is included as well.

Similarly, we can consider the *binomial* distribution

$$P(Y = y) = p_y(\theta) = Bi(y|\theta, m) = \binom{m}{y} \theta^y (1-\theta)^{m-y},$$

where $\theta \in (0, 1)$ is the event parameter and m the trial size parameter. A consistent estimate of θ can easily constructed as $\hat{\theta}_n = \frac{\sum y_i}{nm}$, so that $nBi(y|\hat{\theta}_n)$ and f_y can be compared. To illustrate, following Goldberg[3] and Der and Everitt[4] (p. 187), consider count data arising from a study of the psychiatric screening questionnaire GHQ (general health questionnaire) which delivers an integer score from 0 to 10, in this case. The distribution of this score for a group of 131 women without any evidence of psychiatric disorder is given in Figure 2 ($f_y = 80, 29, 15, 3, 2, 1, 1$ for $y = 0, 1, 2, 3, 4, 5, 6$, respectively).

Figure 2 shows, besides f_y, also the fitted binomial model. Clearly, the binomial does not fit well, in particular for counts of 0. The χ^2-statistic $\sum_{y=0}^{4} [f_y - np_y(\hat{\theta}_n)]^2 / [np_y(\hat{\theta}_n)]$ provides 26.04, highly significant on a χ^2-scale with 3 df (note that frequencies larger or equal to 4 have been collapsed). The lack-of-fit can easily be coped with by modelling a zero-inflated binomial model defined

$$ZBi(y|\alpha, \theta, m) = \begin{cases} \alpha \binom{m}{y} \theta^y (1-\theta)^{m-y}, & \text{if } y > 0 \\ (1-\alpha) + \alpha(1-\theta)^m, & \text{if } y = 0 \end{cases} \quad (1)$$

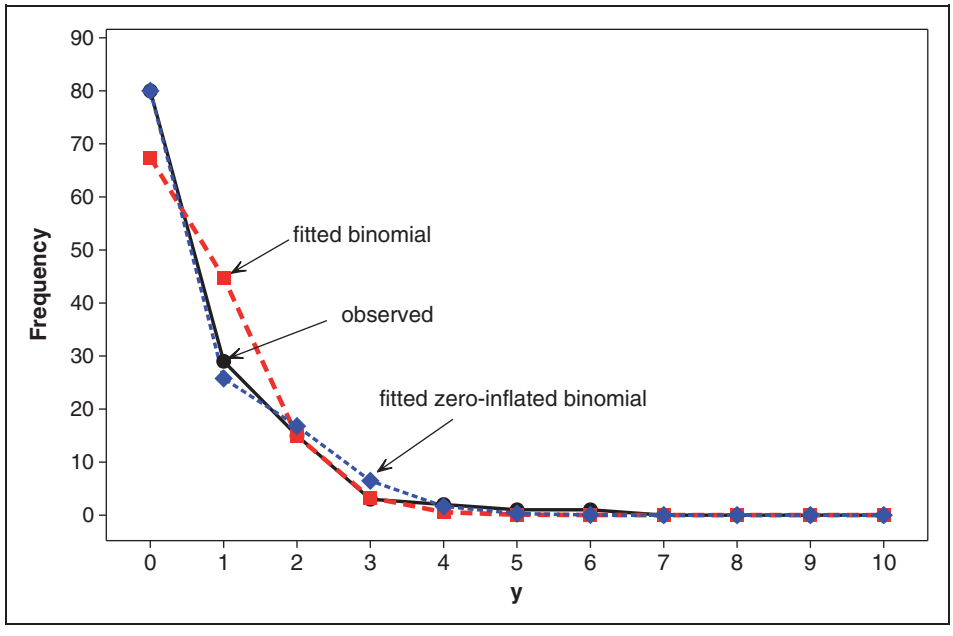

Figure 2. Empirical f_y and fitted binomial and zero-inflated binomial distribution $Bi(y|\hat{\theta}) \times n$ for counts Y_i of GHQ scores for 131 healthy females.
GHQ: general health questionnaire.

The ZBi-model has simply an extra parameter α for the amount of additional zero counts in the data and leads to a considerable improvement in goodness-of-fit as can be seen in Figure 2. The associated goodness-of-fit is $\chi^2 = 4.59$ by 2 df which is no longer significant. In addition, the likelihood ratio test comparing (1) with the standard Poisson is highly significant.

These simple examples in connections with Figures 1 and 2 show that fitted frequency plots are helpful in illustrating goodness-of-fit and potentially illuminating observations or regions of good or poor fit. They are clearly of supplementary value not replacing a thorough diagnostic analysis based on appropriate statistical measures. Hence it appears of interest to construct fitted frequency plots for more complex situations that arise in modelling where it is less clear how a fitted frequency plot should be constructed, but an overall graphical assessment of goodness-of-fit is even more desirable than in situations with one simple parameter. We illustrate the difficulties in the following with two examples. In section 2 we then outline a general way how a covariate-adjusted frequency plot should be constructed, followed in section 3 with some more complex examples for illustration and close in the final section with a brief discussion.

1.2 Some complications

In the Poisson case, the situation becomes more complex if, in addition to the sample of counts y_1, y_2,..., y_n, we have a sample of associated values e_1, e_2,..., e_n. These very often arise in situations where each count y_i is connected with some baseline expected value e_i. In the the oilwell discovery example these e_is could represent different detection efforts which might vary over the years or seasons. In epidemiology, these baseline values represent expected number of cases, calculated from a background population and which are then compared to the observed number of cases as y_i/e_i, the so-called *standardised mortality ratio* (SMR). The typical model of interest for count y_i is

$$P(Y_i = y) = Po(y|\theta, e_i) = \exp(-\theta e_i)(\theta e_i)^y / y!,$$

for every $y = 0, 1, 2,\ldots$ Although a consistent estimator $\hat{\theta}_n = \frac{\sum_i y_i}{\sum_i e_i}$ of θ can easily be constructed, it is less clear to which object f_y should be compared to since there is not a unique $Po(y|\hat{\theta}_n e)$ but there are many fitted values $Po(y|\hat{\theta}_n e_i)$, for $i = ,\ldots, n$ and a fixed count y.

Similarly, if the y_i arise out of a set of possible values $\{0,\ldots, m_i\}$ with varying upper bound m_i. This situation occurs frequently when samples are taken from clusters such as households, communities, small areas, herds or farms – just to mention a few. This situation would also occur in the example of the GHQ score if an individual would only answer a part of the given items. Here interest would be to investigate the validity of the binomial model for count y_i out of m_i:

$$P(Y_i = y) = p_y(\theta) = Bi(y|\theta, m_i) = \binom{m_i}{y}\theta^y(1-\theta)^{m_i - y}.$$

Again, a consistent estimator of θ is constructed easily as $\hat{\theta}_n = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n m_i}$, but it remains again unclear to which object f_y should be compared to. To illustrate how misleading graphs of $y \to f_y$ for the binomial with varying binomial denominator m_i can be, we consider Figure 3. Here 500 counts have been sampled from binomials with size parameter m_i equal to 100 whereas the other 500 were sampled from a binomial with size parameter 200. In all cases, the binomial event parameter is $\theta = 0.25$, hence it is representing a homogeneous binomial distribution. The graph, however, gives a different impression, for example, one might be lead to the impression that a two-component

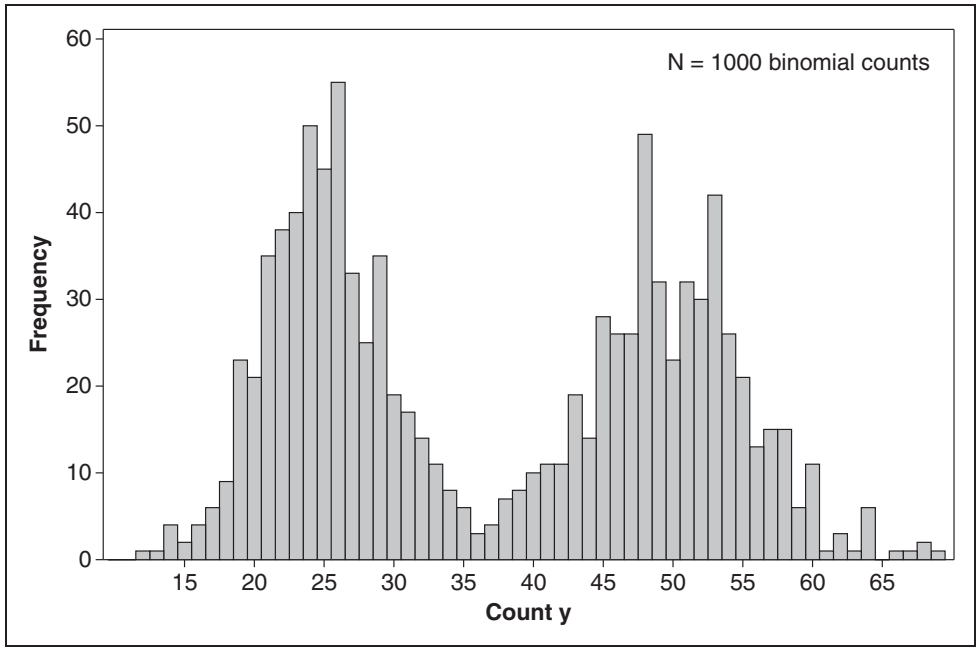

Figure 3. Sample of $n = 1000$ counts from a binomial with $\theta = 0.25$, 50% have size parameter $m_i = 100$ whereas the remaining 50% have size parameter $m_i = 200$.

mixture of binomials might be a likely mechanism behind the data generating process. Note that it is the specific distribution of the m_i which distorts the graph. Typically, it would not be known if a homogeneous event parameter θ is the valid model. Hence it is an important issue how the distribution of the m_i can be taken into account in constructing a fit of the model. Simple approaches such as choosing the mean of the distribution of the m_i as the trial size parameter also give misleading answers. To achieve a valid graph we propose to construct for every $y = 0, 1, 2, \ldots, m$

$$\hat{f}_y(\hat{\theta}_n) = \sum_{i=1}^{n} Bi(y|\hat{\theta}_n, m_i) \tag{2}$$

the margin over all n binomial fits. Note that n corresponds to the size of the sample, so that every sample point contributes to the estimator defined in equation (2). Also, in (2) we are using the convention $Bi(y|\hat{\theta}_n, m_i) = 0$ for $y > m_i$. Figure 4 indicates that this is leading to the right conclusion, e.g. all counts come from a binomial distribution with homogeneous event parameter.

2 The proposal

We assume that count Y_i follows a distributional model $p(\lambda(\theta, \eta_i))$ where θ is an unknown parameter or parameter vector, η_i is a known number or vector, and $\lambda(.,.)$ is a known function. To illustrate in the Poisson, we have that $p_y(\lambda(\theta, \eta_i)) = Po(y|\theta \times e_i)$ with $\eta_i = e_i$ and $\lambda(\theta, e) = \theta e$ in the case of specific Poisson means θe_i or $p_y(\lambda(\theta, \eta_i)) = Po(y|\exp(\beta^T \mathbf{x}_i))$ with $\lambda(\theta, \eta_i) = \exp(\beta^T \mathbf{x}_i)$, $\eta_i = \mathbf{x}_i$, and $\theta = \beta$ in the case of Poisson regression.

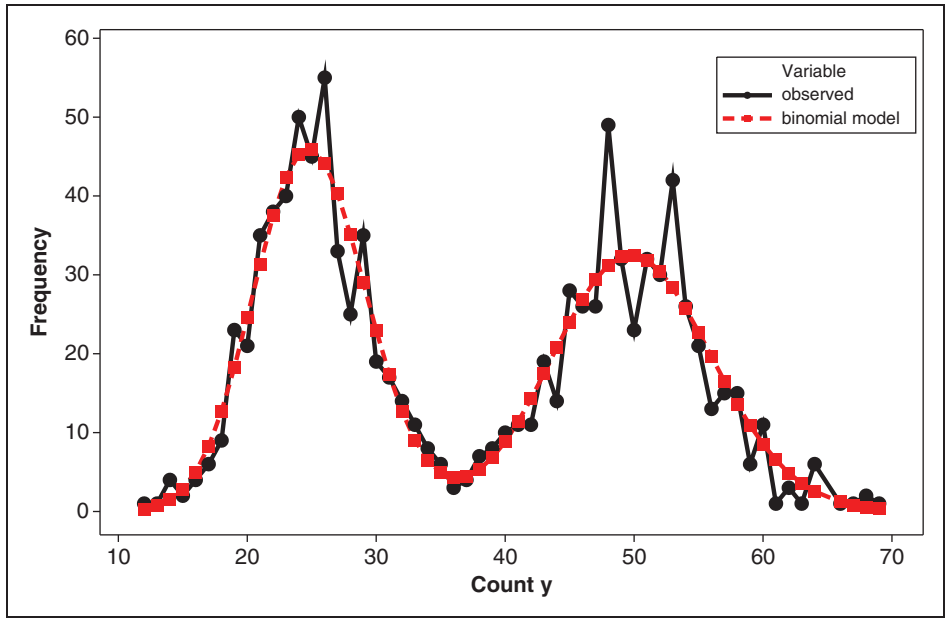

Figure 4. f_y and $\sum_{i=1}^{n} Bi(y|\hat{\theta},m_i)$ for sample of $n = 1000$ counts from a binomial with $\theta = 0.25$, 50% have size parameter $m_i = 100$ whereas the remaining 50% have size parameter $m_i = 200$.

Definition 1: *Given the situation above, we define the* covariate adjusted frequency plot *as*

$$\hat{f}_y(\hat{\theta}_n) = \sum_{i=1}^{n} p_y(\hat{\lambda}_i)$$

and the covariate adjusted probability plot *as*

$$\hat{p}_y(\hat{\theta}_n) = \frac{1}{n}\sum_{i=1}^{n} p_y(\hat{\lambda}_i)$$

where $\hat{\lambda}_i = \lambda(\hat{\theta}_n, \eta_i)$ *for* $i = 1,\ldots, n$ *and* $\hat{\theta}_n$ *is a consistent estimator of* θ.

This definition constructs the object $\hat{f}_y(\hat{\theta}_n)$, a marginal operation over the distribution of the η_i, which allows comparison to f_y. The rationale for this procedure is the following argument. Assume there is an infinite sequence $(\eta_i)_{i\geq 1}$ and each η_i is sampled with a probability w_i, $w_i \geq 0$ and $\sum_{i=1}^{\infty} w_i = 1$. Assume further that, conditional on η_i and θ, the count Y_i follows the density $p_y(\lambda_i)$ where $\lambda_i = \lambda(\theta, \eta_i)$. We will write also for the sake of brevity $Y_i \sim p_y(\lambda_i)$. We can imagine this process as a two-stage procedure where in the first stage the η_is are sampled with probability w_i

$$\eta_1, \eta_2, \ldots$$
$$w_1, w_2, \ldots$$

and, conditional on sampled η_i, we sample Y_i in the second stage

$$Y_{i|\eta_i} \sim p_y(\lambda_i), \lambda_i = \lambda(\theta, \eta_i).$$

Hence, it is clear that marginally

$$Y \sim \sum_{i=1}^{\infty} p_y(\lambda_i) w_i.$$

We assume that $s_y(\theta) = \sum_{i=1}^{\infty} p_y(\lambda_i) w_i < \infty$.

Theorem 1: *Let $\lambda(.,.)$ be continuous and $\hat{\theta}_n$ be a consistent estimator of θ. Then*

$$\hat{f}_y(\hat{\theta}_n)/n = \frac{1}{n} \sum_{i=1}^{n} p_y(\hat{\lambda}_i) \rightarrow \sum_{i=1}^{\infty} p_y(\lambda_i) w_i \tag{3}$$

$$f_y/n \rightarrow \sum_{i=1}^{\infty} p_y(\lambda_i) w_i \tag{4}$$

where $\hat{\lambda}_i = \lambda(\hat{\theta}_n, \eta_i)$.

We defer the short proof to the appendix. The major point of this theorem is that both, f_y and the covariate adjusted frequency $\hat{f}_y(\hat{\theta}_n)$, converge to the same object if the model is correct. Hence they are comparable.

3 More complex applications

3.1 A case study on perinatal mortality

In epidemiology and public health, the *standardised mortality ratio* is an important and basic concept of relative risk estimation and frequently used.[5] It also plays an important role in geographical epidemiology and disease mapping as well in disease surveillance (see the recent special issue in this journal edited by Lawson et al.[6]). The standardised mortality ratio is defined as $SMR_i = y_i/e_i$, where y_i is the observed death count in unit i and e_i is expected death count in unit i, where the latter is calculated on the basis of a process called indirect standardisation. Here we assume that the expected death counts are given and refer to the process of indirect standardisation to Woodward.[5] In geographical epidemiology, the question of *equity of health* arises: are all regions under the identical risk of disease occurrence or are some regions exposed to higher and others to lower risk? These question are often discussed in the framework of disease occurrence risk *homogeneity* versus disease occurrence risk *heterogeneity*. Such a question is discussed by Martuzzi and Hills.[7] They investigate the question of heterogeneity in perinatal mortality in the North West Thames Health Region for the 5-year period of 1986 to 1990. The North West Thames Health Region consists out of 515 wards (small geographical units also used in local elections) with approximately 3 million people. There were 2051 perinatal deaths in total in the observational period. Martuzzi and Hills[7] also consider the standardised mortality ratio for relative risk estiamtion. Here it is defined as $SMR_i = y_i/e_i$, where y_i is the observed perinatal death count in ward i and e_i is expected perinatal death count in ward i. The full data set has kindly been provided by Marco Martuzzi and is available online (www.personal.soton.ac.uk/dab1f10/home.htm).

When investigating the question of heterogeneity in perinatal mortality, it is clear that considering only the observed counts y_i is confounded by the effect of the expected cases. Or, in other words, considering only y_i irrespectively of the e_i is meaningless. Figure 5 shows the distribution of the observed cases, but it is not clear what it tells the reader. An obvious alternative would be to

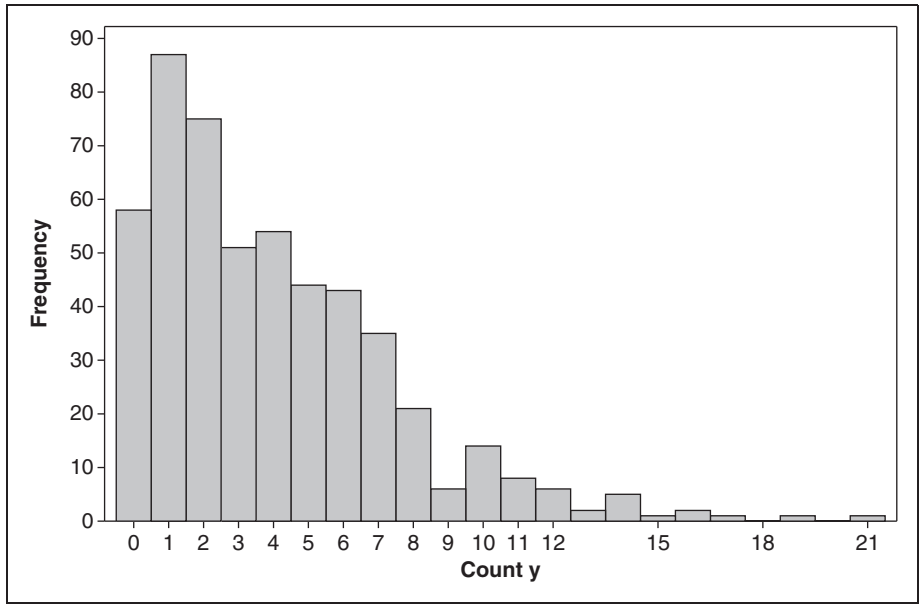

Figure 5. f_y for the perinatal mortality counts in the North West Thames Health Region, 1986–1990.

consider the SMR distribution as provided with Figure 6. However, the SMR distribution has problems as well:

- it is truncated at zero with many values of the SMR being exactly zero which makes modelling of a continuous distribution more difficult,
- it is potentially misleading as it might indicate zero inflation,
- and it ignores in general the count nature of the observed death counts.

The question of heterogeneity in spatial SMR distributions is usually formulated in the following way. Conditional upon the i-th area (here the ward) the observed count is assumed to follow[6,7] the region-specific Poisson model

$$Y_i \sim Po(y|\theta_i e_i) = e^{-\theta_i e_i}(\theta_i e_i)^y / y!,$$

where each ward i has specific mean $\theta_i e_i$. The parameter θ_i is the region-specific theoretical standardised mortality ratio which could be estimated as $\hat{\theta}_i = y_i/e_i$. Rewrite $y_i = e_i\hat{\theta}_i$ and the assumption that y_i arises from a Poisson with parameter $e_i\theta_i$ becomes clear. Various forms of hypotheses on the heterogeneity distribution of θ_i can be formulated and one of the simplest is the hypothesis of homogeneity $\theta_i = \theta$ for all i. An estimate of θ is readily available as $\hat{\theta}_n = \sum_i y_i / \sum_i e_i$. However, this leads to the difficulty of dealing with 515 Poisson models graphically. This can be easily accomplished with the covariate-adjusted frequency plot. In this situation, we have the fitted frequency

$$\hat{f}_y(\hat{\theta}_n) = \sum_{i=1}^{n} Po(y|\hat{\theta}e_i) = \sum_{i=1}^{n} e^{-\hat{\theta}e_i}(\hat{\theta}e_i)^y / y!,$$

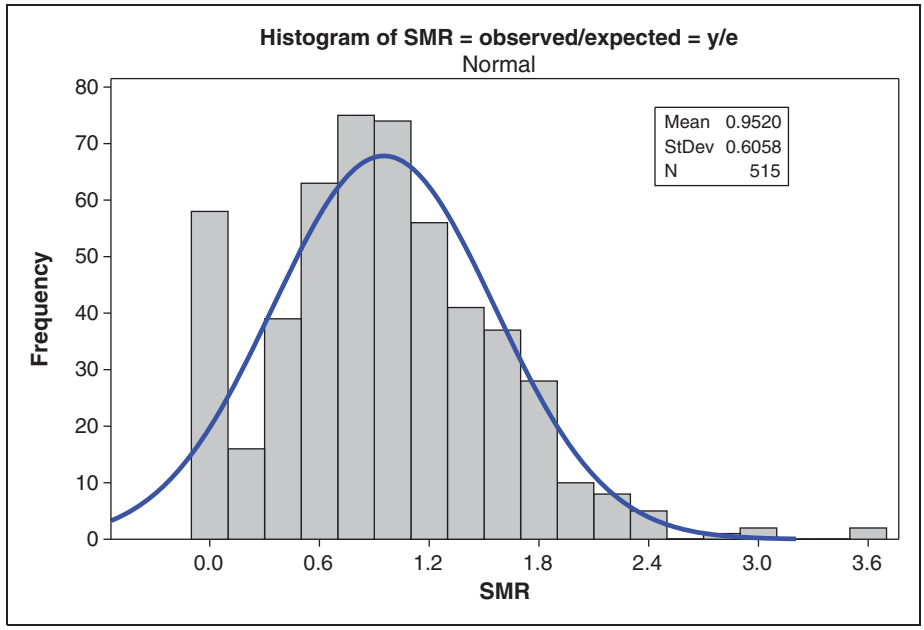

Figure 6. Histogram of the SMR distribution for the perinatal mortality data of the North West Thames Health Region, 1986–1990.
SMR: standardised mortality ratio.

which we call for this situation the expected-cases adjusted frequency plot. Here $\hat{\theta}_n = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} e_i}$ is the maximum likelihood estimate of θ. Figure 7 shows the distribution of the counts of perinatal deaths f_y and the fitted Poisson model $\hat{f}_y(\hat{\theta})$. Also included in Figure 7 for comparison is a finite mixture model allowing for heterogeneity in the parameter θ. There is no evidence of any heterogeneity as also indicated by the BIC values for the two models (Figure 7) with the adjusted frequency plot giving a visual impression of this fact. It also shows that at certain points the goodness-of-fit is not perfect, due to the non-smooth character of the observed data. We further note that the SMR model can be also formulated as a Poisson regression model $Y_i \sim \mu_i$ where $\log \mu_i = \log e_i + \log \theta_i$ and the $\log e_i$ enter as an *offset*. Unobserved heterogeneity could enter as random effects distribution for $\log \theta_i$, but the same principle of the covariate adjusted frequency plot would apply. A benefit of the regression model formulation is that easily observed covariates of the region could be entered into the modelling. We continue the discussion on regression modelling in the next application. An evident alternative to the covariate adjusted frequency plot is considering residual analysis. For example, Cameron and Trivedi[8] consider covariate modelling for count outcomes in detail though model evaluation is focusing more on residual analysis. Pearson residuals have been discussed by many authors including Lindsey,[9] Zelterman[10] or Winkelmann.[11] However, index-plots or Q-Q-plots on the basis of Pearson residuals can be misleading, since even if the model is correct in terms of covariates and distributional assumption the graph might still indicate some deficiencies. Figure 8 shows two kinds of residual plots for the perinatal mortality data of the North West Thames Health Region. The upper panels show the Pearson residuals $r_i = (y_i - e_i)/\sqrt{e_i}$ against i (index plot) and against the expected value e_i. Most observations should lie in the $[-2, 2]$-segment[10]

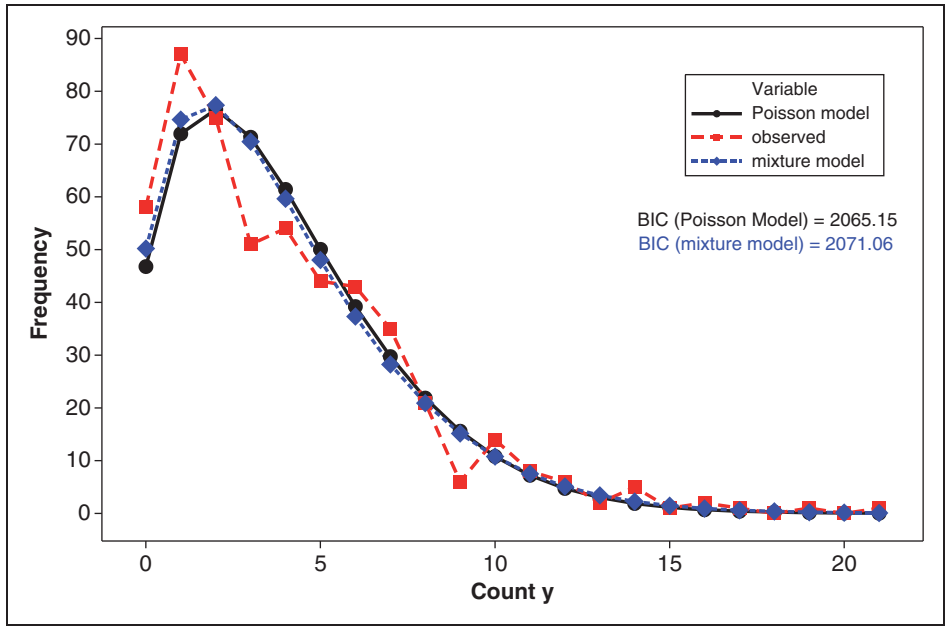

Figure 7. f_y and the expected-cases adjusted frequency plot for the perinatal mortality data of the North West Thames Health Region, 1986–1990.

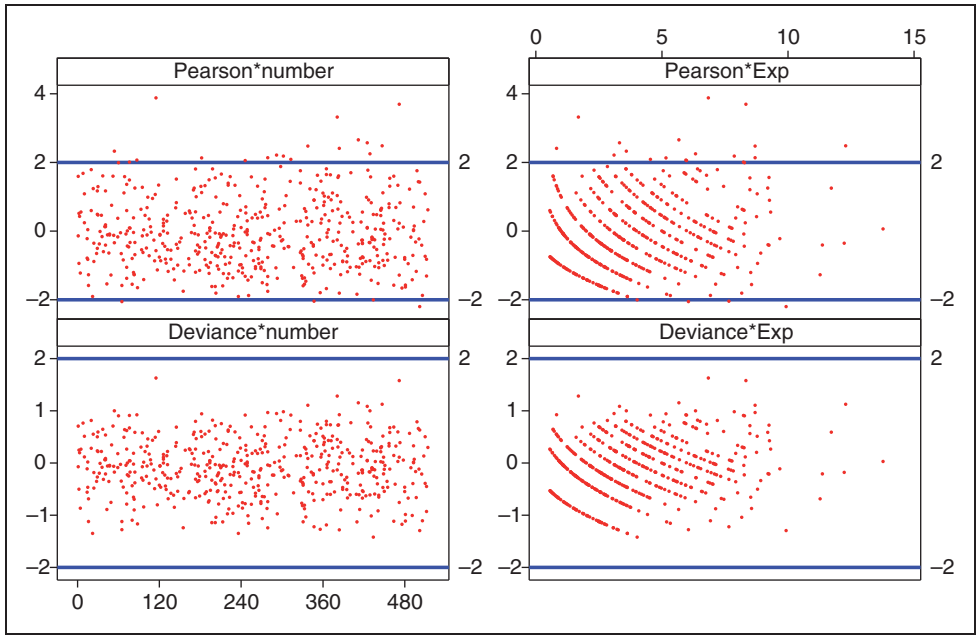

Figure 8. Residual plots for the perinatal mortality data of the North West Thames Health Region, 1986–1990; upper panels are based upon Pearson residuals whereas lower panels are based upon the deviance residuals; left panels are index plots whereas right panels are plots of residuals against expected values.

though a certain percentage can be expected to lie outside this segment. It is also emphasised that it is important to monitor these plots for extreme deviations or particular patterns.[10] Alternatively, it is recommended[10,11] to consider the deviance residuals defined as

$$d_i = \text{sign}(y_i - e_i)\sqrt{2}\sqrt{y_i \log(y_i/e_i) - (y_i - e_i)}.$$

The deviance residuals are expected to have an improved approximations if the model is true. In the lower panels of Figure 8 the deviance residuals are plotted against the ward number and the expected value. All observations now fall into the [−2, 2]-segment, but it appears that lower tail observations are shrunk too much towards the center, reflecting the asymmetric nature of the distribution. Again, there is no evidence for any extreme deviations or pattern. An additional way of investigating goodness-of-fit is the Q-Q-plot as suggested in Lindsey.[9] This procedure is questionable since it screens the residuals for normality which we would not expect when we work with count data. However, at least in the Poisson case with the Poisson parameter becoming large, we can expect the residuals to be approximately normal. Hence we have looked in our perinatal mortality data application at the fit to the normal distribution based on the Pearson and on the deviance residuals. See Figure 9 for details. There is clear evidence that the Pearson residuals show a lack of fit to the normal distribution whereas this is much improved for the deviance residuals. All in all, we see no evidence for a violation of Poisson homogeneity for the risk ratio of perinatal mortality in this health region. Finally, one could argue to work with the standardised mortality ratio directly as it is a continuous quantity. Hence a normal distribution could be valid candidate distribution and

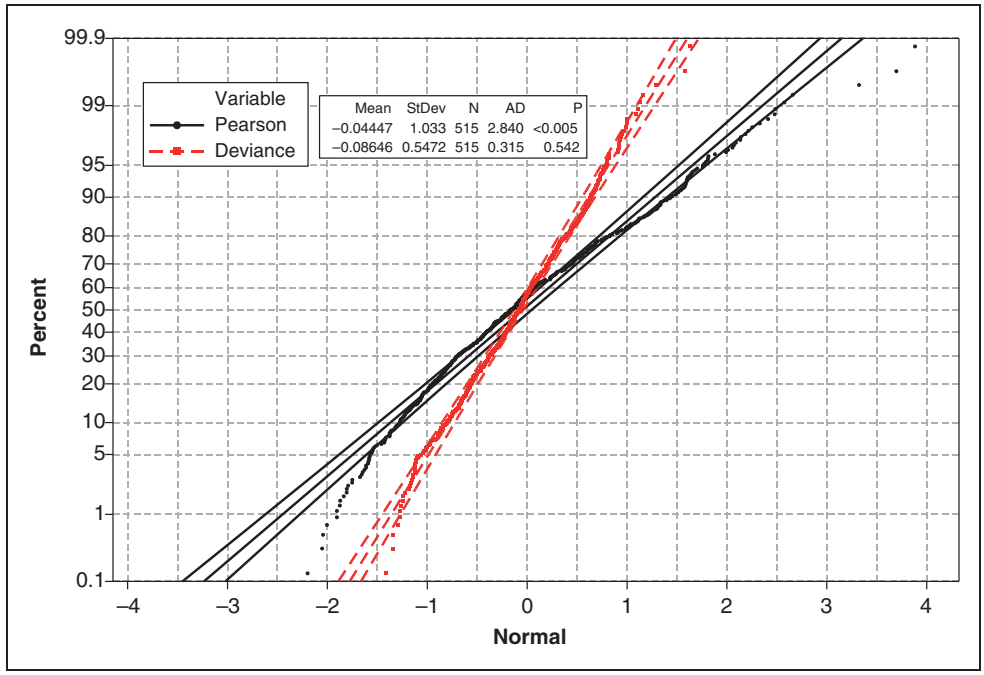

Figure 9. Probability plots for the perinatal mortality data of the North West Thames Health Region, 1986–1990; black dots refer to the Pearson residuals, red squares refer to deviance residuals.

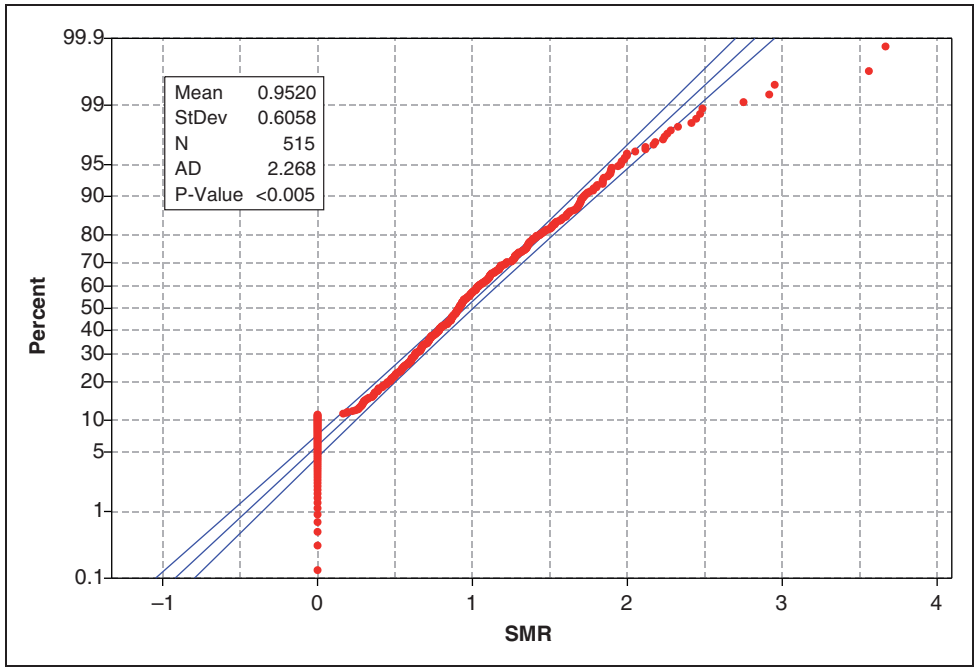

Figure 10. Probability plot for the SMR of the perinatal mortality data of the North West Thames Health Region, 1986–1990.
SMR: standardised mortality ratio.

the Q-Q-plot appropriate to evaluate this. Figure 10 shows this plot for the SMR distribution of the North West Thames Health Region indicating that a normal distribution is not appropriate due to the high amount of zeros caused by the count nature of numerator of the SMR.

3.2 A case study on cerebrovascular mortality in Berlin 1989

In this application, we consider the cerebrovascular mortality in Berlin (West) 1989. The year 1989 is the last year before the fall of the Berlin wall so that the region can be considered as a relative closed population. The data stem from the Berlin Statistical Office and are publicly available. They consist of daily counts of cause-specific mortality (366 days), separately for the male and female population. There are three covariates: $x_1 =$ gender (G) (male $= 1$, female $=0$), $x_2 =$ month of the year (MOY) (Jan $= 1$, Feb $=2,\ldots$, Dec $=12$) and $x_3 =$ day of the month (DOM). The entire data set is available online (www.personal.soton.ac.uk/dab1f10/home.htm) and is partly reproduced here as Table 1.

A natural approach for analysis is to consider a generalised linear model, here a Poisson distribution as the error distribution and a log-link as the link function so that the following Poisson regression model arises

$$Y_i \sim Po(y|\lambda_i) = e^{-\lambda_i}\lambda_i^y/y! \tag{5}$$

$$\lambda_i = \lambda(\beta, \mathbf{x_i}) = \exp(\mathbf{x_i}^T\beta) \tag{6}$$

Table 1. Daily cerebrosvascular mortality in Berlin (West), 1989.

y	0	1	2	3	4	5	6	7
f_y (♀)	28	68	82	84	59	25	12	6
f_y (♂)	1	4	15	31	39	55	54	49
All	29	72	97	115	98	80	66	55
y	8	9	10	11	12	13	14	Total
f_y (♀)	1	1	0	0	0	0	0	366
f_y (♂)	47	31	16	9	8	4	3	366
All	48	32	16	9	8	4	3	732

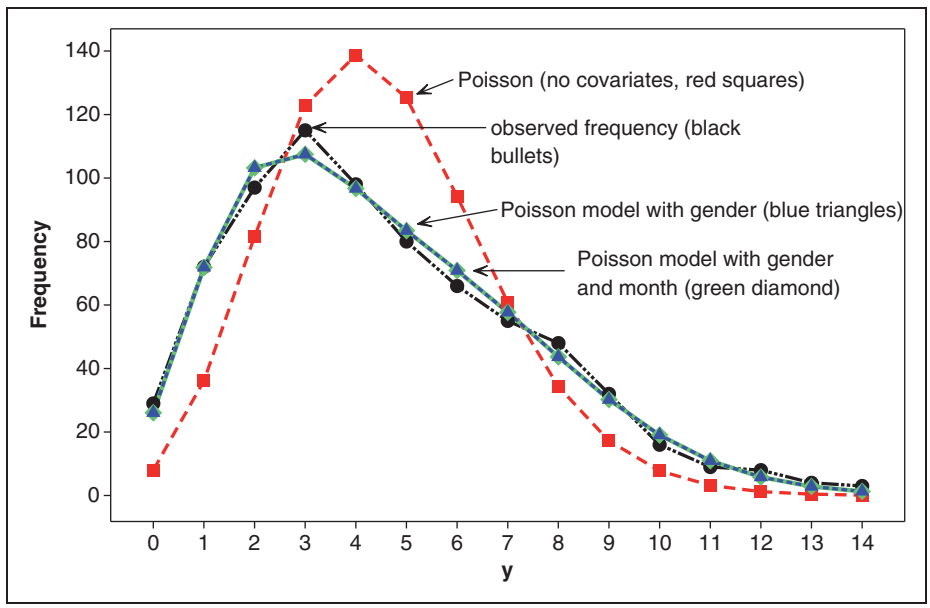

Figure 11. f_y and the covariate adjusted frequency plot for the cerebrosvascular mortality in Berlin (West), 1989; three models are considered: intercept only, including gender, and including gender and month of the year.

where $i = 1, 2, \ldots, 732$. Here $\mathbf{x_i}$ represents all or part of the above-mentioned covariates G, MOY, and DOM.

Fitting models using maximum likelihood provides $\hat{\lambda}_i = \lambda(\hat{\beta}, \mathbf{x_i})$ so that the covariate-adjusted frequency plot

$$f_y(\hat{\beta}) = \sum_{i=1}^{n} Po(y|\hat{\lambda}_i)$$

can be constructed. We look at covariate-adjusted frequency plots for various covariates in Figure 11. Clearly, the intercept-only model does not provide a good fit. The gender difference is important and leads to a considerable improvement of goodness-of-fit. Including other covariates

does not improve the fit to any visual degree. Hence a gender-specific Poisson model appears to be a reasonable model for these mortality data.

4 Discussion

Frequency and fitted frequency plots are provided occasionally, but do not involve covariates (see, for example, Lindsey,[9] p. 133). Modelling count data using covariate information is of wide interest and of great practical value and potential impact. Hence is model diagnostics and goodness-of-fit analysis. The covariate-adjusted frequency plot offers a summarising diagnostic tool. The question arises what the plot offers more than, say, a goodness-of-statistic would provide. We argue that both instruments offer important but different dimensions of diagnostic evidence. The goodness-of-fit statistics offers a numerical value which can be interpreted on a numerical scale whereas the covariate-adjusted frequency plot offers a graphical analysis giving an overall impression of model fit with potential regions in the count data range where the fit is particular good or bad.

In connection with the application 3.1 we have already discussed the alternative of residual diagnostics and plotting, and the various problems associated with it. One of the sources for the problems with Q-Q-plotting lie in the fact that the ordered residuals are plotted against the normal distribution quantiles. It has been observed by Ben and Yohai[12] that these plots can be far away from a straight line even when the model is correct and the distribution is correctly specified. It is then suggested by Ben and Yohai[12] to plot the ordered deviance residual against their theoretical quantile, assuming the model is correct. This makes them more useful for model checking. In a recent paper, Augustin et al.[13] point out that the Ben–Yohai plots are relatively complex and computational expensive in their construction and suggest an alternative approach based on simulation. Augustin et al.[13] also show the validity of the approach and illustrate the method with count data from a cancer registry covering a region in North-East France. As an interesting point they conclude that an analysis based upon normal Q-Q-plots would have erroneously lead to proposing a zero-inflated model, an error that has been avoided using the new (and right) form of Q-Q-plot. This is very similar to our analysis of the perinatal mortality data where we observed a potential for falsely identifying a zero-inflated distribution (Figure 6), in particular when working directly with the SMR distribution, which is quite tempting but leading into serious distributional problems as the probability plot in Figure 10 shows. We can avoid these artifacts using the covariate-adjusted frequency plot which is, of course, also very simple in its construction.

Of course, the covariate adjusted frequency plot is a summarising construction. The adjusted frequency $\hat{f}_y$ averages over all covariate values available in the sample and collects how much they contribute to the fit given the model under consideration. Since it is a summary measure it will only provide a global assessment and will not be able to identify particular observations that experience a poor fit. Here we believe that residual analysis is more appropriate. We view the covariate-adjusted frequency plot as useful tool in a supplementary diagnostic role for count data modelling with possibly complex covariate structures.

Acknowledgement

The authors are grateful to the editor and two anonymous reviewers for their helpful comments.

Funding

This work was funded by the *German Research Foundation* (GZ: Ho1286/7-1).

References

1. Pawitan Y. *In all likelihood. Statistical modelling and inference using likelihood*. Oxford: Clarendon Press, 2001.
2. Maritz JS and Lwin T. *Empirical Bayes methods*, 2nd edn. London: Chapman & Hall, 1989.
3. Goldberg D. *The detection of psychiatric illness by questionnaire*. Oxford: Oxford University Press, 1972.
4. Der G and Everitt BS. *Statistical analysis of medical data using SAS*. Boca Raton: Chapman & Hall/CRC, 2006.
5. Woodward M. *Epidemiology: study design and data analysis*. Boca Raton: Chapman & Hall/CRC, 1999.
6. Lawson AB, Cocchi D, Banerjee S, et al. Special issue GEOMED 2011. *Stat Meth Med Res* 2011; **21**: 431.
7. Martuzzi M and Hills M. Estimating the degree of heterogeneity between event rates using likelihood. *Am J Epidemiol* 1995; **141**: 369–374.
8. Cameron AC and Trivedi PK. *Regression analysis of count data*. Cambridge: Cambridge University Press, 1996.
9. Lindsey JK. *Modelling frequency and count data*. Oxford: Clarendon Press, 1995.
10. Zelterman D. *Models for discrete data*. Oxford: Oxford University Press, 2006.
11. Winkelmann R. *Econometric analysis of count data*. New York: Springer, 2003.
12. Ben MG and Yohai V. Quantile-quantile plot for deviance residuals in the generalized linear model. *J Comput Graph Stat* 2004; **13**: 36–47.
13. Augustin NH, Sauleau EA and Wood SN. On quantile-quantile plots for generalized linear models. *Comput Stat Data Anal* 2012; **56**: 2404–2409.
14. Shao J. *Mathematical statistics*. New York: Springer, 2003.

Appendix

Proof of Theorem 1: Let again be $\lambda_i = \lambda(\theta, \eta_i)$. We have that

$$s_{n,y}(\theta) = \frac{1}{n}\sum_{i=1}^{n} p_y(\lambda_i) = \sum_{i=1}^{n} p_y(\lambda_i)\hat{w}_i,$$

where $\hat{w}_i = \frac{\#\{\eta_j \,|\eta_j = \eta_i\}}{n}$. Since the empirical proportion of sampled η_i converges to the theoretical proportion of sampling η_i, in other words $\hat{w}_i \to w_i$, it follows that

$$s_{n,y}(\theta) = \frac{1}{n}\sum_{i=1}^{n} p_y(\lambda_i) \to_{n\to\infty} \sum_{j=1}^{\infty} p_y(\lambda_j)w_j = s_y(\theta) < \infty.$$

It remains to show that

$$s_{n,y}(\hat{\theta}_n) \to_{n\to\infty} s_y(\theta).$$

This follows from the continuity of $\lambda(., \eta_i)$ for all i and the fact that $\hat{\theta}_n \to_{n\to\infty} \theta$ (see, for example, Shao,[14] p. 59) and ends the proof.