Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2011)
60, *Part* 5, *pp.* 723–741

Capture–recapture estimation by means of empirical Bayesian smoothing with an application to the geographical distribution of hidden scrapie in Great Britain

Dankmar Böhning,

University of Reading, UK

Ronny Kuhnert

Robert Koch Institute, Berlin, Germany

and Victor Del Rio Vilas

Department for Environment, Food and Rural Affairs, London, UK

[Received August 2010. Revised April 2011]

Summary. The paper discusses population size estimation on the basis of a frequency distribution of zero-truncated counts and is motivated by a study on the geographical distribution of hidden scrapie in Great Britain. Aggregation of scrapie cases is considered at the county level and results in sparse zero-truncated count distributions which make the application of conventional capture–recapture procedures for estimating the hidden part of the scrapie-affected population difficult. We suggest a smoothed generalization of Zelterman's estimator of population size which overcomes the overestimation bias of the conventional Zelterman estimator and instead produces a lower bound, which is typically larger than Chao's lower bound estimator. The estimator uses an empirical Bayes approach with various choices for the prior distribution including a parametric choice of the gamma distribution as well as various non-parametric distributions. A simulation study investigates the performance of the new estimators, and also in comparison with conventional estimators. The empirical Bayes estimator with a non-parametric mixture model as prior performs well and the boundary problem of the conventional non-parametric discrete mixture model estimator leading to spurious population size is avoided. In the application to hidden scrapie in Great Britain the new estimators lead to maps of scrapie of observed–hidden ratios as well as completeness of the current surveillance system.

Keywords: Capture–recapture; Empirical Bayes methods; Geographical analysis; Non-parametric mixture model

1. Introduction

For integer N, we consider a sample of counts $x_1, x_2, \ldots, x_N \in \{0, 1, 2, \ldots, \}$ arising from a count random variable X having a mixture probability density function

$$p_x = \int_0^\infty p(x|\lambda)\, q(\lambda)\, \mathrm{d}\lambda \tag{1}$$

with unspecified mixing density $q(\lambda)$ and a mixture kernel $p(x|\lambda)$ which needs to be specified.

Address for correspondence: Dankmar Böhning, Department of Mathematics and Statistics, School of Mathematical and Physical Sciences, University of Reading, Whiteknights, Reading, RG6 6BX, UK.
E-mail: d.a.w.bohning@reading.ac.uk

 0035–9254/11/60723

In this paper, a typical choice for the mixture kernel is the Poisson kernel $p(x|\lambda) = \mathrm{Po}(x|\lambda) = \exp(-\lambda)\lambda^x/x!$ though other choices are possible as well. Whenever $x_i = 0$ unit i remains unobserved, so only a zero-truncated sample of size $n = \Sigma_{j=1}^{m} f_j$ is observed, where f_j is the frequency of counts with value $x = j$ and m is the largest observed count. Hence, f_0 and consequently N are unknown. The purpose is to find an estimate of the size N. Since frequently the count variable X represents repeated identifications of an individual in an observational period, the problem at hand is a special form of the capture–recapture problem (see Bunge and Fitzpatrick (1993) and Collins and Wilson (1992) for a review on the topic).

The sample of counts $x_1, x_2, \ldots, x_N$ can occur in several ways. A target population which might be difficult to count consists of N units. This population might be a wildlife population, a population of homeless people, drug addicts, software errors or animals with a specific disease. Furthermore, let an identification device (a trap, a register or a screening test) be available that identifies unit i at occasion t where $t = 1, \ldots, T$. Let the binary result be x_{it} where $x_{it} = 1$ means that unit i has been identified at occasion t and $x_{it} = 0$ means that unit i has not been identified at occasion t. The indicators x_{it} might be observed or not, but it is assumed that $x_i = \Sigma_{t=1}^{T} x_{it}$ is observed if at least one $x_{it} > 0$ for $t = 1, \ldots, T$. Only if $x_{i1} = x_{i2} = \ldots = x_{iT} = 0$ and, consequently $x_i = 0$, does the unit i remain *unobserved*. In this kind of situation the *clustering* occurs by repeated identifications of the same unit.

In another setting, which is also the basis for this work, the clustering occurs by means of a grouping variable such as herds, holdings, households or villages. In this case, x_{it} denotes whether the tth element in cluster i is identified ($x_{it} = 1$) or not ($x_{it} = 0$). In the example that is given in Section 2 the clusters are holdings of sheep and x_{it} informs about the disease status of the tth animal in holding i. Note that $x_i = \Sigma_t x_{it}$ is observed only if it is positive. In other examples the cluster corresponds to villages or households; one of the earliest applications of this kind is the cholera outbreak in a community in India that was studied by McKendrick (1926) in which the cluster corresponds to households in a village. A more recent example involves the occurrence of cholera in rural East Pakistan where the cluster structure consists of villages (see also Mosley *et al.* (1972)).

The paper is organized as follows. Section 2 introduces the data on scrapie in Great Britain. In Section 3 we review some of the existing approaches in the capture–recapture methodology for the setting of interest. Section 4 describes the development of a new set of empirical Bayes estimators which are then further evaluated by means of a simulation study. The application of the empirical Bayes estimator to the spatial data on scrapie in Great Britain, including the development of maps at county level of completeness and the observed–hidden ratio, ends the paper in Section 5.

The data that are analysed in the paper can be obtained from

`http://www.blackwellpublishing.com/rss`

2. Scrapie data in Great Britain

We now consider as a specific case-study the spatial distribution of scrapie in Great Britain. Classical scrapie, which is a fatal neurological disease of small ruminants, is endemic in Great Britain (see Del Rio Vilas *et al.* (2006) for more details). There is ample evidence to support the occurrence of under-reporting affecting the clinical notification of scrapie cases (Hoinville *et al.*, 2000; Del Rio Vilas *et al.*, 2005; Böhning and Del Rio Vilas, 2008). Although not established to date, there is reason to believe that, reflecting population and surveillance-related heterogeneities, under-reporting presents an uneven distribution across Great Britain. The spatial analysis

that is presented in what follows uses county-specific disease data from the scrapie notifications database (SND) (see Del Rio Vilas *et al.* (2006) for more details): more specifically the number of confirmed clinical cases per scrapie-affected holding. Table 1 shows the frequency f_x of the number of holdings with confirmed clinical cases x for $x = 1, 2, 3, \ldots$ by county. The entire data set is provided in the on-line supporting information. Evidently, there is a considerable range in the number of scrapie-affected holdings per county, ranging from counties with only one affected holding to counties with a large number of affected holdings, the largest number occurring in county 37 with 75 affected holdings.

Our main interest in the following analysis is to investigate the performance of the SND surveillance stream as measured in the observed–hidden ratio (the larger the ratio the better the system) as well as in the *completeness rate*, which is defined as the proportion of observed affected holdings among observed and hidden scrapie-affected holdings. If the case count per holding is collapsed over all counties we find the distribution as given at the bottom of Table 1. With $f_1 = 298$, $f_2 = 89$ and $f_3 = 42$ most of the distribution is concentrated on counts of one, two and three cases with the largest count occurring at 29.

3. Background on capture–recapture estimation

Before we go into the details of the suggested novel approach we give a brief review of the existing capture–recapture methodology for the setting of interest.

3.1. Heterogeneity

The importance of the mixture $p_x = \int_0^\infty p(x|\lambda)\, q(\lambda)\, \mathrm{d}\lambda$ can be seen in the fact that it is a natural model for the population heterogeneity. There appears to be consensus (see for example Pledger (2005) for the discrete mixture model approach and Dorazio and Royle (2005) for the continuous mixture model approach) that a simple model $p(x|\lambda)$ is not sufficiently flexible to capture the variation in the recapture probability for the different members of most real life populations. There has also recently been a debate on the identifiability of the binomial mixture model (see Link (2003, 2006) and Holzmann *et al.* (2006)). Furthermore, using the non-parametric maximum likelihood estimator (NPMLE) $\hat{q}(\lambda)$ of the mixing density $q(\lambda)$ in constructing an estimate of the population size

$$\hat{N} = n \Big/ \left\{ 1 - \int_0^\infty \exp(-\lambda)\, \hat{q}(\lambda)\, \mathrm{d}\lambda \right\}$$

leads to the *boundary problem*. This results in often unrealistic high values for the estimate of the population size (Wang and Lindsay, 2005, 2008). Hence, renewed interest has re-emerged in the lower bound approach for population size estimation suggested by Chao (1987). In this approach neither is there need to specify a mixing distribution, nor is there need to estimate it. In this sense it is completely non-parametric. To give some details of the lower bound approach consider the Poisson mixture kernel $\exp(-\lambda)\lambda^x/x!$. It follows from the Cauchy–Schwarz inequality that

$$\left\{ \int_0^\infty \exp(-\lambda)\lambda\, q(\lambda)\, \mathrm{d}\lambda \right\}^2 \leqslant \int_0^\infty \exp(-\lambda)\, q(\lambda)\, \mathrm{d}\lambda \int_0^\infty \exp(-\lambda)\lambda^2\, q(\lambda)\, \mathrm{d}\lambda,$$

or, equivalently, $p_1^2 \leqslant p_0(2p_2)$. Replacing the theoretical probabilities p_j by their sample estimates f_j/N for $j = 0, 1, 2$, the Chao lower bound estimate $f_1^2/2f_2$ for f_0 follows (see Chao (1987, 1989)) since the unknown denominator N cancels out. The estimate for the population size N is $\hat{N}_C = n + f_1^2/2f_2$. Since the Chao estimator uses only frequencies with counts of 1 and 2, a

Table 1. Distribution of confirmed scrapie-affected holdings from the SND 2002–2006 by county

County	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10+}	n
1	2	1	1	0	0	0	0	0	0	0	4
2	1	1	1	0	1	0	0	0	0	0	4
3	1	0	0	0	0	0	0	0	0	0	1
4	1	0	0	0	0	0	0	0	0	0	1
5	2	0	0	0	1	0	0	1	0	3	7
6	4	1	0	1	0	1	0	1	0	3	11
7	12	1	0	2	3	0	0	0	0	1	19
8	7	2	2	0	0	0	0	0	0	0	11
9	25	8	5	1	1	1	2	0	0	2	45
10	4	1	0	0	0	0	0	0	0	0	5
11	1	0	0	0	0	0	0	0	0	0	1
12	0	0	1	0	0	0	0	0	0	0	1
13	2	0	0	1	0	0	0	0	0	0	3
14	1	2	0	0	0	0	0	0	0	0	3
15	0	1	0	0	0	0	0	0	0	0	1
16	5	2	1	1	0	0	0	1	0	0	10
17	1	0	0	0	0	0	0	0	0	0	1
18	5	0	0	0	0	0	0	0	0	0	5
19	1	1	0	0	0	0	0	0	0	0	2
20	1	0	0	0	0	0	0	0	0	0	1
21	2	1	1	0	0	0	0	0	0	0	4
22	3	3	0	0	1	0	1	0	1	0	9
23	5	0	1	0	0	0	0	0	0	1	7
24	2	0	0	0	0	0	0	0	0	0	2
25	1	1	0	0	0	0	0	0	1	0	3
26	6	2	0	0	0	0	0	0	0	0	8
27	5	1	0	0	1	0	0	0	0	0	7
28	1	0	0	0	0	0	0	0	0	2	3
29	2	0	1	0	1	0	0	0	0	0	4
30	1	0	0	0	0	0	0	0	0	0	1
31	2	1	0	0	0	0	0	0	0	0	3
32	1	0	0	0	0	0	0	0	0	0	1
33	1	0	1	0	0	0	0	0	0	0	2
34	14	10	3	1	3	0	2	1	0	3	37
35	2	0	0	0	0	0	0	0	0	0	2
36	2	0	0	0	0	0	0	0	0	0	2
37	51	11	5	5	0	1	1	1	0	0	75
38	6	3	1	0	0	0	0	0	1	0	11
39	24	9	1	1	3	1	2	1	0	2	44
40	6	4	4	1	2	1	2	1	0	4	25
41	3	5	2	0	1	0	0	0	0	0	11
42	6	1	3	1	0	0	0	0	0	0	11
43	1	1	0	0	0	0	0	0	0	0	2
44	4	0	1	0	0	1	0	0	0	1	7
45	1	0	0	0	0	0	0	0	0	0	1
46	3	0	0	0	0	0	0	0	0	0	3
47	1	0	0	0	1	0	0	0	0	0	2
48	0	0	1	0	0	0	0	0	0	0	1
49	1	0	0	0	0	0	0	0	0	0	1
50	2	0	0	0	0	0	0	0	0	0	2
51	1	0	0	0	0	0	0	0	0	0	1
52	13	2	1	0	0	0	1	0	0	0	17
53	0	0	0	0	1	0	0	0	0	0	1
54	1	0	0	0	0	0	0	0	0	0	1
55	47	10	5	2	0	1	0	0	0	0	65
56	1	3	0	0	0	0	0	0	0	0	4
All	298	89	42	17	20	7	11	7	3	22	516

truncated sample consisting only of counts of 1s and 2s might be considered. This truncated sample leads to a binomial log-likelihood $f_1 \log(p_1) + f_2 \log(p_2)$ which is uniquely maximized for $\hat{p}_2 = 1 - \hat{p}_1 = f_2/(f_1 + f_2)$. Since $p_2 = \lambda/(\lambda + 2)$ and $p_1 = 2/(\lambda + 2)$ is a Poisson kernel that truncates all counts except 1s and 2s, the estimate $\hat{\lambda} = 2f_2/f_1$ for the Poisson parameter λ that was suggested by Zelterman (1988) arises. In the approach of Zelterman the homogeneous Poisson model serves only as a working model and it was suggested by Zelterman that the estimate $\hat{N}_Z = n/(1 - \hat{p}_0) = n/\{1 - \exp(-\hat{\lambda})\}$ is more robust against misspecifications of the Poisson model than the usual maximum likelihood estimate.

3.2. A reanalysis of Zelterman estimation

We are interested in developing a generalization of the Zelterman estimator. Consider the Horvitz–Thompson-type estimate of the population size that was suggested by Zelterman (1988):

$$\hat{N}_Z = \frac{n}{1 - \exp(-2f_2/f_1)}. \tag{2}$$

Although estimator (2) is popular among practitioners it has two disadvantages:

(a) it uses only the frequencies f_1 and f_2 and ignores f_3–f_m;
(b) it can experience *severe overestimation bias*.

The first issue is evident and results in large variance. The second issue is less evident but becomes clear in what follows where we consider a discrete version of equation (1), namely a two-component mixture $p_x = q\,p(x|\lambda) + (1 - q)\,p(x|\mu)$. Note that we can think of the second component $p(x|\mu)$ as the contaminating part in this model. These contaminated models have a tradition in robust statistics (see Hampel *et al.* (1986)) since the mean $E(X) = q\lambda + (1 - q)\mu$ is sensitive to contaminating observations which are generated by a large value of μ. Note that this model generates positive frequencies of counts of 0s, 1s and 2s even for small N as long as q is bounded away from 0 and λ in a reasonable range (0.1–2, say). Under this contamination model the Zelterman estimator can experience infinite overestimation bias if μ becomes large.

Theorem 1. Let $p_x = q\,p(x|\lambda) + (1 - q)\,p(x|\mu)$ be a discrete, two-component mixture with $p(x|\theta) = \mathrm{Po}(x|\theta)$ being the Poisson kernel and $0 < q < 1$. Then,

$$E(\hat{N}_Z) \approx$$

$$N[1 - \{q\exp(-\lambda) + (1 - q)\exp(-\mu)\}] \Big/ \left[1 - \exp\left\{ - \frac{q\exp(-\lambda)\lambda^2 + (1 - q)\exp(-\mu)\mu^2}{q\exp(-\lambda)\lambda + (1 - q)\exp(-\mu)\mu} \right\}\right]$$

$$\to_{\mu \to \infty} N \frac{1 - q\exp(-\lambda)}{1 - \exp(-\lambda)} \geqslant N.$$

Theorem 1 is proved by replacing sample frequencies by their theoretical values. Note that the biasing factor $\{1 - q\exp(-\lambda)\}/\{1 - \exp(-\lambda)\}$ can become arbitrarily large since it is a monotone decreasing function of q and λ (see also Fig. 1). But even for realistic values of q and λ the factor can be considerably larger than 1. For example if $q = 0.5$ and $\lambda \leqslant 0.4$ the factor is larger than 2, so the Zelterman estimate would overestimate severely. The question arises about what is the source of this overestimation bias. We approach this question in the next theorem which states that the Zelterman estimator uses the *wrong* expected value in predicting f_0.

Suppose that all counts are truncated except counts of 1 and 2. The associated truncated probabilities are

$$p_1 = \exp(-\lambda)\lambda / \{\exp(-\lambda)\lambda + \exp(-\lambda)\lambda^2/2\} = 2/(\lambda+2)$$

and

$$p_2 = \frac{\exp(-\lambda)\lambda^2/2}{\exp(-\lambda)\lambda + \exp(-\lambda)\lambda^2/2} = \frac{\lambda}{\lambda+2}.$$

Hence, the associated likelihood is a *binomial* likelihood $L = p_1^{f_1} p_2^{f_2}$ which is maximized for $\hat{p}_2 = f_2/(f_1+f_2)$ and the maximum likelihood estimator is found from the invariance principle for maximum likelihood estimators as $\hat{\lambda} = 2\hat{p}_2/(1-\hat{p}_2)$.

Theorem 2.

(a) Let $\log\{L(\lambda)\} = f_1 \log(p_1) + f_2 \log(p_2)$ with $p_2 = \lambda/(\lambda+2)$ and $p_1 = 1 - p_2$. Then $\log\{L(\lambda)\}$ is maximized for $\hat{\lambda} = 2f_2/f_1$.
(b) $E(f_0 | f_1, f_2; \hat{\lambda}) = f_1^2/2f_2$, for $\hat{\lambda} = 2f_2/f_1$ and $n + E(f_0 | f_1, f_2; \hat{\lambda}) = n + {f_1}^2/2f_2 = N_C$.

A proof of theorem 2 is provided in Appendix A. Theorem 2 establishes a close connection between the approach by Zelterman and Chao's estimator. It shows that Zelterman's estimator of the Poisson parameter λ arises when all counts are truncated except counts of 1s and 2s and when the resulting likelihood is maximized. If the correct prediction for f_0 is used, namely the conditional expectation for the truncated Poisson model, the Chao estimator arises. Hence the strong overestimation of the original Zelterman estimator stems from using a *wrong* conditional expectation.

3.3. Comparing some conventional estimators in a simulation

Before we continue developing the generalized adjusted version of the Zelterman estimator, we consider the performance of Chao and Zelterman estimators in a small simulation study. In the case of a homogeneous Poisson model the maximum likelihood estimate is found by maximizing the likelihood of zero-truncated Poisson observations in λ:

$$\prod_{j=1}^{m} \left(\frac{p_j}{1-p_0} \right)^{f_j} = \prod_{j=1}^{m} \left\{ \frac{1}{1-\exp(-\lambda)} \exp(-\lambda) \frac{\lambda^j}{j!} \right\}^{f_j},$$

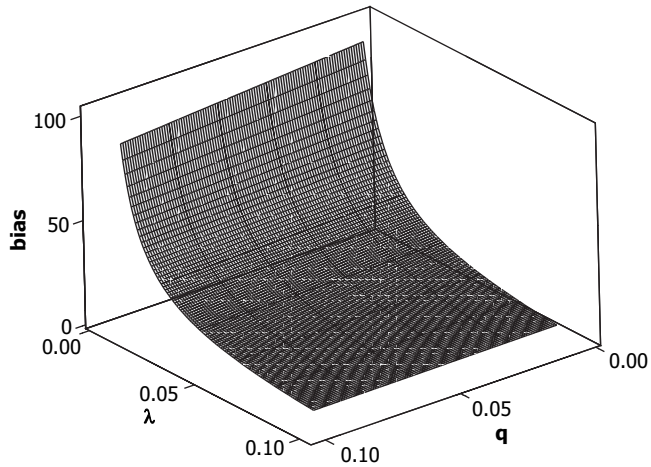

Fig. 1. Biasing factor $\{1 - q\exp(-\lambda)\}/\{1 - \exp(-\lambda)\}$ as a function of q and λ for $\mu \to \infty$

or, equivalently, in solving the following equation in $\hat{N}_{\text{hom}}$:

$$\hat{N}_{\text{hom}} = n\left\{1 - \exp\left(-\frac{S}{\hat{N}_{\text{hom}}}\right)\right\}^{-1},$$

where $S = \Sigma_{x=1}^{m} x f_x$. We must maximize the zero-truncated Poisson mixture likelihood in Q to find the non-parametric maximum likelihood of the mixing distribution

$$L(Q) = \prod_{j=1}^{m}\left(\frac{p_j}{1-p_0}\right)^{f_j} = \prod_{j=1}^{m}\left\{\sum_{l=1}^{k}\frac{\text{Po}(j|\lambda_l)q_l}{1-\sum_i \exp(-\lambda_i)q_i}\right\}^{f_j}$$

where Q is the discrete mixing distribution giving k weights q_j to Poisson parameters λ_j:

$$Q = \begin{pmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_k \\ q_1 & q_2 & \ldots & q_k \end{pmatrix}.$$

Note that we must maximize $L(Q)$ in terms of $\lambda_1, \ldots, \lambda_k$ and $q_1, \ldots, q_k$ but also in k to find the NPMLE. The NPMLE maximizes the likelihood globally and has a finite number of components (Lindsay, 1983). Maximizing the likelihood is typically done in a stepwise manner by fixing k to be $1, 2, 3, \ldots,$ and conditionally on k using a gradient-function-modified version of the EM algorithm for finding the maximum likelihood estimate. For details see Böhning (2003) and Böhning and Kuhnert (2006). If k is fixed we denote the associated maximum likelihood estimator by NPMLE(k). Only a finite number of mixture models in the sequence $k = 1, 2, \ldots, k^*$ need to be considered since for some $k = k^*$ the global maximum will be achieved. The latter case we shall simply denote with NPMLE. Occasionally, we might be interested in comparing mixture models with different numbers of components k by means of the Bayesian information criterion BIC defined as $-\log\{L(Q)\} + (2k-1)\log(n)$. After the NPMLE $\hat{Q}$ of Q has been identified, we can define

$$\hat{N}_{\text{NPMLE}(k)} = \frac{n}{1 - \sum_{j=1}^{k}\exp(-\hat{\lambda}_j)\hat{q}_j}, \tag{3}$$

where NPMLE(k) denotes that the maximum likelihood estimator of the mixing distribution with k components is used. If the global maximum likelihood estimator is used we simply write $\hat{N}_{\text{NPMLE}}$.

To illustrate the performance of these estimators we consider the following simulation experiments. Samples of counts $X_1, \ldots, X_N$ were drawn from a two-component mixture of Poisson densities: $X \sim 0.5\,\text{Po}(1) + 0.5\,\text{Po}(\lambda)$, evidently with equal weights $q_1 = q_2 = 0.5$. The population size was set to $N = 100$ and 10000 replications were used. Ignoring zero counts the estimators of Chao $\hat{N}_{\text{C}} = n + f_1^2/2f_2$ and Zelterman $\hat{N}_{\text{Z}} = n/\{1 - \exp(-2f_2/f_1)\}$ were determined as well as the maximum likelihood estimator under homogeneity $\hat{N}_{\text{hom}}$ and the NPMLE under heterogeneity $\hat{N}_{\text{NPMLE}}$. The results can be found in Table 2. When heterogeneity increases, the Zelterman estimator overestimates whereas the maximum likelihood estimator under homogeneity underestimates—both as expected. The Chao lower bound estimator does well under heterogeneity—again as expected. Most dominant in Table 2 is the drastic failure of the NPMLE which leads to spurious overestimate values.

4. New empirical Bayes estimator of population size

Although it is clear that $2f_2/f_1$ estimates the Poisson parameter in the case that $p_x = \text{Po}(x|\lambda)$, it

Table 2. Simulation using $X \sim 0.5\,\mathrm{Po}(1) + 0.5\,\mathrm{Po}(\lambda)$ and $N = 100$†

λ	*Estimator*	*Mean*	*SD*	*RMSE*
1	MLE-hom	102	13	13
	NPMLE	484	12098	20028
	Chao	104	19	19
	Zelterman	105	21	22
	EB-NPMLE	105	15	15
	EB-Robbins	108	21	22
2	MLE-hom	94	7	9
	NPMLE	4599	35	21328
	Chao	99	12	12
	Zelterman	101	16	16
	EB-NPMLE	98	8	9
	EB-Robbins	102	12	12
3	MLE-hom	88	5	13
	NPMLE	12517	52425	23955
	Chao	97	10	11
	Zelterman	102	15	16
	EB-NPMLE	93	7	10
	EB-Robbins	96	9	10
4	MLE-hom	85	4	16
	NPMLE	11715	54501	23114
	Chao	97	10	10
	Zelterman	108	20	20
	EB-NPMLE	92	7	11
	EB-Robbins	95	9	10
5	MLE-hom	84	4	17
	NPMLE	4657	33069	17373
	Chao	98	10	10
	Zelterman	115	23	27
	EB-NPMLE	92	8	11
	EB-Robbins	95	9	10

†Provided are estimates of $E(\hat{N})$, $\mathrm{var}(\hat{N})^{1/2}$ and $\{E(\hat{N} - N)^2\}^{1/2}$ as the mean, standard deviation SD and root-mean-squared error RMSE.

is not clear what it estimates when there is a mixing distribution instead of Poisson homogeneity. Here, a Bayesian perspective is helpful. We think of the mixing distribution $q(\lambda)$ as a prior distribution on λ so that

$$E(\lambda|x) = \int_0^\infty \lambda \frac{\mathrm{Po}(x|\lambda)\, q(\lambda)}{\int_0^\infty \mathrm{Po}(x|\theta)\, q(\theta)\, \mathrm{d}\theta} \mathrm{d}\lambda \tag{4}$$

is the *posterior mean* with respect to the prior $q(\lambda)$ and Poisson likelihood for observation x. Note that equation (4) can be further simplified to

$$\lambda_x = E(\lambda|x) = \frac{\int_0^\infty \lambda\, \mathrm{Po}(x|\lambda)\, q(\lambda)\, \mathrm{d}\lambda}{\int_0^\infty \mathrm{Po}(x|\lambda)\, q(\lambda)\, \mathrm{d}\lambda}$$

$$= (x+1)\frac{\int_0^\infty \mathrm{Po}(x+1|\lambda)\, q(\lambda)\, \mathrm{d}\lambda}{\int_0^\infty \mathrm{Po}(x|\lambda)\, q(\lambda)\, \mathrm{d}\lambda} = (x+1)\frac{p_{x+1}}{p_x},$$

where p_x is the marginal density (1). We note here in passing the similarity of equation (4) to Good–Turing frequency estimation (Good, 1953) which is commonly used in linguistics and text analysis (Gale and Sampson, 1995). Before we continue on the ways to estimate the ratio of marginals we point out an important monotonicity property.

Theorem 3.

$$\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_m.$$

A proof of theorem 3 is found in Appendix A. Theorem 3 has an important application. Since under heterogeneity we have that $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_m$, we expect that the graph $x \to \hat{\lambda}_x = (x+1)\, f_{x+1}/f_x$ shows a monotone increasing pattern if heterogeneity is present. Hence we can develop a *diagnostic device* for the presence of heterogeneity by plotting $(x+1)\, f_{x+1}/f_x$ against x, which we call the *ratio plot*. The ratio plot for the SND data for the years 2002–2006 is presented in Fig. 2. There is clear evidence for a monotone increasing trend; hence a mixture model coping with the presence of heterogeneity appears appropriate.

Since $1/\{1-\exp(-\lambda)\}$ is monotone non-increasing in λ we have the following corollary which we state without further proof.

Corollary 1.

$$\sum_{x=1}^{m} \frac{f_x}{1-\exp(-\lambda_x)} \leqslant \frac{\sum_{x=1}^{m} f_x}{1-\exp(-\lambda_1)} = \frac{n}{1-\exp(-2p_2/p_1)}. \tag{5}$$

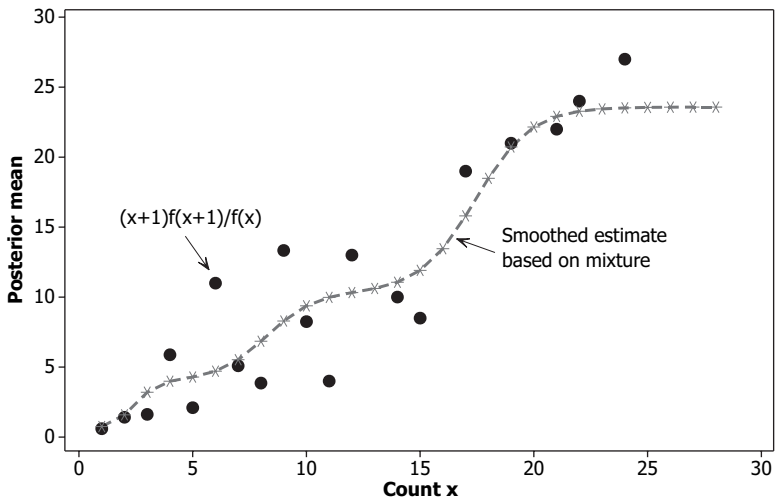

Fig. 2. Ratio plot for the SND data 2002–2006, unstratified by county, for the Robbins estimate of the posterior mean as well as the discrete-mixture- (four components) based EB estimate of the posterior mean

Note that the Zelterman estimator occurs on the right-hand side of expression (5) if p_2/p_1 is replaced by its sample version f_2/f_1. Hence we expect that the overestimation bias of the Zelterman estimate is reduced if λ_x on the left-hand side of expression (5) is appropriately estimated. Furthermore, if

$$\frac{f_1}{1-\exp(-\lambda_1)} = f_1 + \frac{f_1}{\exp(\lambda_1)-1},$$

the first element in the sum on the left-hand side of expression (5), is replaced by its first-order Taylor series expansion $f_1 + f_1/\lambda_1$ and again $\lambda_1 = 2p_2/p_1$ estimated by $2f_2/f_1$, we find that

$$f_1 + \frac{f_1}{\exp(\hat{\lambda}_1)-1} \approx f_1 + \frac{f_1}{\hat{\lambda}_1} = f_1 + \frac{f_1^2}{2f_2},$$

where $f_1^2/2f_2$ is the lower bound estimator of Chao (1987, 1989) for f_0. Hence we expect that the left-hand side of expression (5) provides an improved lower bound estimator if λ_x is estimated appropriately since

$$\sum_{x=2}^{m} \frac{f_x}{1-\exp(-\lambda_x)} \geqslant \sum_{x=2}^{m} f_x.$$

We now consider estimation of λ_x.

The marginal density p_x can be estimated by the relative empirical frequency f_x/N so

$$\widehat{E(\lambda|x)} = \hat{\lambda}_x = (x+1)\frac{f_{x+1}}{f_x}$$

provides an estimate of the posterior mean $E(\lambda|x) = \lambda_x$, using the fact that the unknown denominators N cancel out. Hence, the Zelterman estimate occurs as a special case of the non-parametric, empirical Bayes estimator for observation x (Robbins, 1955; Carlin and Louis, 1997).

The understanding of Zelterman's original estimator of λ as $\hat{\lambda}_1 = 2f_2/f_1$ as an empirical Bayes estimator for observation $x = 1$ is useful, since it helps to find ways to eliminate the overestimation bias. We need to define a Horvitz–Thompson estimator that takes into account the different counts $x = 1, 2, \ldots$ separately. This can be accomplished by defining

$$\hat{N} = \frac{f_1}{1-\exp(-\hat{\lambda}_1)} + \frac{f_2}{1-\exp(-\hat{\lambda}_2)} + \ldots + \frac{f_m}{1-\exp(-\hat{\lambda}_m)}. \tag{6}$$

The question arises about which way the estimator $\hat{\lambda}_x$ should be constructed. A naive estimator would follow Robbins-type estimation to arrive at

$$\hat{N}_{\text{EB-Robbins}} = \frac{f_1}{1-\exp(-2f_2/f_1)} + \frac{f_2}{1-\exp(-3f_3/f_2)} + \ldots + \frac{f_{m-1}}{1-\exp(-m f_m/f_{m-1})} + f_m, \tag{7}$$

where we define

$$\frac{f_j}{1-\exp\{-(j+1)f_{j+1}/f_j\}} = \begin{cases} 0, & \text{if } f_j = 0, \\ f_j, & \text{if } f_{j+1} = 0. \end{cases}$$

Although estimator (7) is intuitively attractive, it has some considerable difficulties. Not only is it unclear what to do with the largest count m (in equation (7) it is not upweighted), but also various counts could have frequencies 0 which would leave some of the frequencies f_x unweighted. More importantly, most of the observed count data will lie on the lower counts, resulting in highly unstable estimates for larger counts.

It is more attractive to consider a *smoothed* version of the Bayes estimator. This can be accomplished by constructing an estimate of the marginal distribution $p_x = \int_0^\infty p(x|\lambda)\, q(\lambda)\, \mathrm{d}\lambda$ by using a discrete finite mixture $p_x = \Sigma_{j=1}^k \mathrm{Po}(x|\lambda_j) q_j$, where $\lambda_j > 0$ and the non-negative weights q_j sum to 1. Estimates of the mixing distribution can be constructed as described in Section 3.3 although their usage here is entirely different in the sense that the mixing distribution is only used to construct a smooth estimate of p_x. We arrive at some estimate of the marginal distribution

$$\hat{p}_x = \sum_{j=1}^{k} \mathrm{Po}(x|\hat{\lambda}_j)\hat{q}_j \tag{8}$$

leading to smoothed estimates of the population size

$$\hat{N}_{\text{EB-NPMLE}} = \sum_{l=1}^{m} \frac{f_l}{1 - \exp\{-(l+1)\,\hat{p}_{l+1}/\hat{p}_l\}}, \tag{9}$$

where we attach a subscript EB to the population size estimate N_{NPMLE} in equation (9) to point out that a smoothed estimate $\hat{p}_x$ of the marginal distribution is used to construct an estimate of N.

Other ways of estimating the mixing distribution $q(\lambda)$ in $\int_0^\infty \mathrm{Po}(x|\lambda)\, q(\lambda)\, \mathrm{d}\lambda$ are possible as well. For example, we could use the empirical distribution itself as an estimator of the mixing distribution. Alternatively, we could think of a parametric mixing distribution such as a gamma distribution for $q(\lambda)$. We do not follow up on this here since our simulation work has indicated that these approaches perform less satisfactorily than the purely non-parametric $\hat{N}_{\text{EB-Robbins}}$ and the smoothed non-parametric $\hat{N}_{\text{EB-NPMLE}}$.

In what follows we continue the simulation study and provide evidence that the suggested empirical Bayes estimator performs better than the conventionally used estimators $\hat{N}_{\text{C}}$ and, in particular, $\hat{N}_{\text{Z}}$. Besides these two conventional estimators we shall consider the non-parametric estimator $\hat{N}_{\text{EB-Robbins}}$ and the smoothed mixture model version $\hat{N}_{\text{EB-NPMLE}}$. The design of the simulation corresponds to that used previously. Samples of counts $X_1, \ldots, X_N$ were drawn from a two-component mixture of Poisson densities: $X \sim 0.5\,\mathrm{Po}(1) + 0.5\,\mathrm{Po}(\lambda)$, evidently with equal weights $q_1 = q_2 = 0.5$. The population size was set to $N = 100$ and 1000 replications were used. Here, we shall concentrate on the main findings. More details are available in the on-line supporting information. We see from Table 2 that both empirical Bayes estimators perform better with respect to their standard error and root-mean-square error than the other estimators adjusting for heterogeneity. If we compare the two empirical Bayes estimators it appears that the estimator that is based on the non-parametric mixture model has smaller variance, which is reflected also in a better mean-squared error.

5. Application to spatial analysis of scrapie in Great Britain

Following the results of the previous section we shall concentrate on using the NPMLE of the mixing distribution as the smoothed empirical Bayes estimate of the prior distribution for further analysis, in particular $\hat{p}_x = \Sigma_{j=1}^k \mathrm{Po}(x|\hat{\lambda}_j)\hat{q}_j$, as derived in equation (8). In the first step, this will be done using the entire SND data, unstratified by county. Once an estimate for the mixing distribution has been achieved, a smoothed *county-specific* estimate of the population size can be developed as follows:

$$\hat{N}_{\text{EB-NPMLE},i} = \sum_{l=1}^{m} \frac{f_{l,i}}{1 - \exp\{-(l+1)\,\hat{p}_{l+1}/\hat{p}_l\}}, \tag{10}$$

Table 3. Estimated mixture models for k equal to 1, 2, 3, 4 and 5 (NPMLE) components with associated estimator of the size of the scrapie–affected population of holdings from the unstratified SND database 2002–2006†

k	$\hat{\lambda}_j$	$\hat{q}_j$	$\log\{L(\hat{Q})\}$	*BIC*	*Discrete mixture model based results*	
					$\hat{N}_{\text{EB-NPMLE}(k)}$	$\hat{N}_{\text{NPMLE}(k)}$
1	2.33	1.00	−1279.0	2561.4	572 (9.4)	572 (9.4)
2	0.97	0.88	−865.4	1740.8	776 (32.4)	793 (34.6)
	9.80	0.12				
3	0.67	0.80	−807.8	1632.4	869 (44.8)	946 (65.8)
	5.46	0.17				
	19.10	0.03				
4	0.56	0.75	−802.3	1628.2	896 (48.0)	1036 (60102)
	4.03	0.19				
	10.35	0.05				
	23.58	0.01				
5	0.01	0.27	−801.2	1632.7	916 (25.5)	528694 (419663)
	1.08	0.54				
	5.13	0.14				
	11.76	0.03				
	23.98	0.01				

†Standard errors are given in parentheses.

where $f_{l,i}$ is the frequency of holdings with l cases in the ith county and $\hat{p}_l$ is taken from equation (8).

5.1. Determining the non-parametric maximum likelihood estimate for the scrapie notification database data

We have seen in Section 4 by using the ratio plot that there is strong evidence for heterogeneity captured by a mixing distribution. We consider the marginal distribution over all counties as available from Table 1: $f_1 = 298$, $f_2 = 89$, $f_3 = 42, \ldots, f_{29} = 2$. We use this (truncated) count distribution to determine the maximum likelihood estimators for the various possible mixture models. The results are provided in Table 3. For each number of components k, starting with the homogeneous case $k = 1$, the estimated mixture model $\hat{Q}$ is provided, the Poisson parameters $\hat{\lambda}_j$ and associated component weights $\hat{q}_j$. This is followed by the log-likelihood $\log\{L(\hat{Q})\}$ and the BIC-value $-2\log\{L(\hat{Q})\} + (2k-1)\log(n)$. Note that two estimates of the population size of scrapie-affected holdings are given. One is based on the direct computation using the mixture model estimated as provided in equation (3); the other is the empirical Bayes estimate by using the estimated mixture as prior distribution (8). It is evident from the last two columns in Table 3 that the empirical Bayes estimate of the population size is less sensitive to the choice of the number of components. Furthermore, the empirical Bayes estimate is not prone to spurious estimates as is the conventional mixture-model-based estimator. We have already mentioned that Fig. 2 supports that there is considerable evidence for a monotone increasing pattern. In addition, the estimate of the posterior mean based on the estimated mixture model with four components (this is what BIC suggests) shows that this monotone pattern is met. Note that the last two columns in Table 3 contain also (in parentheses) an estimate of the standard error of the respective population size estimate. This was achieved by applying the non-parametric bootstrap as adapted to capture–recapture situations by van der Heijden *et al.* (2003) and

Böhning (2008). It is evident from the last column in Table 3 that the conventional mixture-model-based estimator is prone to extreme variance inflation when the number of components becomes large.

5.2. Estimating the number of hidden scrapie-affected holdings per county

We now apply these results to the individual counties by using equation (10). Note that we are using the same mixture distribution in equation (10) estimated from the entire SND data. This is necessary since the county-specific case distributions are frequently very sparse. Take for example county 1 in Table 1: we find $f_{1,1}=2$, $f_{2,1}=1$ and $f_{3,1}=1$, so $n_1=4$. It is clear that a reliable estimation of a mixing distribution is not possible from this count distribution. Hence we use the mixing distribution that is estimated from the entire data set and assume that the heterogeneity found for the entire data set is also valid in each county. Then we compute the *predicted* number of scrapie-affected holdings by applying the weight $[1-\exp\{-(l+1)\hat{p}_{l+1}/\hat{p}_l\}]^{-1}$ to the frequency $f_{l,i}$ of count l in the ith county and summing over all observed frequencies $f_{l,i}$, leading to equation (10). This process is very similar to indirect standardization that is used in epidemiologic methodology (see Waller and Gotway (2004), page 17). The results are provided in Table 4. Details on the computations of standard errors for the estimated population size are found in Appendix B. Note that in a county i with all $f_{l,i}=0$ except one, say $f_{l,j}>0$, there is no variation in the count distribution and, hence, there is no estimated standard error (at least not without making further assumptions). In addition, two further measures are computed. The observed–hidden ratio defined as $n_i/(\hat{N}_i-n_i)$ and the completeness measure defined as $n_i/\hat{N}_i$, provided as the last two columns in Table 4. The completeness ranges between 48% and 99%. Fig. 3 shows a scatter plot of the completeness against the observed count (on the log-scale) of scrapie-affected holdings. There is no evidence for a specific pattern, though the variation of completeness seems to decrease with increasing observed count of scrapie-affected holdings. The median observed–hidden ratio is 1.29 with 95% non-parametric confidence interval (1.11, 1.43) and the completeness is 56.36 with 95% non-parametric confidence interval (52.62%, 58.83%).

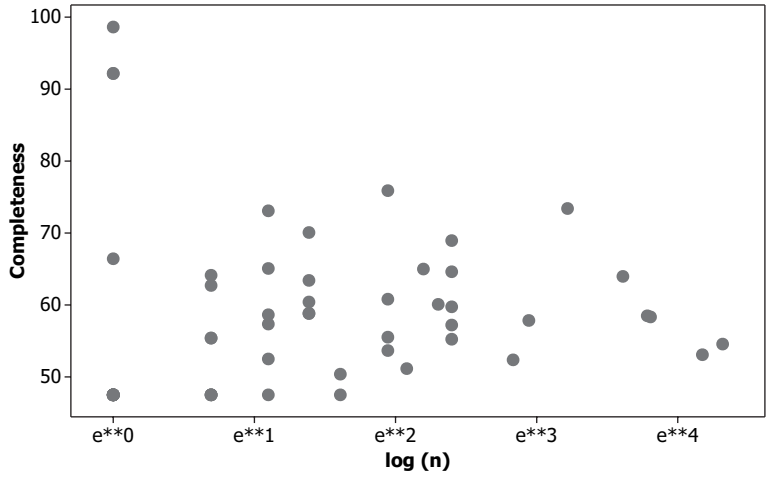

Fig. 3. Scatterplot of completeness of the surveillance stream per county against the observed count of scrapie-affected holdings per county

Table 4. Sizes of observed scrapie-affected holdings (n), estimates of the total population size of scrapie-affected holdings ($\hat{N}$), observed–hidden ratio o/h and completeness of the SND by county, 2002–2006

County	n	$\hat{N}$	*Confidence interval for N*	o/h	*Completeness (%)*
1	4	7	5.3–8.3	1.4	59
2	4	6	4.3–7.1	2.3	70
3	1	2	—†	0.9	48
4	1	2	—†	0.9	48
5	7	9	7.1–11.4	3.1	76
6	11	16	13.2–18.7	2.2	69
7	19	33	29.2–36.5	1.4	58
8	11	20	17.5–22.3	1.2	55
9	45	77	71.8–82.5	1.4	58
10	5	10	8.8–11.1	1.0	50
11	1	2	—†	0.9	48
12	1	1	—†	11.8	92
13	3	5	3.8–6.7	1.3	57
14	3	5	4.1–6.2	1.4	59
15	1	2	—†	2.0	66
16	10	17	14.1–19.2	1.5	60
17	1	2	—†	0.9	48
18	5	11	—†	0.9	48
19	2	4	2.7–4.5	1.2	55
20	1	2	—†	0.9	48
21	4	7	5.3–8.3	1.4	59
22	9	14	11.6–16.1	1.9	65
23	7	13	10.5–14.7	1.2	56
24	2	4	—†	0.9	48
25	3	5	3.3–5.9	1.9	65
26	8	16	14.1–17.2	1.0	51

(*continued*)

Finally, note that the map is based on an estimated size of the scrapie population in county i, given as $\hat{N}_{\text{EB-NPMLE},i} = \Sigma_{l=1}^{m} \hat{w}_l f_{l,i}$, where the estimated weights

$$\hat{w}_l = \frac{1}{1 - \exp\{-(l+1)\hat{p}_{l+1}/\hat{p}_l\}}$$

do not depend on the county index i. Hence, we have that

$$\sum \hat{N}_i = \sum_i \sum_{l=1}^{m} \hat{w}_l f_{l,i} = \sum_{l=1}^{m} \hat{w}_l \sum_i \hat{f}_{l,i} = \sum_{l=1}^{m} \hat{w}_l f_l = \hat{N},$$

where $f_l = \Sigma_i f_{l,i}$, so the margin (over counties) of the county-specific estimates of the size of the scrapie population and the estimate of the size of the scrapie population, unstratified by county, coincide.

Fig. 4 shows the geographical distribution of county-specific completeness. The completeness is fairly stable with most counties in the 50–59% category and fewer counties in the upper completeness categories. Note that, as well as providing completeness and observed–hidden ratios, we can also estimate adjusted measures of occurrence of disease for each county. However, for our particular case, this would not have a clear biological interpretation as annual data were pooled to increase the power of our analyses.

Table 4 (*continued*)

County	*n*	$\hat{N}$	*Confidence interval for N*	*o/h*	*Completeness (%)*
27	7	13	11.2–14.9	1.2	54
28	3	4	2.6–5.6	2.7	73
29	4	6	4.6–8.1	1.7	63
30	1	2	—†	0.9	48
31	3	6	4.7–6.8	1.1	52
32	1	2	—†	0.9	48
33	2	3	2.0–4.4	1.7	63
34	37	58	53.1–62.6	1.8	64
35	2	4	—†	0.9	48
36	2	4	—†	0.9	48
37	75	137	131.1–143.8	1.2	55
38	11	19	16.7–21.7	1.3	57
39	44	75	70.0–80.5	1.4	58
40	25	34	30.4–37.7	2.8	73
41	11	17	14.8–19.3	1.8	65
42	11	18	15.7–21.1	1.5	60
43	2	4	2.7–4.5	1.2	55
44	7	12	9.2–13.8	1.6	61
45	1	2	—†	0.9	48
46	3	6	—†	0.9	48
47	2	3	1.9–4.4	1.8	64
48	1	1	—†	11.8	92
49	1	2	—†	0.9	48
50	2	4	—†	0.9	48
51	1	2	—†	0.9	48
52	17	32	29.8–35.2	1.1	52
53	1	1	—†	71.6	99
54	1	2	—†	0.9	48
55	65	122	117.0–127.9	1.1	53
56	4	7	5.5–7.7	1.5	60

†No estimated standard error available.

6. Discussion

We have developed an estimator here with lower bound properties similar to those of Chao's estimator but with potential gains in precision. In addition, the estimator is very suitable for dealing with highly stratified data such as in our application. However, the approach requires assumptions and one of them is that the heterogeneity in the strata is similar to the heterogeneity observed and estimated in the unstratified population. This assumption seems reasonable for the application that was discussed here. It would be desirable to investigate possible extensions of the model, including the incorporation of covariates in the model.

In addition the estimator can be used in other applications where estimates of population size for a large number of strata are of interest. This was the motivation behind the application to the scrapie data where county-specific estimates of completeness of surveillance are of interest. We note that different spatially specific stratifications are possible, provided that our simplified assumption of homogeneity across the units of interest remains generally valid or extensions to the model to account for heterogeneity are incorporated, for scrapie or other conditions. One application would be to assess the completeness of surveillance by geographical catchment areas representing the organizational units of a surveillance system.

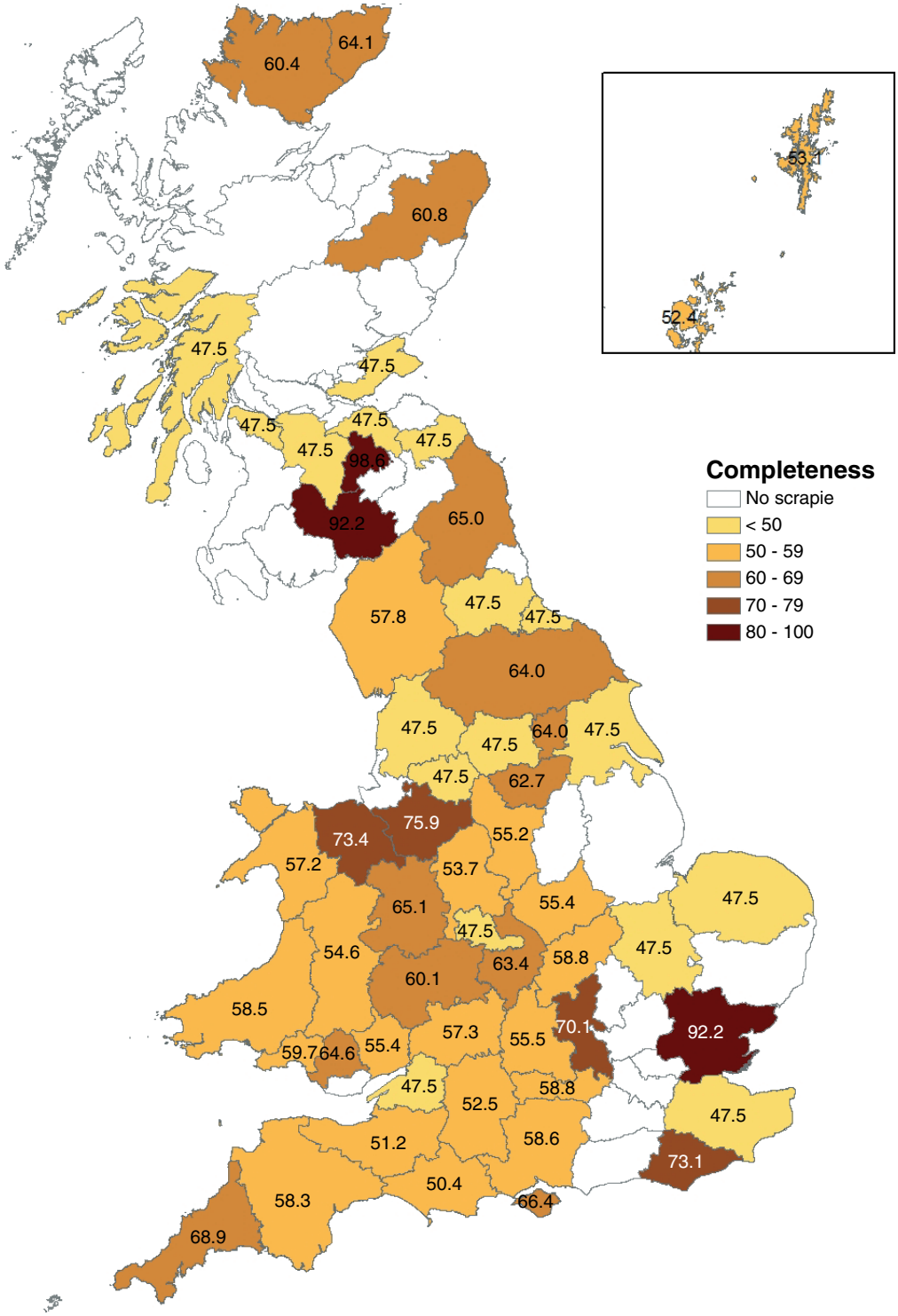

Fig. 4. Map of estimated completeness at the county level for the SND data, 2002–2006

Appendix A: Proofs of theorems 2 and 3

A.1. Proof of theorem 2

For the first part, it is clear that $f_1 \log(p_1) + f_2 \log(p_2)$ is maximal for $\hat{p}_1 = f_1/(f_1+f_2)$, which is attained for $\hat{\lambda} = 2f_2/f_1$. For the second part, we see that, with $e_x = E(f_x|f_1, f_2; \lambda) = \mathrm{Po}(x|\lambda)N$,

$$e_x = \mathrm{Po}(x|\lambda)N = \mathrm{Po}(x|\lambda)\left(e_0 + f_1 + f_2 + \sum_{j=3}^{\infty} e_j\right)$$

so that

$$e_0 + e_3^+ = \{1 - \mathrm{Po}(1|\lambda) - \mathrm{Po}(2|\lambda)\}\,(e_0 + e_3^+) + \{1 - \mathrm{Po}(1|\lambda) - \mathrm{Po}(2|\lambda)\}\,(f_1 + f_2)$$

with $e_3^+ = \Sigma_{j=3}^{\infty} e_x$. Hence

$$e_0 + e_3^+ = \frac{1 - \mathrm{Po}(1|\lambda) - \mathrm{Po}(2|\lambda)}{\mathrm{Po}(1|\lambda) + \mathrm{Po}(2|\lambda)}(f_1 + f_2)$$

and

$$\begin{aligned} e_0 = \mathrm{Po}(0|\lambda)(f_1 + f_2 + e_0 + e_3^+) &= \mathrm{Po}(0|\lambda)(f_1 + f_2) + \mathrm{Po}(0|\lambda)\frac{1 - \mathrm{Po}(1|\lambda) - \mathrm{Po}(2|\lambda)}{\mathrm{Po}(1|\lambda) + \mathrm{Po}(2|\lambda)}(f_1 + f_2) \\ &= \frac{\mathrm{Po}(0|\lambda)}{\mathrm{Po}(1|\lambda) + \mathrm{Po}(2|\lambda)}(f_1 + f_2) = \frac{f_1 + f_2}{\lambda + \lambda^2/2}. \end{aligned}$$

Plugging in the maximum likelihood estimate $\hat{\lambda} = 2f_2/f_1$ for λ yields the desired result.

A.2. Proof of theorem 3

Consider

$$p_j = \int_0^{\infty} \exp(-\lambda)\lambda^j/j!\, q(\lambda)\, \mathrm{d}\lambda$$

with unknown $q(\lambda)$ for $\lambda > 0$. Then, by means of the *Cauchy–Schwarz inequality* for random variables X and Y,

$$E(XY)^2 \leqslant E(X^2)\, E(Y^2)$$

we have that

$$\left\{ \int_0^{\infty} \overbrace{\sqrt{\exp(-\lambda)\lambda^{(j-1)/2}}}^{X} \overbrace{\sqrt{\exp(-\lambda)\lambda^{(j+1)/2}}}^{Y} \mathrm{d}\lambda \right\}^2$$

$$\leqslant \int_0^{\infty} \overbrace{\exp(-\lambda)\lambda^{(j-1)}}^{X^2} \mathrm{d}\lambda \int_0^{\infty} \overbrace{\exp(-\lambda)\lambda^{(j+1)}}^{Y^2} \mathrm{d}\lambda,$$

or

$$(j!\, p_j)^2 \leqslant (j-1)!\, p_{j-1}(j+1)!\, p_{j+1}$$

or, finally, $j p_j / p_{j-1} \leqslant (j+1) p_{j+1}/p_j$.

Appendix B: Standard error estimates of county-specific population size estimates

It is also possible to derive estimates for the standard errors of $\hat{N}_{\text{EB-NPMLE,i}} = \Sigma_{l=1}^{m} \hat{w}_l f_{l,i} = \hat{\mathbf{w}}^{\mathrm{T}} \mathbf{f}_i$. Here the estimated weight w_l is given by $\hat{w}_l = 1/\{1 - \exp(-(l+1)\hat{p}_{l+1}/\hat{p}_l)\}$. The variance conditional on $\hat{\mathbf{w}}$ is simply $\hat{\mathbf{w}}^{\mathrm{T}} \mathrm{cov}(\mathbf{f}_i)\hat{\mathbf{w}}$ with $\mathrm{cov}(\mathbf{f}_i) = \Lambda_{\mathbf{f}_i} - \mathbf{f}_i \mathbf{f}_i^{\mathrm{T}}/n_i$, where $n_i = \Sigma_{l=1}^{m} f_{l,i}$ and $\Lambda_{\mathbf{f}_i}$ the diagonal matrix with elements $f_{l,i}$, $l = 1, \ldots, m$, on the diagonal. This variance estimate is dependent on the vector $\mathbf{f}_i$ and will be different for each county, but it is conditional on

$$\hat{w}_l = \frac{1}{1 - \exp\{-(l+1)\hat{p}_{l+1}/\hat{p}_l\}}$$

for $l = 1, \ldots, m$, which is identical for each county. Although a conditional variance estimate seems appropriate for comparison of variation within the county strata, it might sometimes be desirable to provide an unconditional variance estimate. This can be achieved by adding an additional variance component due to the random error involved in the estimate $\hat{\mathbf{w}}$ (for more details on variance computations in the capture–recapture setting see Böhning (2008)), so that the unconditional variance estimate becomes

$$\widehat{\text{var}}(\hat{N}_{\text{EB-NPMLE},i}) = \hat{\mathbf{w}}^{\mathrm{T}} \text{cov}(\mathbf{f}_i)\hat{\mathbf{w}} + \mathbf{f}_i^{\mathrm{T}} \text{cov}(\hat{\mathbf{w}})\hat{\mathbf{f}}_i,$$

where $\text{cov}(\hat{\mathbf{w}})$ is the covariance matrix for the vector $\mathbf{w}$. This needs to be determined only once for the entire data set but will depend on the estimator that is used to estimate $\hat{p}_l$ in

$$\hat{w}_l = \frac{1}{1 - \exp\{-(l+1)\hat{p}_{l+1}/\hat{p}_l\}}$$

and it is best done by using the non-parametric bootstrap that was mentioned in Section 4.

References

Böhning, D. (2003) The EM algorithm with gradient function update for discrete mixtures with known (fixed) number of components. *Statist. Comput.*, **13**, 257–265.

Böhning, D. (2008) A simple variance formula for population size estimators by conditioning. *Statist. Methodol.*, **5**, 410–423.

Böhning, D. and Del Rio Vilas, V. J. (2008) Estimating the hidden number of scrapie affected holdings in Great Britain using a simple, truncated count model allowing for heterogeneity. *J. Agric. Biol. Environ. Statist.*, **13**, 1–22.

Böhning, D. and Kuhnert, R. (2006) The equivalence of truncated count mixture distributions and mixtures of truncated count distributions. *Biometrics*, **62**, 1207–1215.

Bunge, J. and Fitzpatrick, M. (1993) Estimating the number of species: a review. *J. Am. Statist. Ass.*, **88**, 364–373.

Carlin, B. P. and Louis, T. A. (1997) *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman and Hall.

Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783–791.

Chao, A. (1989) Estimating population size for sparse data in capture-recapture experiments. *Biometrics*, **45**, 427–438.

Del Rio Vilas, V. J., Guitian, J., Pfeiffer, D. U. and Wilesmith, J. W. (2006) Analysis of data from the passive surveillance of scrapie in Great Britain between 1993 and 2002. *Veter. Rec.*, **159**, 799–804.

Del Rio Vilas, V. J., Sayers, R., Sivam, K., Pfeiffer, D. U., Guitian, J. and Wilesmith, J. W. (2005) A case study of capture-recapture methodology using scrapie surveillance data in Great Britain. *Prev. Veter. Med.*, **67**, 303–317.

Dorazio, R. M. and Royle, J. A. (2005) Mixture models for estimating the size of a closed population when capture rates vary among individuals. *Biometrics*, **59**, 351–364.

Gale, W. A. and Damson, G. (1995) Good-Turing frequency estimation without tears. *J. Quant. Ling.*, **2**, 217–237.

Good, I. J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237–264.

Hampel, F. R., Rousseeuw, P. J., Ronchetti, E. M. and Stahel, W. A. (1986) *Robust Statistics—the Approach based on Influence Functions*. New York: Wiley.

van der Heijden, P. G. M., Cruyff, M. and van Houwelingen, H. C. (2003) Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Statist. Neerland.*, **57**, 1–16.

Hoinville, L. J., Hoek, A. R., Gravenor, M. B. and McLean, A. R. (2000) Descriptive epidemiology of scrapie in Great Britain: results of a postal survey. *Veter. Rec.*, **146**, 455–461.

Holzmann, H., Munk, A. and Zucchini, W. (2006) On identifiability in capture-recapture models. *Biometrics*, **62**, 934–939.

Lindsay, B. G. (1983) The geometry of mixture likelihoods, part I: a general theory. *Ann. Statist.*, **11**, 783–792.

Link, W. A. (2003) Nonidentifiability of population size from capture-recapture data with heterogeneous detection probabilities. *Biometrics*, **59**, 1123–1130.

Link, W. A. (2006) Response to a paper by Holzmann, Munk and Zucchini. *Biometrics*, **62**, 936–939.

McKendrick, A. G. (1926) Application of mathematics to medical problems. *Proc. Edinb. Math. Soc.*, **44**, 98–130.

Mosley, W. H., Bart, K. J. and Sommer, A. (1972) An epidemiological assessment of cholera control programs in rural East Pakistan. *Int. J. Epidem.*, **1**, 5–11.

Pledger, S. A. (2005) The performance of mixture models in heterogeneous closed population capture-recapture. *Biometrics*, **61**, 868–876.

Robbins, H. (1955) An empirical Bayes approach to statistics. In *Proc. 3rd Berkeley Symp. Mathematical Statistics and Probability* (ed. J. Neyman), vol. 1, pp. 157–164. Berkeley: University of California Press.

Waller, L. A. and Gotway, C. A. (2004) *Applied Spatial Statistics for Public Health Data*. Hoboken: Wiley.
Wang, J.-P. and Lindsay, B. G. (2005) A penalized nonparametric maximum likelihood approach to species richness estimation. *J. Am. Statist. Ass.*, **100**, 942–959.
Wang, J.-P. and Lindsay, B. G. (2008) An exponential partial prior for improving nonparametric maximum likelihood estimation in mixture models. *Statist. Methodol.*, **5**, 30–45.
Wilson, R. M. and Collins, M. F. (1992) Capture-recapture estimation with samples of size one using frequency data. *Biometrika*, **79**, 543–553.
Zelterman, D. (1988) Robust estimation in truncated discrete distributions with applications to capture-recapture experiments. *J. Statist. Planng Inf.*, **18**, 225–237.

Supporting information
Additional 'supporting information' may be found in the on-line version of this article:

'Supporting information for Capture-recapture estimation by means of empirical Bayesian smoothing with an application to spatial analysis of hidden scrapie in Great Britain'.