The Annals of Applied Statistics
2019, Vol. 13, No. 2, 1198–1211
https://doi.org/10.1214/18-AOAS1232

THE IDENTITY OF THE ZERO-TRUNCATED, ONE-INFLATED LIKELIHOOD AND THE ZERO-ONE-TRUNCATED LIKELIHOOD FOR GENERAL COUNT DENSITIES WITH AN APPLICATION TO DRINK-DRIVING IN BRITAIN

BY DANKMAR BÖHNING AND PETER G. M. VAN DER HEIJDEN

University of Southampton and University of Utrecht

For zero-truncated count data, as they typically arise in capture-recapture modelling, we consider modelling under one-inflation. This is motivated by police data on drink-driving in Britain which shows high one-inflation. The data, which are used here, are from the years 2011 to 2015 and are based on DR10 endorsements. We show that inference for an arbitrary count density with one-inflation can be equivalently based upon the associated zero-one truncated count density. This simplifies inference considerably including maximum likelihood estimation and likelihood ratio testing. For the drink-driving application, we use the geometric distribution which shows a good fit. We estimate the total drink-driving as about 2,300,000 drink-drivers in the observational period. As 227,578 were observed, this means that only about 10% of the drink-driving population is observed with a bootstrap confidence interval of 9%–12%.

1. Introduction. This work is motivated by the following application. Drink-driving (DD) is a serious problem in many countries in the world, including the UK. It relates to driving (or attempting to drive) while being above the legal alcohol limit. Typically, persons involved in an accident are sampled using a breath-testing device. According to the Guardian from 30/December/2016 [The Guardian (2016)], in Britain, there were 227,578 Driver & Vehicle License Agency (DVLA) reported motorists between 2011 and 2015 of which 219,000 motorists were caught once by the police, 8068 twice, etc. (see Table 1 below). These numbers are based on DR10 endorsements. A DR10 is a United Kingdom motoring endorsement issued by the Driver and Vehicle Licensing Agency (DVLA) and UK Police which means driving or attempting to drive with a blood alcohol level above the allowable limit. Following an arrest for a DR10 one can expect a ban from driving and a fine. In more serious cases, a Community Order, or a prison term of up to six months may be issued. It is clear that DD is largely a hidden activity as many DD-drivers remain unidentified. Hence the cell f_0 is empty as it is unknown from the collected data.

Hence we model the count X of identifications of a driver with some count density $p(x, \theta)$ where $x = 0, 1, 2, \dots$ and θ is a parameter or parameter vector. The

Received May 2018; revised November 2018.

Key words and phrases. Capture–recapture, Chao estimator, behavioral response, power series distribution, mixture model, zero-truncated model, nonparametric estimator of population size.

TABLE 1

Frequency distribution of the count (per person) of DVLA reported drink-driving (DD) in Britain between 2011 and 2015 (figures are based on DR10 endorsements; see also explanations in the text); here f_x is the frequency of drivers identified exactly x times

Count of DD	f_0	f_1	f_2	f_3	f_4	f_5	f_6	n
Frequency		219,008	8068	449	46	5	2	227,578

background population of drink-drivers can be enumerated from 1 to N, the latter being unknown. Hence we have a sample of counts of identifications $X_1, \dots, X_N$ arising in the observational period, where we do not observe $X_i = 0$: any zero-counts remain hidden. Hence we consider the associated zero-truncated density $p_+(x, \theta) = p(x, \theta)/[1 - p(0, \theta)]$ to model the observational process. Let n denote the size of the observed zero-truncated counts with f_x being the frequency of observing exactly x counts. The largest observed count is denoted as m. Drink-drivers (in general terminology called *units*) that are identified only once are also called singletons, units that occur twice are called doubletons, and so forth. In Table 1, there are 219,008 singletons, whereas there are only 8068 doubletons and 449 tripletons. This huge number of singletons might be easily explained as being caught once by the police might change the DD behavior considerably so that such an event will not occur again (because of potential legal consequences). To incorporate extra-ones or one-inflation into the modeling we consider

$$p_{+1}(x, \theta) = \begin{cases} (1 - \omega) + \omega p_+(x, \theta), & \text{if } x = 1, \\ \omega p_+(x, \theta), & \text{if } x > 1, \end{cases} \tag{1}$$

where $1 - \omega$, $\omega \in [0, 1]$, is an extra mass at $x = 1$ that controls the amount of one-inflation. We call this the zero-truncated, one-inflated density $p_{+1}(x, \theta)$ arising from count density $p(x, \theta)$ which we call the *baseline* density. We provide two examples of baseline densities which are used frequently in count data modeling. One is the Poisson distribution given as

$$p(x, \theta) = \exp(-\theta)\theta^x/x!$$

for $x = 0, 1, 2, \dots$ and $\theta > 0$. The other is the geometric given as

$$p(x, \theta) = (1 - \theta)^x \theta$$

for $x = 0, 1, 2, \dots$ and $\theta \in (0, 1)$. The geometric distribution is typically introduced as a discrete time-to-event distribution although it also can be thought of as flexible count data distribution as it occurs when a Poisson distribution is mixed with an exponential distribution:

$$(1 - \theta)^x \theta = \int_0^\infty \exp(-\lambda)\lambda^x/x! \times \frac{1}{\vartheta} \exp(-\lambda/\vartheta)\, d\lambda,$$

where $\theta = 1/(1+\vartheta)$. Note that the geometric distribution has the simple property that $p_+(x,\theta) = (1-\theta)^{x-1}\theta$ for $x = 1, 2, \dots$ and $p_{++}(x,\theta) = p(x,\theta)/[1 - p(0,\theta) - p(1,\theta)] = (1-\theta)^{x-2}\theta$ for $x = 2, 3, \dots$, implying that truncated geometric distributions are also geometric distributions.

The associated likelihood is

$$L_{+1} = \prod_{x=1}^{m} p_{+1}^{f_x} = \bigl[(1-\omega) + \omega p_+(1,\theta)\bigr]^{f_1} \prod_{x=2}^{m} \bigl[\omega p_+(x,\theta)\bigr]^{f_x}. \tag{2}$$

We also consider the zero-one-truncated density

$$p_{++}(x,\theta) = p_+(x,\theta)/\bigl[1 - p_+(1,\theta)\bigr] = p(x,\theta)/\bigl[1 - p(0,\theta) - p(1,\theta)\bigr]$$

for $x = 2, 3, \dots$ with associated likelihood

$$L_{++} = \prod_{x=2}^{m} p_{++}(x,\theta)^{f_x}. \tag{3}$$

We will show in the following that the zero-truncated one-inflated likelihood is identical (up to a constant independent of θ) to the zero-one truncated likelihood. Working with the zero-one truncated likelihood simplifies inference considerably as computational tools are readily available for standard count densities. We then utilize these results in estimation the amount of drink-driving in Britain, including the frequency f_0 of hidden units in the population of drink-drivers.

In many applications, one-inflation can be explained as behavioral change after first identification as it is a plausible explanation in our application of drink-driving. For other applications of one-inflation see Godwin and Böhning (2017). Other ways of one-inflation occurrence are discussed in Bunge, Willis and Walsh (2014). For a general introduction into capture-recapture problems see Borchers, Buckland and Zucchini (2004), Bunge and Fitzpatrick (1993), McCrea and Morgan (2015) or Böhning, Bunge and van der Heijden (2018).

The paper is organized as follows. In the next section, we show a more general result. For any discrete density with one sample element, x_0 say, truncated and another element, x_1 say, inflated it is shown that the x_0-truncated, x_1-inflated likelihood is identical to the x_0-x_1-truncated likelihood. In Section 3, we consider a likelihood ratio test for testing the null hypothesis of no x_1-inflation against the alternative of x_1-inflation. For count densities, we then utilize these results in estimation of the total population size, including the frequency f_0 of hidden units in the population in Section 4. For confidence interval estimation, we use the imputed bootstrap in Section 5. Finally, we apply these results to estimate the magnitude of DD in Britain in Section 6.

2. Identity. We intend to show the identity of the likelihoods (2) and (3) (up to a parameter-independent constant). In fact, we prove a more general result. Let X take one of the values in the sequence $x_0, x_1, x_2, \dots,$ $x_i \neq x_j$ for $i \neq j$, not necessarily count values. Also, the associated probabilities $P(X = x_i)$ are given by some parametric model $p(x_i, \theta)$ where θ is a real parameter or vector. Let us assume that there is one value of X which is not observable and we take this value without limitation of generality to be x_0. The associated x_0-truncated discrete density is then $p_+(x_i, \theta) = p(x_i, \theta)/[1 - p(x_0, \theta)]$ for any $x_i \neq x_0$. Furthermore, we assume that there is a second value, again without limitation of generality let this value be x_1, which is inflated:

$$p_{+1}(x_i, \theta) = \begin{cases} (1-\omega) + \omega p_+(x_i, \theta), & \text{if } x_i = x_1, \\ \omega p_+(x_i, \theta), & \text{if } x_i \neq x_0, x_1, \end{cases} \tag{4}$$

which we call the x_0-truncated, x_1-inflated density of the baseline distribution $p(x, \theta)$.

Let $f_1, f_2, \dots, f_m$ be the observed sample frequencies of $x_1, x_2, \dots, x_m$. Here m denotes the last observed sample element in the lexicographical order. Then, we have the associated likelihood as

$$L_{+1} = \big[(1-\omega) + \omega p_+(x_1, \theta)\big]^{f_1} \prod_{i=2}^{m} \big[\omega p_+(x_i, \theta)\big]^{f_i}. \tag{5}$$

In addition, we consider the x_0-x_1-truncated density

$$p_{++}(x, \theta) = p_+(x, \theta)/\big[1 - p_+(x_1, \theta)\big] = p(x, \theta)/\big[1 - p(x_0, \theta) - p(x_1, \theta)\big]$$

for $x = x_2, x_3, \dots$ with associated likelihood

$$L_{++} = \prod_{i=2}^{m} p_{++}(x_i, \theta)^{f_i}. \tag{6}$$

We have the following general result.

THEOREM 1. *Let $n = f_1 + f_2 + \cdots + f_m$ and $n_1 = n - f_1$. Then*

$$\log L_{+1} = f_1 \log(f_1/n) + n_1 \log(1 - f_1/n) + \log L_{++}. \tag{7}$$

Furthermore, for any fixed θ,

$$\hat{\omega} = \frac{1 - f_1/n}{1 - p_+(x_1, \theta)} \tag{8}$$

and

$$1 - \hat{\omega} + \hat{\omega} p_+(x_1, \theta) = f_1/n. \tag{9}$$

The result implies that fitting the zero-truncated one-inflated likelihood, for an arbitrary count density, can be accomplished using the zero-one-truncated likelihood. Furthermore, the fit of the one-inflated component will be identical to the observed proportion f_1/n. Hence this implies also that the fit of the model will entirely depend on the zero-one truncated part and any analysis can be restricted to this.

We now give a proof of Theorem 1. From the x_0-truncated, x_1-inflated likelihood (5) we get the log-likelihood

$$\log L_{+1}(\omega,\theta) = f_1 \log\bigl[(1-\omega) + \omega p_+(x_1,\theta)\bigr] + \sum_{i=2}^{m} f_i \log\bigl[\omega p_+(x_i,\theta)\bigr]. \tag{10}$$

Maximizing (10) for fixed θ leads to $\hat{\omega} = \hat{\omega}(\theta) = (1 - f_1/n)/(1 - p_+(x_1,\theta))$ which is (8) from where (9) follows. Hence the *profile log-likelihood* $\log L_{+1}(\theta) = \log L_{+1}(\hat{\omega}(\theta),\theta)$ is

$$\begin{aligned}\log L_{+1}(\theta) &= f_1 \log(f_1/n) + n_1 \log(1 - f_1/n) - n_1 \log\bigl[1 - p_+(x_1,\theta)\bigr] \\ &\quad + \sum_{i=2}^{m} f_i \log\bigl[p_+(x_i,\theta)\bigr].\end{aligned} \tag{11}$$

Clearly, the second, θ-dependent part of (11) corresponds to

$$\sum_{i=2}^{m} f_i \log\bigl[p_+(x_i,\theta)/\bigl(1 - p_+(x_1,\theta)\bigr)\bigr] = \log L_{++}(\theta),$$

which shows (7) of Theorem 1 and ends the proof.

It is clear that the estimate $\hat{\omega}$ in (8) must be nonnegative. However, it might be larger than one in which case it will be truncated to one and the one-inflation model reduces to the standard zero-truncated model $p_+(x,\theta)$. Hence interest can focus on the case that $0 < \hat{\omega} < 1$.

3. Likelihood ratio test. Clearly, the question of whether there is one-inflation will depend on the choice of $p(x,\theta)$. Hence it needs to be investigated whether there are more singletons than compatible with the baseline density $p(x,\theta)$. A simple likelihood ratio test can be developed from the results of the previous section. Under the alternative model (1) of one-inflation the log-likelihood is given as $\log L_{+1}(\hat{\theta}) = f_1 \log(f_1/n) + n_1 \log(1 - f_1/n) + \log L_{++}(\hat{\theta})$ where $\hat{\theta}$ maximizes the likelihood $L_{++}(\theta) = \prod_{i=2}^{m} p_{++}(x_i,\theta)^{f_i}$. The log-likelihood under the null-hypothesis of no one-inflation is given as $\log L_+(\hat{\theta}_0)$ where $\hat{\theta}_0$ maximizes $\log L_+(\theta) = \sum_{i=1}^{m} f_i \log p_+(x_i,\theta)$. Hence the likelihood ratio statistics is given as

$$\begin{aligned}\lambda &= 2\bigl[\log L_{+1}(\hat{\theta}) - \log L_+(\hat{\theta}_0)\bigr] \\ &= 2\bigl[f_1 \log(f_1/n) + n_1 \log(1 - f_1/n) + \log L_{++}(\hat{\theta}) - \log L_+(\hat{\theta}_0)\bigr].\end{aligned} \tag{12}$$

Care has to be taken when referring to the null distribution of λ. In our case, the one-inflation parameter ω is restricted to be not larger than one. Because the null hypothesis is that this parameter is indeed one, which is on the boundary of the parameter space, the distribution of λ is a 50 : 50 mixture of a one-point distribution which puts all its point mass at zero and a $\chi^2_{(1)}$ distribution with 1 degree of freedom. Therefore, significance levels in the one-parameter case can be adjusted accordingly. See Self and Liang (1987) for the appropriate theory. For example, the p-value can be computed simply by dividing the p-value by 2 which is obtained when applying the conventionally $\chi^2_{(1)}$-distributional result.

To demonstrate the likelihood ratio test for the case of a count density we use the geometric density $p(x, \theta) = (1 - \theta)^x \theta$ for $x = 0, 1, \ldots$. [We will see in Section 6 that the geometric provides a good fit here.] Hence we have that $p_+(x, \theta) = (1 - \theta)^{x-1}\theta$ for $x = 1, 2, \ldots$ and $p_{++}(x, \theta) = (1 - \theta)^{x-2}\theta$ for $x = 2, 3, \ldots$. The relevant log-likelihoods are readily obtained as $\log L_+(\theta) = S_1 \log(1 - \theta) + n \log\theta$ with maximum likelihood estimate $\hat{\theta}_0 = 1/(S_1/n + 1)$ and $\log L_{++}(\theta) = S_2 \log(1 - \theta) + n_1 \log\theta$ and maximum likelihood estimate $\hat{\theta} = 1/(S_2/n_1 + 1)$, where $S_1 = \sum_{x=1}^m f_x(x - 1)$ and $S_2 = \sum_{x=2}^m f_x(x - 2)$. In the case of the DD data, we find a value 117.70 for λ, supporting a strong evidence for one-inflation.

4. Population size estimation. We are interested in estimating the total size N of the population, in the application the total amount of DD in Britain in the observational period. The population size N consists of the observed part n and the unobserved part f_0. The conventional Horvitz–Thompson estimator would estimate N as $\hat{N} = n/[1 - p(x_0, \theta)]$, or in detail $\hat{N} = \sum_{i=1}^N I_i/[1 - p(x_0, \theta)]$ where I_i is the indicator function for the ith unit of the population. Every observed member of the target population is up-weighted by the probability being observed, for each observed unit an estimate of the associated unobserved units is computed.

The problem in the x_1-inflated case is that also the extra-singletons, or in more general terms, the extra-x_1's, would be up-weighted while only the nonextra-singletons (nonextra-x_1's) should be up-weighted. We solve this problem by removing the singletons (x_1's) completely and construct a Horvitz–Thompson estimator for the target population with no extra-singletons (x_1's)

$$\hat{N}_{\text{nes}} = \frac{n_1}{1 - p(x_0, \theta) - p(x_1, \theta)}.$$

$\hat{N}_{\text{nes}}$ is an unbiased estimator of the population size N_{nes} of the target population with no extra-x_1's. Hence we are able to construct an estimator of the hidden units f_0 as

$$\hat{f}_0 = p(x_0, \theta) \frac{n_1}{1 - p(x_0, \theta) - p(x_1, \theta)}.$$

Recall that $n_1 = n - f_1$, the observed sample size reduced by the frequency of singletons. Ultimately, we achieve the Horvitz–Thompson estimator of the target population of interest as

$$\hat{N} = \hat{f}_0 + f_1 + n_1 = p(x_0, \theta) \frac{n_1}{1 - p(x_0, \theta) - p(x_1, \theta)} + f_1 + n_1. \tag{13}$$

Again, this is an unbiased estimator of the population size N of interest. Typically, θ is unknown and is best replaced with the maximum likelihood estimator $\hat{\theta}$ under $p_{+1}(x,\theta)$ leading to

$$\hat{N} = \hat{f}_0 + f_1 + n_1 = p(x_0,\hat{\theta})\frac{n_1}{1 - p(x_0,\hat{\theta}) - p(x_1,\hat{\theta})} + f_1 + n_1. \tag{14}$$

5. Standard errors. Every population size estimator needs to be accompanied by a measure of certainty. To keep the level of generality, we wish to apply the nonparametric bootstrap as discussed, for example, in Efron and Tibshirani (1993). However, bootstrapping is more complex in the capture-recapture setting, as a simple random sample of size n with replacement will typically underestimate the variability of the quantity of interest. In Böhning, van der Heijden and Bunge (2018) an example is given where the conventional bootstrap delivers in all cases the original sample. A correct application of nonparametric bootstrap would imply taking a random sample of *size* N from the observed sample *including* the unobserved zero counts. In Anan, Böhning and Maruotti (2017), this is called the *true* bootstrap. As N is unknown, it is replaced by $\hat{N}$. Hence, a random sample of size $\hat{N}$ with replacement is drawn from the observed sample including the estimated frequency $\hat{f}_0$ of zero counts. This is called the *imputed* bootstrap. In Anan, Böhning and Maruotti (2017) it is demonstrated, by means of extensive simulation work, that the true and imputed bootstrap give an accurate estimate of the true standard error. However, for imputed bootstrap it has to be assumed that the model, under which θ is estimated, is correct. The idea has been originally suggested in van der Heijden et al. (2003) in the Poisson setting. Formally, we apply the bootstrap in this setting as follows.

1. Draw a sample of size $\|\hat{N}\|$ from the observed distribution defined by the probabilities $\frac{\hat{f}_0}{\hat{N}}, \frac{f_1}{\hat{N}}, \frac{f_2}{\hat{N}}, \dots, \frac{f_m}{\hat{N}}$. (Here $\|x\|$ denotes the rounding of x to the nearest integer.)
2. Derive $\hat{\theta}$ and $\hat{N}$ for the bootstrap sample in step 1.
3. Repeat steps 1 and 2 B times, leading to a sample of estimates $N^{(1)}, \dots, N^{(B)}$
4. Calculate the bootstrap standard error as

$$SE^* = \frac{1}{B}\sum_{b=1}^{B}(N^{(b)} - \bar{N^*})^2,$$

where $\bar{N^*} = \frac{1}{B}\sum_{b=1}^{B} N^{(b)}$.

We now apply these concepts to the drink-driving data from Britain.

6. Modeling and application to DD. We have seen that there is overwhelming evidence for a case of one-inflation in the data on drink-driving. Ignoring one-inflation would lead to a serious overestimation of the population size. This can most easily be seen in the case of a Poisson distribution. Ignoring one-inflation would pull the Poisson parameter estimate towards zero and thus increase the population size estimate, potentially quite largely. Fortunately, one-inflation can now easily be coped with using the one-inflation model for any base distribution. As there is perfect fit for the frequency of singletons for any base model, the fit of the model will entirely rest on the base distribution. Modeling the base distribution is an area in itself as the there are a diversity of approaches. We restrict ourselves here on some simple, parametric distributions. If these are valid for the observed and unobserved parts of the data, the associated population size estimates are consistent and asymptotically unbiased under regularity assumptions outlined in Sanathanan (1977). However, simple models like the Poisson are seldom valid models for the count of identifications, as they are not able to cope with heterogeneity in the parameters across the target populations, for example in the case of the DD data there might be differences in the distributional parameter according to the area or time of the day or to demographic characteristics of the driver. Hence models that incorporate these unobserved heterogeneities are likely to be more appropriate. Mixture models are one class of models that naturally incorporate heterogeneity in the modeling process. A typical mixture model for count distributions is

$$\int_0^\infty \exp(-\theta)\theta^x/x!g(\lambda)\,d\lambda, \tag{15}$$

where the mixture kernel is a Poisson density and the mixing distribution is left unspecified with density $g(\lambda)$. We have seen before that specific choices for the mixing density g leads to particular distributions such as the exponential distribution leads to the geometric as marginal distribution, and, more generally, the choice of a gamma distribution leads to the negative-binomial as marginal. We will concentrate on these simple distributions in the following. However, we also want to point out that one can leave g unspecified and estimate it nonparametrically which leads to discrete mixture distributions of the form

$$\sum_{j=1}^{k} \exp(-\theta_j)\theta_j^x/x!g_j, \tag{16}$$

where weights g_j are giving positive mass to a finite number k of subpopulations. Unfortunately, discrete mixtures have to be considered with care as there is a risk of a boundary problem in which case large and spurious population size estimators are generated by the mixture [Wang and Lindsay (2005, 2008)] indicating a consistency problem as the regularity conditions given in Sanathanan (1977) fail to hold for discrete mixture models. See also the comments by McCrea and Morgan [(2015), pages 43, 52]. We will discuss this further below. Nevertheless, it is obvious that a mixture model is more likely to be close to a valid model than a model just based on the mixture kernel, for example, a homogeneous Poisson. However, it is shown in Puig and Kokonendji (2018) and van der Heijden et al. (2003) that,

TABLE 2
Log-likelihood and model selection criteria (AIC and BIC) for drink-driving (DD) data in Britain between 2011 *and* 2015 *for the Poisson, geometric and negative-binomial distribution*

Model	Log-L	AIC	BIC	$\hat{N}$
Poisson	−2127.9	4257.8	4264.9	666,746
Geometric	−2116.8	4235.6	4242.6	2,333,519
Negative-binomial	−2116.8	4237.6	4251.7	2,337,892

in the case of a valid mixture model, the population size estimator based on the homogeneous model provides a lower bound to the true population size. Hence simple models can provide still useful information about the population size even though they might be not valid.

In Table 2, we consider model log-likelihoods as well as the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) for the zero-one-truncated Poisson, geometric, and negative-binomial distribution. Note that the negative-binomial contains both, the Poisson and the geometric, as special cases. For completeness we provide the density function of the negative-binomial with $\theta = (\mu, \alpha)$:

$$p(x, \theta) = \frac{\Gamma(x + \frac{1}{\alpha})}{\Gamma(x + 1)\Gamma(\frac{1}{\alpha})} \left(\frac{1/\alpha}{\mu + 1/\alpha}\right)^{1/\alpha} \left(\frac{\mu}{\mu + 1/\alpha}\right)^{x}$$

for $x = 0, 1, 2, \ldots$ using the mean parameterization, so that $\mu > 0$ is the mean and $\alpha > 0$ is the dispersion parameter. The geometric occurs for $\alpha = 1$ and the Poisson for $\alpha \to 0$.

As Table 2 shows there is clear evidence to support the geometric distribution. The fit of this model confirms evidence for the geometric density as also a 95% confidence interval for the dispersion parameter is found as $0.9998 - 1.0002$ which very narrowly ensconce the value of one which corresponds to the geometric as special case in the negative-binomial distribution.

Despite the difficulties involved in fitting discrete mixture models we have looked into estimating discrete Poisson mixture models for the DD data set. Table 3 shows the results for fitting Poisson mixture models for $k = 1, 2, 3$ components. It appears that a two-component mixture provides good model selection criteria values with an estimate of the population size of 993,790. However, if we consider the model with $k = 3$ components the population size estimate jumps to an unrealistic and spurious value. This occurs as the maximum likelihood estimate attaches high mass of 0.999974 to a Poisson component with very low parameter value of 0.000006 leading to very large estimate of 0.999987 for $p(x_0, \theta)$ generating a largely inflated population size estimate as can be seen from (14). For more details on this aspect see Wang and Lindsay (2005), pages 943–944.

TABLE 3
Log-likelihood and mixture model selection criteria (AIC and BIC) for drink-driving (DD) data in Britain between 2011 and 2015 using $k = 1, 2, 3$ subpopulations

k	Log-L	AIC	BIC	$\hat{N}$
1	−2127.9	4257.8	4264.9	666,746
2	−2106.3	4218.7	4239.8	993,790
3	−2106.2	4222.3	4257.6	7.66×10^9

As we reach with the BIC-value of 4242.64 for the geometric distribution a value which compares favorable with all others, we will use the geometric distribution for the further inference. The two-component Poisson model achieves also a BIC-value close to the BIC-value of the geometric model, but uses two more parameters to achieve this. Note also that the BIC is the better suitable criterion for the selection of mixture models in comparison to the AIC [Keribin (2000), Ray and Lindsay (2008)]. Hence we are focusing here on the BIC for the selection of the model. The geometric distribution provides also a reasonable fit here as Table 4 shows, at least clearly better than those achieved by fitting a Poisson distribution. Only observed and fitted values are shown for x larger than one as the distribution is one-inflated. The improved fit of the geometric distribution upon the Poisson distribution might be explained as it occurs as a mixture of the Poisson with an exponential distribution as was mentioned previously. Hence the geometric is able to adjust for some of the potential, unobserved heterogeneity involved in a Poisson distribution.

The estimate of the total of DD in Britain in the observational period is found to be 2,333,519. This corresponds to 9.7% observed DD, so that about 90% of DD remains hidden. We use the bootstrap with $B = 10{,}000$ to find a 95% confidence interval. The histogram of the bootstrap distribution of $\hat{N}$ is given in Figure 1 and appears fairly symmetric so that a normal-based confidence interval seems reasonable and is provided as 1,975,820–2,727,610, corresponding to a 95% confidence interval of 8.7%–12.0% of the observed DD. The associate percentiles based 95% confidence interval is 2,008,895–2,756,244, corresponding to a 95% confi-

TABLE 4
Observed and fitted distributions for the count x of repeated drink-driving identifications

x	Observed	Geometric	Poisson
2	8068	8040.8	8032.5
3	449	496.5	512.1
≥ 4	53	32.7	25.5
χ^2		17.3	37.8

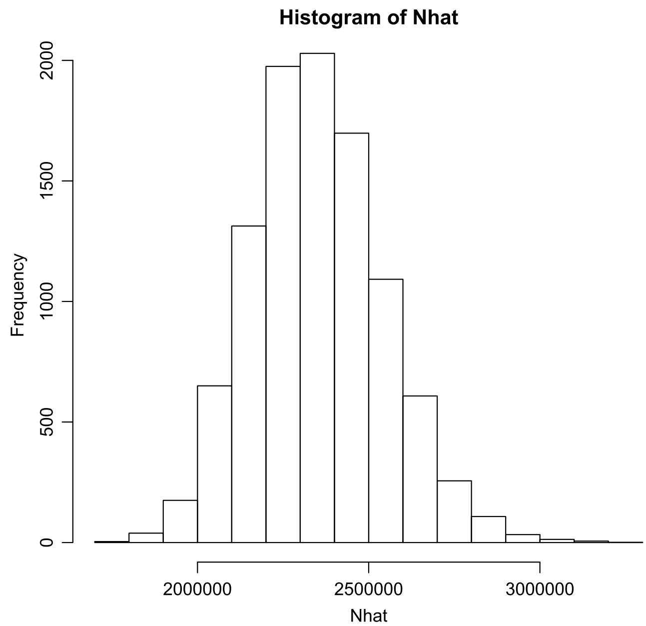

FIG. 1. *Histogram of the bootstrap distribution of the population size estimator.*

dence interval of 8.8%–12.1% of the observed DD which is almost identical to the normal-based confidence interval.

In the following, we are taking some plausibility checks on the estimated 9.7% of observed DD. In 2017, 74% of the population of England over 17 years of age were in the possession of a valid driving license leading to a number of 32.9 million [Department of Transport (2018)]. For Wales, this number corresponds to 2.3 million, so that in total we have 35.9 million driving licenses. Using a growth rate of 1% we achieve for the 5-year period from 2011 to 2015 a number of $34.3 + 33.9 + 33.6 + 33.3 + 32.9 = 168$ million drivers with valid driving licenses. Hence we can estimate the proportion of DD among the licensed population in Britain as $2.3/168 \times 100\% = 1.37\%$. In a similar period, namely for the year 2016, road traffic casualties of all severity are given as 181,386 for Great Britain of which 6080 are determined as DD-accidents leading to a percentage of 3.35% of DD-driving among those with accidents (as for all those accidents samples on alcohol levels are taken) [Road accidents and safety statistics (2018)]. Clearly, our estimated percentage of drink-driving is about 2.5 times lower than the number derived from the road traffic accident statistics. However, drivers with an accident, in particular those with several accidents, might not be representative for the general population of licensed drivers. In addition, drink-driving and accident occurrence are positively associated, so that it is quite plausible that the percentage from the road traffic accident statistics lies higher than our estimate. On the other hand, the denominator used in determining the 1.37% percent drink-driving among the general population might be too large as it will include drivers that rarely drive or do not drive at all and are hence not prone to an accident. All in all, the derived number of 2.3 million drink-drivers in the 5-year period or 460,000 p.a. do not appear unrealistic.

In the article of the Guardian on the topic [The Guardian (2016)], the president of the AA, the British Automobile Association, was cited as follows:

> The fact that more than 8000 drivers have been caught twice in five years is all the more astonishing when they should have been off the road for a year or more.

In addition, it is clearly potentially even more worrisome that a large number of uncaught individuals exists as our model suggests that only 10% of DD is seen.

7. Discussion. We think that the result in Theorem 1 is interesting and useful as it allows inference to focus on the x_0-x_1-truncated part, for count densities the zero-one truncated part, of the distribution. If the latter is fairly simple, inference can be straightforward as shown in the application of DD in Britain.

The argument used in proving Theorem 1 shows also that for any discrete density, with one value x_1 inflated but not necessarily another one truncated, the inflated likelihood could be constructed from the x_1-truncated likelihood. For example, a zero-inflated Poisson or binomial likelihood could be found from the associated zero-truncated Poisson or binomial likelihood. Unfortunately, Theorem 1 does not generalize for continuous densities as truncation with the associated truncated densities can only be defined for regions with positive mass and not for single points as it would be required.

Coming to the application study, we use the geometric distribution as it provides an acceptable fit, in fact, considerably better than the Poisson in this case. In addition, we find often that the geometric does provide reasonable fits in these kind of applications. We see one reason for this finding in the fact that the geometric is an exponential mixture of a Poisson density, hence it adjusts already for some unobserved heterogeneity of the target population of drink-drivers in a natural way.

We need to emphasize that the population size estimate depends on the specification of the model. The appropriateness of this distributional model can only be evaluated on the basis of the observed counts, and the model, even if it provides an acceptable fit, might not be appropriate for the unobserved part. Hence it is always advisable to use other mechanisms, such as the plausibility check we have done in the case study here, to investigate if the achieved population size is reasonable at all. In addition, it would be helpful if covariate information would be available which could improve the modelling considerably. Even demographic information on age and gender composition would be useful to obtain as it is known that traffic accident risks are differential with age and gender. We recommend that routinely collected data by the police on DD do record these and other characteristics and that these can be obtained on request. Clearly, especially with capture-recapture data, it is important to be aware of the assumptions underlying the approach and its limitations, especially, if its application is considered for other case studies in different settings.

Acknowledgements. We are grateful to the Editor, the Associate Editor, and two reviewers for several helpful comments.

REFERENCES

ANAN, O., BÖHNING, D. and MARUOTTI, A. (2017). Uncertainty estimation in heterogeneous capture-recapture count data. *J. Stat. Comput. Simul.* **87** 2094–2114. MR3647589

BÖHNING, D., BUNGE, J. and VAN DER HEIJDEN, P. G. M. (2018). Uncertainty assessment in capture-recapture studies and the choice of sampling effort. In *Capture-Recapture Methods for the Social and Medical Sciences* (D. Böhning, P. G. M. van der Heijden and J. Bunge, eds.) 389–394. CRC Press/CRC, Boca Raton, FL.

BÖHNING, D., VAN DER HEIJDEN, P. G. M. and BUNGE, J. (2018). *Capture–Recapture Methods for the Social and Medical Sciences*. CRC Press/CRC, Boca Raton, FL.

BORCHERS, D. L., BUCKLAND, S. T. and ZUCCHINI, W. (2004). *Estimating Animal Abundance. Closed Populations*. Springer, Heidelberg.

BUNGE, J. and FITZPATRICK, M. (1993). Estimating the number of species: A review. *J. Amer. Statist. Assoc.* **88** 364–373.

BUNGE, J., WILLIS, A. and WALSH, F. (2014). Estimating the number of species in microbial diversity studies. *Ann. Rev. Stat. Appl.* **1** 427–445.

DEPARTMENT OF TRANSPORT (2018). Driving Licences. Available at https://www.ethnicity-facts-figures.service.gov.uk/culture-and-community/transport/driving-licences/latest, accessed 1/10/2018.

EFRON, B. and TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability* **57**. CRC Press, New York. MR1270903

GODWIN, R. T. and BÖHNING, D. (2017). Estimation of the population size by using the one-inflated positive Poisson model. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **66** 425–448. MR3611695

KERIBIN, C. (2000). Consistent estimation of the order of mixture models. *Sankhyā Ser. A* **62** 49–66. MR1769735

MCCREA, R. S. and MORGAN, B. J. T. (2015). *Analysis of Capture-Recapture Data. Chapman & Hall/CRC Interdisciplinary Statistics Series*. CRC Press, Boca Raton, FL. MR3330977

PUIG, P. and KOKONENDJI, C. C. (2018). Non-parametric estimation of the number of zeros in truncated count distributions. *Scand. J. Stat.* **45** 347–365. MR3803593

RAY, S. and LINDSAY, B. G. (2008). Model selection in high dimensions: A quadratic-risk-based approach. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **70** 95–118. MR2412633

ROAD ACCIDENTS AND SAFETY STATISTICS (2018). Statistics and data about reported accidents and casualties on public roads. Available at https://www.gov.uk/government/collections/road-accidents-and-safety-statistics. Accessed 1/10/2018.

SANATHANAN, L. (1977). Estimating the size of a truncated sample. *J. Amer. Statist. Assoc.* **72** 669–672. MR0458693

SELF, S. G. and LIANG, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.* **82** 605–610. MR0898365

THE GUARDIAN (2016). More than 8000 people caught drink-driving twice in five years. (accessed 23/5/2018), in print p. 11. Available at www.theguardian.com/society/2016/dec/30/more-than-8000-people-caught-drink-driving-twice-in-five-years.

VAN DER HEIJDEN, P. G. M., BUSTAMI, R., CRUYFF, M. J. L. F., ENGBERSEN, G. and VAN HOUWELINGEN, H. C. (2003). Point and interval estimation of the population size using the truncated Poisson regression model. *Stat. Model.* **3** 305–322. MR2012155

WANG, J.-P. Z. and LINDSAY, B. G. (2005). A penalized nonparametric maximum likelihood approach to species richness estimation. *J. Amer. Statist. Assoc.* **100** 942–959. MR2201021

WANG, J.-P. and LINDSAY, B. G. (2008). An exponential partial prior for improving nonparametric maximum likelihood estimation in mixture models. *Stat. Methodol.* **5** 30–45. MR2416938

SOUTHAMPTON STATISTICAL SCIENCES
RESEARCH INSTITUTE
UNIVERSITY OF SOUTHAMPTON
HIGHFIELDS CAMPUS
SOUTHAMPTON, SO17 1BJ
UNITED KINGDOM
E-MAIL: d.a.bohning@soton.ac.uk