National Institute for Applied Statistics Research Australia

University of Wollongong, Australia

Working Paper

18-23

A Comparison of Likelihood-Based Methods for Size-Biased Sampling

Victoria L. Leaver, Robert G. Clark, Pavel N. Krivitsky, and Carole L. Birrell

National Institute for Applied Statistics Research Australia, University of Wollongong, Wollongong NSW 2522, Australia T: +61 2 42215076. E: karink@uow.edu.au

A comparison of likelihood-based methods for size-biased sampling

Victoria L. Leaver[a,*], Robert G. Clark[b], Pavel N. Krivitsky[c], Carole L. Birrell[d]

[a]*School of Mathematics and Applied Statistics, University of Wollongong, New South Wales 2522, Australia*
[b]*Research School of Finance, Actuarial Studies and Statistics , Australian National University, Australian Capital Territory 2601, Australia*
[c]*School of Mathematics and Statistics, University of New South Wales, New South Wales 2052, Australia*
[d]*School of Mathematics and Applied Statistics, University of Wollongong, New South Wales 2522, Australia*

Abstract

Three likelihood approaches to estimation under informative sampling are compared using a special case for which analytic expressions are possible to derive. An independent and identically distributed population of values of a variable of interest is drawn from a gamma distribution, with the shape parameter and the population size both assumed to be known. The sampling method is selection with probability proportional to a power of the variable with replacement, so that duplicate sample units are possible. Estimators of the unknown parameter, variance estimators and asymptotic variances of the estimators are derived for maximum likelihood, sample likelihood and pseudo-likelihood estimation. Theoretical derivations and simulation results show that the efficiency of the sample likelihood approaches that of full maximum likelihood estimation when the sample size n tends to infinity and the sampling fraction f tends to zero. However, when n tends to infinity and f is not negligible, the maximum likelihood estimator is more efficient than the other methods because it takes the possibility of duplicate sample units into account. Pseudo-likelihood can perform much more poorly than the other methods in some cases. For the special case when the superpopulation is exponential and the selection is probability proportional to size, the anticipated variance of the pseudo-likelihood estimate is infinite.

Keywords: Maximum likelihood estimation, sample likelihood estimation, pseudo-likelihood estimation, model-based inference, analysis of complex surveys

---

*Corresponding author
Email addresses: `victoria.leaver@abs.gov.au` (Victoria L. Leaver), `robert.clark@anu.edu.au` (Robert G. Clark), `p.krivitsky@unsw.edu.au` (Pavel N. Krivitsky), `cbirrell@uow.edu.au` (Carole L. Birrell)

1. Introduction

Maximum likelihood, sample likelihood and pseudo-likelihood estimation are three frequentist approaches to survey estimation. Maximum likelihood estimation is a standard approach that uses the likelihood function to estimate unknown parameter values from observed data. The sample likelihood method maximizes a likelihood function calculated from a weighted distribution function that incorporates the selection probabilities; it is equivalent to the product of the densities conditional on each unit having been selected in the sample. The widely used pseudo-likelihood method maximises an inverse-probability-weighted estimator of the log-likelihood that would have been obtained if all the population units were selected. Chambers et al. (2012) advocates for the use of full maximum likelihood on the grounds of statistical efficiency. Maximum likelihood and sample likelihood are both straightforward for so-called ignorable designs such as when the selection probabilities depend only on the covariate in a regression model. However, survey data are now frequently being combined with information from other sources such as adminstrative collections, leading to informative designs becoming more common (for example, see Kopra et al. 2015 and Gray et al. 2020). See Sugden and Smith (1984) for a discussion of implications of informative designs on likelihood estimation.

"Informative designs" can be defined as cases where the distributions of the observed variables, the sample inclusion probabilities and the design variables are jointly dependent (page 301 of Chambers et al. 2012). For such designs, estimators for the full maximum likelihood approach can be very difficult to derive in practice, even for relatively simple cases (for example, see Section 3.6.2 of Pfeffermann 2011 and Clark 2020). Sample likelihood is a simpler alternative to full maximum likelihood, but it is not always clear how efficient the sample likelihood estimator will be. The pseudo-likelihood estimator may be less efficient than both full maximum likelihood and sample likelihood (for example, see the "Additional Remarks" section in Lawless 1997), but in many cases it is used because it is the easiest method to implement, with options to apply pseudo-likelihood estimation available in statistical software such as the `survey` package in R (Lumley, 2011). There are currently no general theoretical results on the relative efficiencies of the three approaches, which makes the choice of estimation method even more difficult.

Other approaches have also been suggested, such as weight smoothing (Beaumont, 2008). However, these approaches will generally reduce to pseudo-likelihood in our assumed design, where probabilities are a function of the variable of interest. Another possible approach is empirical likelihood (Rao and Wu, 2009).

This paper presents a case where the relative efficiencies of the three approaches (maximum likelihood, sample likelihood and pseudo-likelihood) can be directly compared. In this case, we estimate the superpopulation parameter θ when a population is drawn from the gamma distribution with rate θ and known shape α. A sample is then drawn with probability proportional to a power of the size y, with replacement. This is an example of a case where there is limited information available to the analyst, because the population parameter θ and therefore the sample inclusion probabilities are assumed to be unknown. This relatively simple example will yield insights into which of the three approaches may be most appropriate for more complex models or designs. Our set-up is related to that of the example in Section 2.3.3 of Chambers et al. (2012), but they also assumed knowledge of at least one inclusion probability, which implies knowledge of all population values. This is a degenerate scenario, as the sample design then becomes irrelevant. In contrast, we assume only that it is known that the inclusion probabilities are proportional to a power of the variable of interest. In our example, the assumption of the known shape parameter α was made so that tractable results could be obtained. However, there are real-life situations where this assumption is reasonable (for example, see Borgos et al. 2002, discussed later in this paper). In practice, it may be necessary to estimate α from the sample data as an initial step.

For the assumed models and designs, we derive asymptotic anticipated variances for all three approaches so that the performance of the estimators can be compared. The anticipated variance is defined as the variance over both probability sampling and realisation of the population from an assumed model; it is an appropriate measure for comparing probability-based and model-based estimators.

This case serves as an illustration of the potential complexity of applying full maximum likelihood methods to even relatively simple examples. However, the comparison of the full maximum likelihood results with results from the sample likelihood and pseudo-likelihood demonstrates that full maximum likelihood estimation can potentially provide better results in practice. In this example, the with replacement sampling approach has important implications for the performance of the different methods, particularly when the sampling fraction is high. The size-biased sampling approach also affects the performance of the pseudo-likelihood estimator, which would not be recommended in this case, given that its anticipated variance can be infinite, and the variance of the estimates obtained from simulations is very high.

Although this case is relatively simple to allow theoretical results to be derived, there are some related real-life problems in the natural sciences literature. For example, Solow and Smith (1997) describe a model in which mean taxonomic duration is assumed to have an exponential distribution and the discovery of preserved finds in the fossil record is assumed to follow a Poisson process. Borgos et al. (2002) find that the fault sizes in a data set of geological faults is "best described by the exponential distribution", and the probability of discovering faults may be proportional to size, depending on the technological method used to scan the ocean floor.

Section 2 describes the assumed model and the notation. Sections 3, 4 and 5 present the estimators and theoretical anticipated variances for maximum likelihood, sample likelihood and pseudo-likelihood respectively. Section 6 compares the performance of the three methods using results from a simulation study. Section 7 applies the sample likelihood and pseudo-likelihood approaches to the geological fault data in Borgos et al. (2002). Section 8 discusses the theoretical properties of the estimators, the simulation results and the case study.

2. Assumed model and notation

Suppose that the population values $\{y_i : i = 1, \ldots, N\}$ are drawn independently and identically from a gamma distribution with shape parameter α and rate parameter θ. It is assumed that α is fixed and known, while the aim is to estimate θ. The probability density function for this distribution is:

$$f(y) = \frac{\theta^{\alpha} y^{\alpha-1} e^{-\theta y}}{\Gamma(\alpha)} \quad 0 \leq y < \infty \, .$$

There is no auxiliary information. The inclusion probabilities are unknown, but the population size N is known. A sample of size n units is selected with replacement, with the probability of selection of a unit y_i being proportional to a power m of y_i, where m is assumed to be known. We observe only the $n \times 1$ vector of values for the sampled units, y_s, which may contain duplicates. While the value y_i is observed for each sampled unit, the probability of selection for y_i remains unknown, because the constant of proportionality is unobserved.

Figure 1 is an illustrative example of the assumed model, showing the theoretical gamma distribution, the expected distribution of the sample, a single population drawn from the gamma distribution and four different samples selected from the single population. For the case shown by Fig. 1, $m = 1$, and so larger population units are more likely to be selected in the samples. Using the framework presented in Patil and Ord (1976), the expected distribution of the sample is calculated by deriving the population distribution $f(y)$ weighted by the selection probability. For the $m = 1$ case, the weighted distribution is $f_w(y) = y f(y) / \int y f(y)\, dy$, which equals

$y\theta^{\alpha} y^{\alpha-1} e^{-\theta y}\theta / (\Gamma(\alpha)(\alpha+1)) = \theta^{\alpha+1} y^{\alpha} e^{-\theta y}/\Gamma(\alpha+1)$ (see Section 4.4 of Pfeffermann et al. 1998). This is the gamma distribution with parameters $(\theta, \alpha + 1)$, and it is shown by the solid line in Fig. 1.

The expected number of times that the ith unit is selected is $\pi_i \propto y_i^m$. The probability that the ith population unit is the jth unit selected in the sample is π_i/n for any j.

Throughout this paper, U refers to the set of units in the population, S contains the sampled units including any duplicates, S_d is the set of distinct units in the sample, and R is the set of units in U that are not in S_d. The $N \times 1$ vector of units in the population is written as y_U. The number of distinct units in the sample is denoted n_{s_d}, and y_{s_d} is the $n_{s_d} \times 1$ vector of distinct units in the sample, while the number of population units not selected in the sample is denoted n_r, and y_r is the vector of non-sampled units in the population. The sum of the values of the sampled units is denoted t_s, t_{s_d} is the sum of the values of the distinct units in the sample, t_r is the sum of the values of the non-sampled units, $t_U = t_{s_d} + t_r$ is the sum of the values of the population units, and $\bar{y}_U$ is the population mean. The design expectation is denoted by E_p and the model expectation is denoted by E_m (see Särndal et al. 1992 for definitions of design expectation and model expectation), and the corresponding variances are denoted var_p and var_m. The anticipated variance is the variance evaluated over both the model and the design, and it can be decomposed as $E_m\left(\text{var}_p\right) + \text{var}_m\left(E_p\right)$. Throughout this paper, integrals are evaluated from 0 to ∞ unless otherwise stated.

While the assumption that α is known may be considered to be a limitation of the theoretical results derived in this paper, in practice it is possible to obtain a reasonably good estimate of α from the observed data, using properties of the gamma distribution such as the result derived in Section 4.4 of Pfeffermann et al. (1998).

3. Maximum Likelihood Estimation

The missing information principle (Orchard and Woodbury, 1972) is often used to calculate the maximum likelihood estimate. However, for the model described in Section 2, the score function for the sample can be derived directly.

Theorem 1. *In the maximum likelihood case, the score function for a given sample y_s is*

$$\text{sc}(\theta) = \frac{\int \left\{ \prod_{k \in R} \frac{\theta^{\alpha} y_k^{\alpha-1} e^{-\theta y_k}}{\Gamma(\alpha)} \right\} \left(\frac{1}{\sum_{k \in U} y_k^m} \right)^n \left(N\alpha\theta^{-1} - t_{s_d} - t_r \right) dy_r}{\int \left\{ \prod_{k \in R} \frac{\theta^{\alpha} y_k^{\alpha-1} e^{-\theta y_k}}{\Gamma(\alpha)} \right\} \left(\frac{1}{\sum_{k \in U} y_k^m} \right)^n dy_r} . \quad (1)$$

For the proof of Theorem 1, see Appendix A.

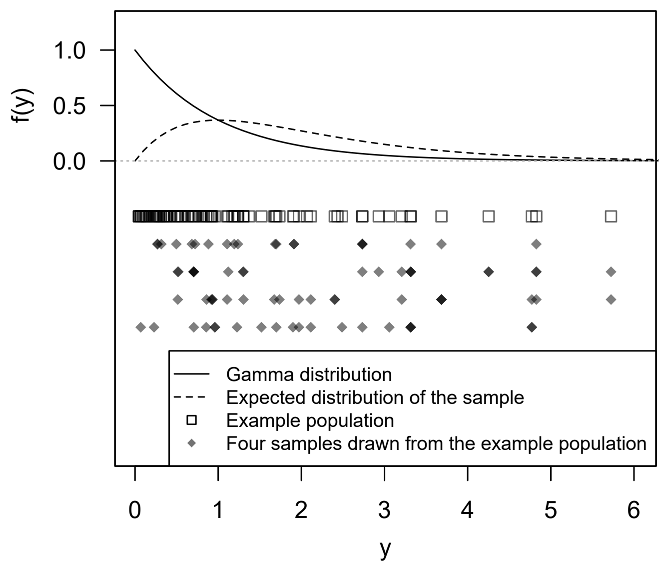

Figure 1: An illustrative example of the assumed model. The gamma distribution used for this example has parameters $\theta = 1$ and $\alpha = 1$, and the sampling scheme uses $m = 1$ (probability proportional to size selection). The population size is 100 and the sample size is 20.

The score function in (1) can also be expressed in the following form, which is useful for setting up simulations:

$$\mathrm{sc}(\theta) = \frac{E_{y_r}\left\{\left(\frac{1}{\sum_{k\in U} y_k^m}\right)^n \left(N\alpha\theta^{-1} - t_{s_d} - t_r\right) : y_r \overset{\text{iid}}{\sim} \mathrm{Gamma}(\alpha,\theta)\right\}}{E_{y_r}\left\{\left(\frac{1}{\sum_{k\in U} y_k^m}\right)^n : y_r \overset{\text{iid}}{\sim} \mathrm{Gamma}(\alpha,\theta)\right\}} . \tag{2}$$

Here, the notation

$$E_{y_r}\left\{\ldots : y_r \overset{\text{iid}}{\sim} \mathrm{Gamma}(\alpha,\theta)\right\}$$

is shorthand for the integral with respect to y_r weighted by this gamma density. It does not imply that y_r follows this distribution; it does not, because of the informative selection.

The maximum likelihood estimate for θ is obtained by setting the score function in (1) to 0.

In practice, we can estimate θ by using numerical optimisation to find a value of $\hat{\theta}$ that makes the right hand side of (1) as close as possible to 0.

In the special case where $m = 1$, we immediately obtain by elementary operations:

$$\mathrm{sc}(\theta) = N\alpha\theta^{-1} - t_{s_d} - \left\{\int \frac{t_r^{n_r\alpha-1} e^{-\theta t_r}}{(t_{s_d} + t_r)^n}\, dt_r\right\}^{-1} \int \frac{t_r^{n_r\alpha} e^{-\theta t_r}}{(t_{s_d} + t_r)^n}\, dt_r\, . \tag{3}$$

The observed information for a given sample and value of θ,

$$\mathrm{info}_s(\theta) = -\frac{\partial}{\partial\theta}\,\mathrm{sc}_s(\theta), \tag{4}$$

can be derived explicitly from the score function, or it can be calculated by numerical differentiation.

3.1. Fisher information for the Maximum Likelihood Estimator

In this section, the Fisher or expected information is derived for the maximum likelihood estimator. The derivation uses the approach of scaling the sample values of y by θ, with the scaled units being denoted by $z_i = \theta y_i$. The sum of the distinct units in the scaled sample z_s is denoted by $t_{z_{s_d}}$, the sum of the scaled non-sampled units in z_r is denoted by t_{z_r} and the sum $\sum_{k\in U} z_k^m$ is denoted by t_{Z^m}.

Theorem 2. *In the maximum likelihood case, the Fisher information is*

$$\mathcal{I}(\theta) = \theta^{-2} E_{z_s} \left(\left[\frac{E_{z_r} \left\{ \frac{\left(N\alpha - t_{z_{s_d}} - t_{z_r}\right)}{(t_{Z^m})^n} : z_r \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, 1) \right\}}{E_{z_r} \left\{ \frac{1}{(t_{Z^m})^n} : z_r \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, 1) \right\}} \right]^2 \right) \tag{5}$$

where $z_s = \theta y_s$ are the scaled sampled units, t_{z_r} is the sum of the scaled non-sampled units $z_r = \theta y_r$, $t_{z_{s_d}}$ is the sum of the distinct units in the scaled sample, and n_d is the number of distinct units in the sample.

See Appendix A for the proof. This result shows that the Fisher information for the maximum likelihood case depends on θ only as a scaling parameter.

For $m = 1$, the Fisher information is

$$\mathcal{I}(\theta) = \theta^{-2} E_{z_s} \left(\left[\frac{E_{t_{z_r}} \left\{ \frac{\left(N\alpha - t_{z_{s_d}} - t_{z_r}\right)}{\left(t_{z_{s_d}} + t_{z_r}\right)^n} : t_{z_r} \sim \text{Gamma}(n_r\alpha, 1) \right\}}{E_{t_{z_r}} \left\{ \frac{1}{\left(t_{z_{s_d}} + t_{z_r}\right)^n} : t_{z_r} \sim \text{Gamma}(n_r\alpha, 1) \right\}} \right]^2 \right). \tag{6}$$

Also, for the $m = 1$ case, if $N \to \infty$ while n and the other parameters remain fixed,

$$\mathcal{I}(\theta) \to \theta^{-2} n(\alpha + 1). \tag{7}$$

For a heuristic derivation of (7), see Appendix A.

4. Sample Likelihood

The sample likelihood method maximizes a likelihood function calculated from a weighted distribution function that incorporates the selection probabilities (Patil and Ord 1976, Krieger and Pfeffermann 1992). This method has been used in a wide range of applications, including multilevel modelling (Pfeffermann et al., 2006).

4.1. Sample likelihood: estimator and theoretical anticipated variance

We begin by calculating

$$\text{pr}\,(I_i = 1) = E\,\{\text{pr}\,(I_i = 1 \mid y_i)\} = nE\left(\frac{ny_i^m}{\sum_{j\in U} y_j^m}\right).$$

To proceed with sample likelihood, it is necessary to assume that the denominator $\sum_{j\in U} y_j^m$ can be treated as a constant (Krieger and Pfeffermann, 1992). This assumption implies that $\text{pr}(I_i = 1) \propto E\left(y_i^m\right) = \Gamma(m+\alpha)/(\Gamma(m)\theta^m)$. Following Krieger and Pfeffermann (1992), the sample likelihood is defined as

$$L_s = \prod_{i \in s} \frac{\text{pr}(I_i = 1 \mid y_i) f(y_i)}{\text{pr}(I_i = 1)} \propto \prod_{i \in s} \frac{\theta^m \Gamma(m) y_i^m \frac{\theta^\alpha}{\Gamma(\alpha)} y_i^{\alpha-1} e^{-\theta y_i}}{\Gamma(m+\alpha)} = \prod_{i \in s} \frac{\theta^{\alpha+m}\Gamma(m) y_i^{\alpha+m-1} e^{-\theta y_i}}{\Gamma(m+\alpha)\Gamma(\alpha)} . \tag{8}$$

Taking the logarithm of (8) gives

$$\log L_s = \log(K) + \sum_{i \in s} \left\{ (\alpha+m)\log(\theta) + \log\Gamma(m) + \log\left(y_i^{\alpha+m-1}\right) - \theta y_i - \log\Gamma(m+\alpha) - \log\Gamma(\alpha) \right\} \tag{9}$$

where K is a constant. Differentiating (9) with respect to θ gives the score function, which is solved to obtain the sample likelihood estimate of θ, denoted by $\hat{\theta}_{SL}$:

$$\sum_{i \in s} \left(\frac{\alpha+m}{\hat{\theta}_{SL}} - y_i \right) = 0 \qquad \Rightarrow \qquad \hat{\theta}_{SL} = \frac{\alpha+m}{\bar{y}_s} \tag{10}$$

where $\bar{y}_s$ is the mean of the sample units y_i.

The information function is:

$$\text{info}_{y_s} = -\frac{\partial\, \text{sc}(\theta)}{\partial \theta} = -\frac{\partial}{\partial \theta} \sum_{i \in s} \left(\frac{\alpha+m}{\theta} - y_i \right) = \frac{n(\alpha+m)}{\theta^2} . \tag{11}$$

An estimated variance for the sample likelihood estimator can be calculated by taking the inverse of (11). However, this formula for the variance does not take into account the loss of information due to the possibility of selecting duplicate units in the sample. The anticipated variance will provide a more accurate measure, especially if the sampling fraction is high.

The theoretical anticipated variance of the sample likelihood estimate of θ is

$$AV\left(\hat{\theta}_{SL}\right) \approx \frac{\theta^2}{(\alpha+m)^2} \left(\frac{\alpha+m}{n} + \frac{3m^2 + 2m\alpha + 2m + \alpha}{N} \right) . \tag{12}$$

See Appendix B for the derivation of (12).

For the $m = 1$ case, when $N \to \infty$ while the other parameters remain fixed, $AV_{SL} \to \theta^2/n(\alpha+1)$.

5. Pseudo-Likelihood

The aim of the pseudo-likelihood approach is to "construct a design-consistent estimate" (Chambers et al., 2012) of the score function that would have been obtained if the population units had been selected, rather than just a sample. The approach was discussed in the context of deriving estimators for complex surveys by Binder (1983), and Godambe and Thompson (1986) presented further theoretical properties.

5.1. Pseudo-likelihood: estimator and theoretical anticipated variance

Let $\hat{\theta}_{PL}$ denote the pseudo-likelihood estimate of θ. The probability weighted score function is

$$\mathrm{sc}_w(\theta) = \sum_{i\in S} \frac{\sum_{j\in S_d} y_j^m + \sum_{k\in R} y_k^m}{n y_i^m} \left(\frac{\alpha}{\theta} - y_i\right) \tag{13}$$

and so

$$\begin{aligned} 0 &= \frac{\alpha}{\hat{\theta}_{PL}} \sum_{i\in S} \frac{1}{y_i^m} - \sum_{i\in S} y_i^{1-m} \\ \Rightarrow \qquad \hat{\theta}_{PL} &= \alpha \frac{\sum_{i\in S} y_i^{-m}}{\sum_{i\in S} y_i^{1-m}} . \end{aligned} \tag{14}$$

The theoretical anticipated variance of $\hat{\theta}_{PL}$ can be approximated as follows:

$$\begin{aligned} AV\left(\hat{\theta}_{PL}\right) \approx \frac{\theta^2}{n}\Bigg[& 1 + \frac{\Gamma(\alpha-m)\Gamma(\alpha+m)}{\Gamma(\alpha)^2} - 2\frac{\Gamma(\alpha+m)\Gamma(\alpha-m+1)}{\Gamma(\alpha)\Gamma(\alpha+1)} \\ & + \frac{\Gamma(\alpha-m+2)\Gamma(\alpha+m) + \left\{(n-1)N^{-1}\alpha - \alpha^2\right\}\Gamma(\alpha)^2}{\Gamma(\alpha+1)^2}\Bigg]. \end{aligned} \tag{15}$$

See Appendix C for the derivation of (15).

For the $m = 1$ case, as $N \to \infty$ while the other parameters remain fixed, $AV_{PL} \to \theta^2/n(\alpha-1)$ (from Corollary 7 in Appendix C).

If $X \sim \mathrm{Gamma}(\alpha, \theta)$ and $m \geq \alpha$, then $E\left[X^{-m}\right]$ is infinite (see Lemma 8 in Appendix C), and so the theoretical anticipated variance from (15) is also infinite. In particular, for the case where the population is drawn from the exponential distribution and the selection is probability proportional to size, the anticipated variance for the pseudo-likelihood estimator is infinite.

6. Simulation results

6.1. Design of simulation study

This section presents results from a simulation study that explores the performances of the maximum likelihood, sample likelihood and pseudo-likelihood estimators as the sampling fraction $f = n/N$ and other parameters are varied. The results in this section have been obtained using the `R` statistical analysis software. In each iteration of the simulation, a population is drawn from the gamma distribution and a sample is drawn with replacement from the population.

For the maximum likelihood case, numerical optimisation and integration methods were used for the calculations. The `integrate` function in `R` was used to obtain values for the integrals, and the `optimize` function was used to find $\hat{\theta}$ from (1). Due to computational limitations, Simpson's rule was used to approximate integral values in some cases. The theoretical anticipated variance for the maximum likelihood case is the inverse of the Fisher information $\mathcal{I}$, which has been calculated using Monte Carlo integration to evaluate (2). The theoretical anticipated variance for the sample likelihood case is (12). For the pseudo-likelihood case, the theoretical anticipated variance is (15).

The weight coefficient of variation has been used in the literature to derive measures of the impact on the variance of an estimate caused by variation of the survey weights (Levy and Lemeshow, 2008). The weight coefficient of variation may be useful as a guide to when the results presented in this paper may be indicative of the relative efficiencies of different likelihood approaches in more complex survey designs. The weights are considered to be the inverse of the selection probabilities. When $\alpha = 1$ and $m = 1$, the weight coefficient of variation ranges from about 1 (for $n = 20$ and $N = 100$) to 1.3 (for $n = 50$ and $N = 1000$) in these simulations, while when $\alpha = 3$ and $m = 1$, the weight coefficient of variation ranges between 0.6 to 0.7, and when $\alpha = 1$ and $m = 2$, the weight coefficient of variation ranges between 1.7 to 2.2. For comparison, the weight coefficient of variation for the first cohort (initially surveyed during October 1999 to December 2000) of the National Survey of Child and Adolescent Wellbeing is 1.92 (Levy and Lemeshow, 2008, pp. 509-512). This suggests that $\alpha = 1$ and $m = 2$ may be a realistic scenario.

6.2. Estimates, variances, estimated variances and theoretical anticipated variances for $\alpha = 1$, $m = 1$

The results in Table 1 show the estimates, simulation variances and theoretical anticipated variances for maximum likelihood, sample likelihood and pseudo-likelihood for the case when $\alpha = 1$ and $m = 1$. These results have been calculated from 10,000 simulations. This table provides a comparison of the estimated variance and the theoretical antipicated variance with the variance of the values of $\hat{\theta}$ from the simulations. For the $\alpha = 1$ and $m = 1$ case, the pseudo-likelihood theoretical antipicated variance is infinite.

The maximum likelihood estimated variance in Table 1 is the inverse of the observed information from (4), which was calculated using numerical differentiation with the `grad` function in the `numDeriv` package (Gilbert and Varadhan, 2016). The sample likelihood estimated variance is the mean of the inverse of the observed information from (11) across the simulations. As discussed in subsection 4.1, this estimated variance will generally underestimate the theoretical anticipated variance. The pseudo-likelihood estimated variance is calculated by taking the mean across the simlulations of the design variance $\text{var}_p\left(\hat{\theta}\right)$, which equals $\left(\alpha^2/n\right)\left[t_Y^{-1}\sum_{k\in U} Y_k^{-1} - (N/t_Y)^2\right]$ in the $\alpha = 1$, $m = 1$ case. These estimated variances can all be calculated using information observed only in the sample, as opposed to the theoretical anticipated variance calculations, which incorporate the true value of θ.

6.3. *Comparison of theoretical anticipated variances for $m = 1$ case as α and the sampling fraction vary*

The plots in Fig. 2 compare the theoretical anticipated variances for maximum likelihood, sample likelihood and pseudo-likelihood as α varies. Each plot shows the results for a particular combination of sample size n and population size N.

The plots illustrate how the pseudo-likelihood anticipated variance increases as α approaches 1. For high sampling fractions (plots B and C in Fig. 2), the performance of the maximum likelihood estimator is better than that of the sample likelihood estimator, while for small sampling fractions (plots A and D), the performance of the maximum likelihood and sample likelihood estimators are similar.

6.4. *Theoretical anticipated variances for maximum likelihood, sample likelihood and pseudo-likelihood for different values of m*

The anticipated variances in Table 2 are calculated from the theoretical formulas derived in this paper. For maximum likelihood, the theoretical anticipated variance is calculated from (2) using Monte Carlo integration, while the theoretical anticipated variances for sample likelihood and pseudo-likelihood can be calculated directly from (12) and (15). The true value of θ is set to 1. This table presents results for various values of m and α. In Table 2, RE_{SL} is the relative efficiency of the SL estimator compared with the maximum likelihood estimator, and it is defined as the anticipated

Table 1: Simulation results for $\theta = 1$, number of simulations = 10,000, $m = 1$. Here, $E(n_d)$ denotes the mean number of distinct sample units across the simulations, $E(\hat{\theta})$ is the mean value of $\hat{\theta}$, $\text{var}(\hat{\theta})$ is the variance of the values of $\hat{\theta}$ from the 10,000 simulations, $E(\hat{\text{var}})$ is the mean value of the variance estimate given by the inverse of the observed information, and AV is the theoretical anticipated variance, which is calculated using Monte Carlo integration for the MLE. In the Method column, 'MLE' indicates maximum likelihood estimation, 'SL' indicates sample likelihood and 'PL' indicates pseudo-likelihood

α	n	N	$\frac{n}{N}$	$E(n_d)$	Method	$E(\hat{\theta})$	$\text{var}(\hat{\theta})$	$E(\hat{\text{var}})$	AV
1	10	50	0.2	8.48	MLE	1.0638	0.0728	0.0547	0.0565
1	10	50	0.2	8.48	SL	1.1097	0.1003	0.0666	0.0900
1	10	50	0.2	8.48	PL	1.0265	0.6447	2.0339	∞
1	25	50	0.5	16.90	MLE	1.0327	0.0349	0.0298	0.0306
1	25	50	0.5	16.90	SL	1.0809	0.0621	0.0246	0.0600
1	25	50	0.5	16.90	PL	1.0141	0.1809	0.8142	∞
1	40	50	0.8	22.50	MLE	1.0272	0.0272	0.0244	0.0254
1	40	50	0.8	22.50	SL	1.0754	0.0530	0.0151	0.0525
1	40	50	0.8	22.50	PL	1.0201	0.1436	0.5109	∞
1	20	100	0.2	16.82	MLE	1.0328	0.0332	0.0282	0.0281
1	20	100	0.2	16.82	SL	1.0572	0.0474	0.0291	0.0450
1	20	100	0.2	16.82	PL	1.0159	0.2691	0.9132	∞
1	50	100	0.5	33.55	MLE	1.0170	0.0164	0.0151	0.0146
1	50	100	0.5	33.55	SL	1.0410	0.0303	0.0111	0.0300
1	50	100	0.5	33.55	PL	1.0110	0.2520	0.3562	∞
1	80	100	0.8	44.70	MLE	1.0133	0.0134	0.0123	0.0102
1	80	100	0.8	44.70	SL	1.0376	0.0265	0.0069	0.0263
1	80	100	0.8	44.70	PL	1.0096	0.0964	0.2195	∞

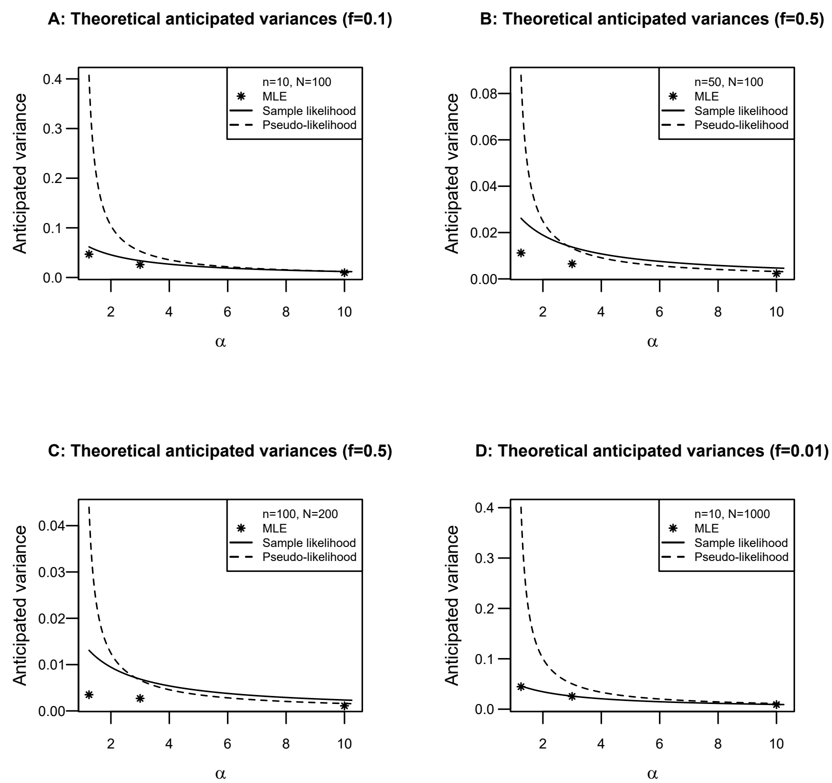

Figure 2: Comparison of theoretical anticipated variances for maximum likelihood, sample likelihood and pseudo-likelihood for varying α. The four plots show the comparison for different combinations of sample size n and population size N ($f = n/N$ is the sampling fraction).

variance of the maximum likelihood estimator divided by the anticipated variance of the sample likelihood estimator. Similarly, RE_{PL} is defined as the anticipated variance of the maximum likelihood estimator divided by the anticipated variance of pseudo-likelihood estimator.

The results in Table 2 illustrate how the relative efficiencies of the sample likelihood and pseudo-likelihood estimators vary across different scenarios. In particular, for higher values of m, the relative efficiency of the pseudo-likelihood estimator becomes very low, especially when α is close to 1. As discussed at the end of subsection 6.1, these results may be indicative of the relative efficiencies of the different likelihood approaches in related but more complex survey designs.

7. Case study using data set of geological faults

Borgos et al. (2002) compare two possible models for the distribution of underwater geological faults. Limitations of the methods for surveying these faults suggest that the observations are likely to be biased. A Pareto distribution has commonly been used in the geophysics literature to model the population of faults, but an exponential distribution has also been proposed. The authors use a data set, considered to be of high quality, collected from the Gullfaks Field in the North Sea, and they evaluate the models using the Bayes factor, which is a ratio that can be used as a criterion for deciding which of two models describes the observed data better. The method for observing the faults is sonar pulses from ships on the ocean surface. This method is more likely to detect larger faults. Table 1 of Borgos et al. (2002) contains the full data set of observed geological faults, which we use to explore an application of the results from this paper to a real-life example.

Borgos et al. (2002) state assumptions about the probability of observing a fault, depending on its length. These assumptions are based on properties of the technique used to collect the data. For faults with a length below 2 metres, the probability of selection is assumed to be constant and close to 0, while for faults with a length above 20 metres, the probability of selection is assumed to be constant and close to 1. For faults with a length between 2 metres and 20 metres, the probability of selection is assumed to be proportional to size. The selection probability as a function of length is continuous.

In the Gullfaks Field data, there were 66 faults between 2 metres and 20 metres, and 103 faults greater than 20 metres. The minimum length was 2 metres, the 25th percentile value was 16, the median value was 24, the 75th percentile value was 56 and the maximum value was 256.

Table 2: Results for $\theta = 1$, number of Monte Carlo realisations = 10,000. Here, $E(n_d)$ denotes the mean number of distinct sample units across the realisations, AV_{MLE} is the anticipated variance for the maximum likelihood estimate, AV_{SL} is the anticipated variance for the sample likelihood estimate, AV_{PL} is the anticipated variance for the pseudo-likelihood estimate, RE_{SL} is the relative efficiency of sample likelihood compared with maximum likelihood and RE_{PL} is the relative efficiency of pseudo-likelihood compared with maximum likelihood

m	α	n	N	$\frac{n}{N}$	$E(n_d)$	AV_{MLE}	AV_{SL}	AV_{PL}	RE_{SL}	RE_{PL}
0	1	20	100	0.2	18.22	0.0547	0.0600	0.0595	0.911	0.919
0	1	50	100	0.5	39.49	0.0257	0.0300	0.0298	0.858	0.864
0	1	80	100	0.8	55.25	0.0181	0.0225	0.0224	0.806	0.810
0	3	20	100	0.2	18.21	0.0180	0.0200	0.0198	0.900	0.908
0	3	50	100	0.5	39.53	0.0084	0.0100	0.0099	0.839	0.844
0	3	80	100	0.8	55.30	0.0060	0.0075	0.0075	0.798	0.803
0	5	20	100	0.2	18.22	0.0110	0.0120	0.0119	0.914	0.922
0	5	50	100	0.5	39.43	0.0051	0.0060	0.0060	0.853	0.859
0	5	80	100	0.8	55.21	0.0038	0.0045	0.0045	0.838	0.843
0.5	1	20	100	0.2	17.79	0.0372	0.0500	0.0684	0.744	0.544
0.5	1	50	100	0.5	37.52	0.0180	0.0300	0.0334	0.600	0.539
0.5	1	80	100	0.8	51.56	0.0140	0.0250	0.0246	0.560	0.569
0.5	3	20	100	0.2	18.07	0.0158	0.0206	0.0200	0.767	0.788
0.5	3	50	100	0.5	38.85	0.0075	0.0120	0.0100	0.620	0.746
0.5	3	80	100	0.8	54.02	0.0055	0.0099	0.0075	0.552	0.727
0.5	5	20	100	0.2	18.13	0.0100	0.0130	0.0119	0.768	0.834
0.5	5	50	100	0.5	39.14	0.0048	0.0075	0.0060	0.640	0.800
0.5	5	80	100	0.8	54.53	0.0034	0.0062	0.0045	0.546	0.749
1	1	20	100	0.2	16.82	0.0281	0.0450	∞	0.624	0
1	1	50	100	0.5	33.55	0.0146	0.0300	∞	0.487	0
1	1	80	100	0.8	44.70	0.0102	0.0263	∞	0.387	0
1	3	20	100	0.2	17.71	0.0139	0.0213	0.0282	0.644	0.486
1	3	50	100	0.5	37.18	0.0065	0.0138	0.0133	0.475	0.492
1	3	80	100	0.8	50.98	0.0047	0.0119	0.0095	0.397	0.494
1	5	20	100	0.2	17.90	0.0090	0.0139	0.0144	0.651	0.628
1	5	50	100	0.5	38.07	0.0042	0.0089	0.0070	0.473	0.604
1	5	80	100	0.8	52.56	0.0031	0.0076	0.0051	0.403	0.604
2	1	20	100	0.2	13.76	0.0208	0.0400	∞	0.520	0
2	1	50	100	0.5	24.73	0.0122	0.0300	∞	0.406	0
2	1	80	100	0.8	31.53	0.0078	0.0275	∞	0.285	0
2	3	20	100	0.2	16.35	0.0101	0.0224	0.1698	0.452	0.060
2	3	50	100	0.5	32.09	0.0051	0.0164	0.0699	0.310	0.073
2	3	80	100	0.8	42.52	0.0033	0.0149	0.0450	0.222	0.073
2	5	20	100	0.2	17.01	0.0073	0.0155	0.0369	0.470	0.198
2	5	50	100	0.5	34.51	0.0036	0.0112	0.0160	0.318	0.224
2	5	80	100	0.8	46.40	0.0025	0.0102	0.0107	0.243	0.230

It should be noted that in this case study, the value of N is assumed to be a fixed but unknown quantity. The sample likelihood and pseudo-likelihood estimators derived in this paper do not rely on N being known, and so they can be adapted for this case study. However, the maximum likelihood result derived in this paper assumes that N is known, and so it is not directly applicable. It may be possible to derive a maximum likelihood estimator for this case study using a different approach such as the missing information principle.

We assume that the population values of fault length y are drawn from the exponential distribution with rate parameter θ, and a sample is drawn from the population, with selection probabilities

$$\mathrm{pr}\,(I_i = 1 \mid y_i) \propto \begin{cases} h_1, & 0 \leq y_i < h_1 \\ y_i, & h_1 \leq y_i < h_2 \\ h_2, & h_2 \leq y_i \end{cases} \tag{16}$$

where $h_1 = 2$ and $h_2 = 20$.

The methods used previously to derive sample likelihood and pseudo-likelihood estimators can be extended in a straightforward way to the model in this case study. The sample likelihood estimate of θ ($\hat{\theta}_{SL}$) can be obtained by finding the solution to:

$$\bar{y}_s = \frac{1}{\hat{\theta}_{SL}} - \frac{1}{\hat{\theta}_{SL}} \left(\frac{-h_1\hat{\theta}_{SL}e^{-h_1\hat{\theta}_{SL}} - e^{-h_1\hat{\theta}_{SL}} + h_2\hat{\theta}_{SL}e^{-h_2\hat{\theta}_{SL}} + e^{-h_2\hat{\theta}_{SL}}}{h_1\hat{\theta}_{SL} + e^{-h_1\hat{\theta}_{SL}} - e^{-h_2\hat{\theta}_{SL}}} \right). \tag{17}$$

See Appendix B for the derivation of (17).

The pseudo-likelihood estimate of θ ($\hat{\theta}_{PL}$) is:

$$\hat{\theta}_{PL} = \frac{\frac{n_1}{h_1} + \sum_{i \in S, h_1 < y_i \leq h_2} \frac{1}{y_i} + \frac{n_3}{h_2}}{\sum_{i \in S, y_i < h_1} \frac{y_i}{h_1} + n_2 + \sum_{i \in S, y_i \geq h_2} \frac{y_i}{h_2}} \tag{18}$$

where n_1 is the number of sample units y_i such that $0 \leq y_i < h_1$, n_2 the number of sample units y_i such that $h_1 \leq y_i < h_2$ and n_3 the number of sample units y_i such that $h_2 \leq y_i$. See Appendix C for the derivation of (18).

Applying these estimators to the actual data set from Table 1 of Borgos et al. (2002) gives $\hat{\theta}_{SL} = 0.028$ and $\hat{\theta}_{PL} = 0.033$, compared with an estimate of 0.027 with a 95% prediction interval of $[0.023, 0.031]$ calculated from the Markov Chain Monte Carlo technique used by Borgos et al. (2002).

Using the values of $\hat{\theta}$ and N estimated by Borgos et al. (2002) to run $10,000$ simulations, the mean parameter estimates and 95% confidence intervals derived from the

Monte Carlo variances are 0.0273 and $[0.0217, 0.0330]$ for sample likelihood, and 0.0272 and $[0.0211, 0.0334]$ for pseudo-likelihood. Even though the simulated sample data sets have a large proportion of observations that are greater than h_1 and so are assumed to be selected with equal probability, the Monte Carlo variance of the sample likelihood estimator is nearly 16% lower than the Monte Carlo variance of the pseudo-likelihood estimator in this simulation study.

It should be noted that the example given in Borgos et al. (2002) contains some additional complexities not discussed here. In particular, some degree of bias is expected in the measurement of the fault length, and there is a small value of y below which observation is impossible in practice.

8. Discussion of theoretical results, simulation study and case study

The comparison of the anticipated variances shown in Table 1 suggests that the maximum likelihood approach and the sample likelihood approach will perform substantially better than pseudo-likelihood when α is close to 1. When this is the case, there are many population units with low probabilities of selection. However, sample likelihood is not always more efficient than pseudo-likelihood, particularly when the sampling fraction is large and α is small. The sample likelihood estimator and the pseudo-likelihood estimator do not include any information about duplicate units in the sample, while the maximum likelihood estimator contains some information about duplicate units in the t_{s_d} term. The sample likelihood estimator is particularly affected by duplicate units and becomes relatively less efficient compared with the full maximum likelihood as the sampling fraction $f = n/N$ increases (see fig. 2). Conversely, as the population size N increases and the other parameters are held constant, the anticipated variance for sample likelihood converges to that of maximum likelihood, while pseudo-likelihood remains less efficient. As $N \to \infty$, $AV_{PL}/AV_{SL} \to (\alpha + 1)/(\alpha - 1)$ (from (12) and Corollary 7 in Appendix C), which goes to infinity as α goes to 1.

The case study using the data from Borgos et al. (2002) gives an example of applying the results derived in this paper to a real-life problem. Even though most of observations in the data set are assumed to be selected with equal probability, the sample likelihood approach is still more efficient than pseudo-likelihood, with the variance of the pseudo-likelihood estimator being nearly twenty percent higher than the variance of the sample likelihood estimator.

Appendix A. Maximum likelihood proofs

Proof of Theorem 1 . The probability function for the sampled units given the population is the design-based probability of selecting the entire sample, which is

$$f(y_s \mid y_U) = \prod_{i \in S} \frac{y_i^m}{\sum_{j \in S_d} y_j^m + \sum_{k \in R} y_k^m} .$$

The likelihood is therefore

$$L(\theta) = f(y_s) = \int f(y_s, y_r)\, dy_r = \int \frac{N!}{n_r!} f(y_s, y_U)\, dy_r = \int \frac{N!}{n_r!} f(y_s \mid y_U) f(y_U)\, dy_r$$

as there are $N!/n_r!$ ways for y_s to be interleaved with the non-sampled units y_r

$$= \int \frac{N!}{n_r!} \left(\prod_{i \in S} \frac{y_i^m}{\sum_{j \in S_d} y_j^m + \sum_{k \in R} y_k^m} \right) \left\{ \prod_{j \in U} \frac{\theta^\alpha y_j^{\alpha-1} e^{-\theta y_j}}{\Gamma(\alpha)} \right\} dy_r$$

$$= E_{y_r} \left[\frac{N!}{n_r!} \left(\prod_{i \in S} \frac{y_i^m}{\sum_{j \in S_d} y_j^m + \sum_{k \in R} y_k^m} \right) \left\{ \prod_{j \in S_d} \frac{\theta^\alpha y_j^{\alpha-1} e^{-\theta y_j}}{\Gamma(\alpha)} \right\} : y_r \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \theta) \right] . \tag{A.1}$$

So

$$\frac{\partial L(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \int \frac{N!}{n_r!} \left(\prod_{i \in S} \frac{y_i^m}{\sum_{j \in S_d} y_j^m + \sum_{k \in R} y_k^m} \right) \left\{ \prod_{j \in U} \frac{\theta^\alpha y_j^{\alpha-1} e^{-\theta y_j}}{\Gamma(\alpha)} \right\} dy_r$$

$$= \int \frac{N!}{n_r!} \left(\prod_{i \in S} \frac{y_i^m}{\sum_{j \in S_d} y_j^m + \sum_{k \in R} y_k^m} \right) \left\{ \prod_{j \in U} \frac{y_j^{\alpha-1}}{\Gamma(\alpha)} \right\} \frac{\partial}{\partial \theta} \left(\theta^{N\alpha} e^{-\theta t_U} \right) dy_r$$

(assuming regularity conditions apply, allowing

the derivative $\partial/\partial\theta$ to be taken within the integral)

$$= \int \frac{N!}{n_r!} \left(\prod_{i \in S} \frac{y_i^m}{\sum_{j \in S_d} y_j^m + \sum_{k \in R} y_k^m} \right) \left\{ \prod_{j \in S_d} \frac{\theta^\alpha y_j^{\alpha-1} e^{-\theta y_j}}{\Gamma(\alpha)} \right\}$$

$$\left\{ \prod_{k \in R} \frac{\theta^\alpha y_k^{\alpha-1} e^{-\theta y_k}}{\Gamma(\alpha)} \right\} \left(N\alpha\theta^{-1} - t_U \right) dy_r$$

$$= E_{y_r}\left[\frac{N!}{n_r!}\left(\prod_{i\in S}\frac{y_i^m}{\sum_{j\in S_d} y_j^m + \sum_{k\in R} y_k^m}\right)\left\{\prod_{j\in S_d}\frac{\theta^\alpha y_j^{\alpha-1}e^{-\theta y_j}}{\Gamma(\alpha)}\right\}\right.$$
$$\left.\left(N\alpha\theta^{-1} - t_{s_d} - t_r\right) : y_r \overset{\text{iid}}{\sim} \text{Gamma}(\alpha,\theta)\right]. \tag{A.2}$$

Let $\sum_{k\in U} y_k^m = \sum_{j\in S_d} y_j^m + \sum_{k\in R} y_k^m$ be denoted by t_{Y^m}. The score function is

$$\begin{aligned}
\text{sc}(\theta) &= \frac{\partial \log L(\theta)}{\partial\theta} = \frac{\frac{\partial}{\partial\theta}L(\theta)}{L(\theta)} \\
&= \frac{\frac{N!}{n_r!}\left(\prod_{i\in S} y_i^m\right)\left\{\prod_{j\in S_d}\frac{\theta^\alpha y_j^{\alpha-1}e^{-\theta y_j}}{\Gamma(\alpha)}\right\} E_{y_r}\left\{\left(\frac{1}{t_{Y^m}}\right)^n\left(N\alpha\theta^{-1} - t_{s_d} - t_r\right) : y_r \overset{\text{iid}}{\sim} \text{Gamma}(\alpha,\theta)\right\}}{\frac{N!}{n_r!}\left(\prod_{i\in S} y_i^m\right)\left\{\prod_{j\in S_d}\frac{\theta^\alpha y_j^{\alpha-1}e^{-\theta y_j}}{\Gamma(\alpha)}\right\} E_{y_r}\left\{\left(\frac{1}{t_{Y^m}}\right)^n : y_r \overset{\text{iid}}{\sim} \text{Gamma}(\alpha,\theta)\right\}} \\
&= \frac{E_{y_r}\left\{\left(\frac{1}{t_{Y^m}}\right)^n\left(N\alpha\theta^{-1} - t_{s_d} - t_r\right) : y_r \overset{\text{iid}}{\sim} \text{Gamma}(\alpha,\theta)\right\}}{E_{y_r}\left\{\left(\frac{1}{t_{Y^m}}\right)^n : y_r \overset{\text{iid}}{\sim} \text{Gamma}(\alpha,\theta)\right\}}.
\end{aligned} \tag{A.3}$$

□

Proof of Theorem 2 . Let $z_i = \theta y_i$. The sum of the distinct units in the scaled sample z_s is denoted by $t_{z_{s_d}}$, the sum of the scaled non-sampled units in z_r is denoted by t_{z_r}, and the sum of the scaled population units in z_U is denoted by t_Z. For a change of integration variable,

$$dy_r = \theta^{-n_r}\, dz_r\,. \tag{A.4}$$

From (A.1) and (A.4), the likelihood of the sample y_s is

$$\begin{aligned}
L(\theta) &= \int \frac{N!}{n_r!}\left(\prod_{i\in S}\frac{\theta^{-m}z_i^m}{\sum_{j\in S_d}\theta^{-m}z_j^m + \sum_{k\in R}\theta^{-m}z_k^m}\right)\left\{\prod_{j\in U}\frac{\theta z_j^{\alpha-1}e^{-z_j}}{\Gamma(\alpha)}\right\}\theta^{-n_r}\, dz_r \\
&= \frac{N!\theta^N}{\Gamma(\alpha)^N}\int \frac{\theta^{-n_r}}{n_r!}\left(\prod_{i\in S}\frac{z_i^m}{\sum_{j\in S_d} z_j^m + \sum_{k\in R} z_k^m}\right)\left(\prod_{j\in U} z_j^{\alpha-1}\right)e^{-t_Z}\, dz_r\,.
\end{aligned} \tag{A.5}$$

Let $\sum_{k=1}^{N} z_k^m = \sum_{j\in S_d} z_j^m + \sum_{k\in R} z_k^m$ be denoted by t_{Z^m}. Then, from (A.5),

$$\frac{\partial L(\theta)}{\partial\theta} = \frac{\partial}{\partial\theta}\int \frac{N!}{n_r!}\left(\prod_{i\in S}\frac{y_i^m}{\sum_{j\in S_d} y_j^m + \sum_{k\in R} y_k^m}\right)\left\{\prod_{j\in U}\frac{\theta^\alpha y_j^{\alpha-1}e^{-\theta y_j}}{\Gamma(\alpha)}\right\} dy_r$$

$$= \int \frac{N!}{n_r!}\left(\prod_{i\in S}\frac{y_i^m}{\sum_{j\in S_d} y_j^m + \sum_{k\in R} y_k^m}\right)\left\{\prod_{j\in U}\frac{y_j^{\alpha-1}}{\Gamma(\alpha)}\right\}\frac{\partial}{\partial\theta}\left(\theta^{N\alpha}e^{-\theta t_U}\right) dy_r$$

(assuming regularity conditions apply, allowing

the derivative $\partial/\partial\theta$ to be taken within the integral)

$$= \int \frac{N!}{n_r!}\left(\prod_{i\in S}\frac{\theta^{-1}z_i^m}{\sum_{j\in S_d}\theta^{-m}z_j^m + \sum_{k\in R}\theta^{-m}z_k^m}\right)\left\{\prod_{j\in U}\frac{\theta^{-(\alpha-1)}z_j^{\alpha-1}}{\Gamma(\alpha)}\right\}$$
$$\times\left(N\alpha\theta^{N\alpha-1}e^{-t_Z} - \theta^{N\alpha}\theta^{-1}t_Z e^{-t_Z}\right)\theta^{-n_r}dz_r$$
$$= \frac{N!\theta^{N-1}}{\Gamma(\alpha)^N}\int \frac{\theta^{-n_r}}{n_r!}\left(\prod_{i\in S}\frac{z_i^m}{t_{Z^m}}\right)\left(\prod_{j\in U} z_j^{\alpha-1}\right)e^{-t_Z}\,(N\alpha - t_Z)\,dz_r\,. \qquad \text{(A.6)}$$

From (A.5) and (A.6), the score function for the sample y_s is

$$\mathrm{sc}(\theta) = \frac{\frac{\partial}{\partial\theta}L(\theta)}{L(\theta)}$$
$$= \frac{\frac{N!\theta^{N-1}}{\Gamma(\alpha)^N}\int\frac{\theta^{-n_r}}{n_r!}\left(\prod_{i\in S}\frac{z_i^m}{t_{Z^m}}\right)\left(\prod_{j\in U}z_j^{\alpha-1}\right)e^{-t_Z}\,(N\alpha - t_Z)\,dz_r}{\frac{N!\theta^N}{\Gamma(\alpha)^N}\int\frac{\theta^{-n_r}}{n_r!}\left(\prod_{i\in S}\frac{z_i^m}{t_{Z^m}}\right)\left(\prod_{j\in U}z_j^{\alpha-1}\right)e^{-t_Z}\,dz_r}$$
$$= \theta^{-1}\frac{\int\frac{\theta^{-n_r}}{n_r!}\,(N\alpha - t_Z)\left(\prod_{i\in S}\frac{z_i^m}{t_{Z^m}}\right)\left(\prod_{j\in U}z_j^{\alpha-1}\right)e^{-t_Z}\,dz_r}{\int\frac{\theta^{-n_r}}{n_r!}\left(\prod_{i\in S}\frac{z_i^m}{t_{Z^m}}\right)\left(\prod_{j\in U}z_j^{\alpha-1}\right)e^{-t_Z}\,dz_r}\,. \qquad \text{(A.7)}$$

It is a standard result that if regularity conditions hold, $\mathcal{I}(\theta) = E\left[\{\mathrm{sc}(\theta)\}^2\right]$. This expectation is across all the possible samples, and the value of n_r depends on the particular scaled sample z_s. We partition the set of possible z_s into groups based on the value of n_r. For each of these groups, n_r can be treated as a constant for expectations calculated across the samples within the group. So from (A.7),

$$\begin{aligned}
\mathcal{I}(\theta) &= E_{z_s}\left[\{\mathrm{sc}(\theta)\}^2\right] = \sum_{n_d=1}^{n} E_{z_s}\left[(\mathrm{sc}(\theta))^2 \mid n_d\right] \mathrm{pr}\,[n_d] \\
&= \sum_{n_d=1}^{n} E_{z_s}\left[\left\{\theta^{-1}\frac{\int \frac{\theta^{-n_r}}{n_r!}(N\alpha - t_Z)\left(\prod_{i\in S}\frac{z_i^m}{t_{Z^m}}\right)\left(\prod_{j\in U} z_j^{\alpha-1}\right)e^{-t_Z}\,dz_r}{\int \frac{\theta^{-n_r}}{n_r!}\left(\prod_{i\in S}\frac{z_i^m}{t_{Z^m}}\right)\left(\prod_{j\in U} z_j^{\alpha-1}\right)e^{-t_Z}\,dz_r}\right\}^2 \mid n_d\right] \mathrm{pr}\,(n_d) \\
&= \theta^{-2}\sum_{n_d=1}^{n} E_{z_s}\left(\left[\frac{\left(\prod_{i\in S} z_i^m\right)\left(\prod_{i\in S_d} z_i^{\alpha-1}e^{-z_i}\right)\Gamma(\alpha)^{-n_r}E_{z_r}\left\{\frac{\left(N\alpha - t_{z_{s_d}} - t_{z_r}\right)}{(t_{Z^m})^n}\right\}}{\left(\prod_{i\in S} z_i^m\right)\left(\prod_{i\in S_d} z_i^{\alpha-1}e^{-z_i}\right)\Gamma(\alpha)^{-n_r}E_{z_r}\left\{\frac{1}{(t_{Z^m})^n}\right\}}\right]^2 \mid n_d\right) \mathrm{pr}\,(n_d) \\
&= \theta^{-2}\sum_{n_d=1}^{n} E_{z_s}\left(\left[\frac{E_{z_r}\left\{\frac{\left(N\alpha - t_{z_{s_d}} - t_{z_r}\right)}{(t_{Z^m})^n} : z_r \overset{\text{iid}}{\sim} \mathrm{Gamma}(\alpha, 1)\right\}}{E_{z_r}\left\{\frac{1}{(t_{Z^m})^n} : z_r \overset{\text{iid}}{\sim} \mathrm{Gamma}(\alpha, 1)\right\}}\right]^2 \mid n_d\right) \mathrm{pr}\,(n_d) \\
&= \theta^{-2}E_{z_s}\left(\left[\frac{E_{z_r}\left\{\frac{\left(N\alpha - t_{z_{s_d}} - t_{z_r}\right)}{(t_{Z^m})^n} : z_r \overset{\text{iid}}{\sim} \mathrm{Gamma}(\alpha, 1)\right\}}{E_{z_r}\left\{\frac{1}{(t_{Z^m})^n} : z_r \overset{\text{iid}}{\sim} \mathrm{Gamma}(\alpha, 1)\right\}}\right]^2\right). \qquad \text{(A.8)}
\end{aligned}$$

□

Heuristic derivation of (7). We assume that the probability of selecting duplicate units is 0, so that $N = n + n_r$ and $t_{z_{s_d}} = t_{z_s}$. We also assume that the sum of the scaled sampled units $t_{z_{s_d}}$ is very small compared with the sum of the scaled non-sampled units t_{z_r}, so that $(t_{z_{s_d}} + t_{z_r})^n \approx t_{z_r}^n$. Under these assumptions, (6) becomes

$$\mathcal{I}(\theta) \approx \theta^{-2}E_{z_s}\left(\left[N\alpha - t_{z_s} - \frac{E_{t_{z_r}}\left\{t_{z_r}^{-n+1} : t_{z_r} \sim \mathrm{Gamma}(n_r\alpha, 1)\right\}}{E_{t_{z_r}}\left\{t_{z_r}^{-n} : t_{z_r} \sim \mathrm{Gamma}(n_r\alpha, 1)\right\}}\right]^2\right). \qquad \text{(A.9)}$$

From properties of the gamma distribution, if $X \sim \mathrm{Gamma}(\alpha, \theta)$ and $q > -\alpha$, then $E\left(X^{q+1}\right)/E\left(X^q\right) = (q + \alpha)/\theta$. So

$$\frac{E_{t_{z_r}}\left\{t_{z_r}^{-n+1} : t_{z_r} \sim \mathrm{Gamma}(n_r\alpha, 1)\right\}}{E_{t_{z_r}}\left\{t_{z_r}^{-n} : t_{z_r} \sim \mathrm{Gamma}(n_r\alpha, 1)\right\}} = -n + n_r\alpha \qquad \text{(A.10)}$$

for $n_r\alpha > n$. Substituting (A.10) into (A.9) gives

$$\mathcal{I}(\theta) \approx \theta^{-2}\, E_{z_s}\left[\{N\alpha - t_{z_s} - (-n + n_r\alpha)\}^2\right] = \theta^{-2}\, E_{z_s}\left[\{n(\alpha+1) - t_{z_s}\}^2\right]. \tag{A.11}$$

In this case, $m = 1$, so the expectation of t_{z_s} is

$$\begin{aligned} E_{z_s}\left(t_{z_s}\right) &= E_m\left\{E_p\left(t_{z_s}\right)\right\} = E_m\left(n \sum_{k\in U} \frac{Z_k}{t_Z} Z_k\right) \\ &= n \frac{E_m\left(\sum_{k\in U} Z_k^2\right)}{E_m\left(\sum_{k\in U} Z_k\right)} = n \frac{\alpha\,(\alpha+1)}{\alpha} = n(\alpha+1)\,. \end{aligned} \tag{A.12}$$

So from (A.11) and (A.12),

$$\begin{aligned} \mathcal{I}(\theta) &\approx \theta^{-2}\, E_{z_s}\left[\{E_{z_s}\left(t_{z_s}\right) - t_{z_s}\}^2\right] = \theta^{-2}\, \mathrm{var}_{z_s}\left(t_{z_s}\right) \\ &= \theta^{-2}\left[E_m\left\{\mathrm{var}_p\left(t_{z_s}\right)\right\} + \mathrm{var}_m\left\{E_p\left(t_{z_s}\right)\right\}\right]. \end{aligned} \tag{A.13}$$

Following the approach in Sections 6.1 and 6.2 of Lohr (2021), the design expectation and variance of t_{z_s} are

$$E_p\left(t_{z_s}\right) = n \sum_{k=1}^{N} \frac{Z_k}{t_Z} Z_k = \frac{n}{t_Z} \sum_{k\in U} Z_k^2 \tag{A.14}$$

and

$$\begin{aligned} \mathrm{var}_p\left(t_{z_s}\right) &= E_p\left(t_{z_s}^2\right) - \left\{E_p\left(t_{z_s}\right)\right\}^2 \\ &= n \sum_{k\in U} \frac{Z_k}{t_Z} Z_k^2 + \left(n^2 - n\right) \sum_{k\in U, l\in U} \frac{Z_k Z_l}{t_Z^2} Z_k Z_l - \frac{n^2}{t_Z^2}\left(\sum_{k\in U}^{N} Z_k^2\right)^2 \\ &= n\left\{\sum_{k\in U} \frac{Z_k^3}{t_Z} - \left(\sum_{k\in U} \frac{Z_k^2}{t_Z}\right)^2\right\}. \end{aligned} \tag{A.15}$$

Combining (A.13), (A.14) and (A.15) gives

$$
\begin{aligned}
&\mathcal{I}(\theta)\\
&\approx \theta^{-2}\left(E_m\left[n\left\{\sum_{k\in U}\frac{Z_k^3}{t_Z}-\left(\sum_{k\in U}\frac{Z_k^2}{t_Z}\right)^2\right\}\right]\right)+E_m\left[\left(\frac{n}{t_Z}\sum_{k\in U}Z_k^2\right)^2\right]-\left(E_m\frac{n}{t_Z}\sum_{k\in U}Z_k^2\right)^2\\
&\approx \theta^{-2}\left\{\frac{\Gamma(\alpha+3)}{\alpha\Gamma(\alpha)}-\frac{\Gamma(\alpha+2)^2}{\alpha^2\Gamma(\alpha)^2}+\frac{n\Gamma(\alpha+4)}{N\alpha^2\Gamma(\alpha)}-\frac{n\Gamma(\alpha+2)^2}{N\alpha^2\Gamma(\alpha)^2}\right\}\\
&\qquad\qquad \text{from properties of the gamma distribution}\\
&=\theta^{-2}n\left\{(\alpha+1)+\frac{n}{N\alpha}\left(4\alpha^2+10\alpha+6\right)\right\}\\
&\to \theta^{-2}n(\alpha+1)\qquad \text{as } N\to\infty \text{ while } n \text{ is fixed.}
\end{aligned}
\tag{A.16}
$$

□

Appendix B. Sample likelihood proofs

Proof of (12). The selection probabilities are $y_i^m / \sum_{k=1}^N y_k^m$. Let $\sum_{k=1}^N y_k^m$ be denoted by t_{Y^m}. The anticipated variance of $\bar{y}_s$ is

$$AV(\bar{y}_s) = E_m\left\{\text{var}_p(\bar{y}_s)\right\} + \text{var}_m\left\{E_p(\bar{y}_s)\right\}. \tag{B.1}$$

The design expectation and variance of $\bar{y}_s$ are

$$E_p(\bar{y}_s) = \sum_{k=1}^N \frac{Y_k^m}{t_{Y^m}} Y_k = \frac{N}{t_{Y^m}} \frac{\sum_{k=1}^N Y_k^{m+1}}{N}, \tag{B.2}$$

$$\text{var}_p(\bar{y}_s) = n^{-1}\left\{\sum_{k=1}^N \frac{Y_k^{m+2}}{t_{Y^m}} - \left(\sum_{k=1}^N \frac{Y_k^{m+1}}{t_{Y^m}}\right)^2\right\}. \tag{B.3}$$

Substituting (B.2) and (B.3) into (B.1) gives

$$
\begin{aligned}
&AV\{\bar{y}_s\} \\
&= E_m\left[n^{-1}\left\{\frac{NY^{m+2}}{t_{Y^m}} - \left(\frac{NY^{m+1}}{t_{Y^m}}\right)^2\right\}\right] + \frac{1}{N}\operatorname{var}_m\left(\frac{N}{t_{Y^m}}Y^{m+1}\right) \\
&= n^{-1}\left\{\frac{(m+1+\alpha)(m+\alpha)}{\theta^2} - \left(\frac{m+\alpha}{\theta}\right)^2\right\} + \frac{1}{N}\left\{\frac{(2m+1+\alpha)(2m+\alpha)}{\theta^2} - \left(\frac{m+\alpha}{\theta}\right)^2\right\} \\
&\qquad\qquad \text{(from properties of the gamma distribution)} \\
&= \frac{\alpha+m}{n\theta^2} + \frac{3m^2+2m\alpha+2m+\alpha}{N\theta^2}\,. \qquad \text{(B.4)}
\end{aligned}
$$

From (10), $\hat{\theta}_{SL} = (\alpha+m)/\bar{y}_s$. From (B.2),

$$
E(\bar{y}_s) = E_m\left\{E_p(\bar{y}_s)\right\} = E_m\left(\frac{N}{t_{Y^m}}\frac{\sum_{k=1}^N Y_k^{m+1}}{N}\right) = \frac{\alpha+m}{\theta}\,. \qquad \text{(B.5)}
$$

Using a Taylor series expansion and combining (B.5) and (B.4) gives

$$
\begin{aligned}
AV\left(\hat{\theta}\right) &= (\alpha+m)^2 AV\left(\frac{1}{\bar{y}_s}\right) \approx (\alpha+m)^2\left\{\frac{1}{E(\bar{y}_s)}\right\}^4 AV(\bar{y}_s) \\
&= (\alpha+m)^2\left(\frac{\theta}{\alpha+m}\right)^4\left(\frac{\alpha+m}{n\theta^2} + \frac{3m^2+2m\alpha+2m+\alpha}{N\theta^2}\right) \\
&= \frac{\theta^2}{(\alpha+m)^2}\left(\frac{\alpha+m}{n} + \frac{3m^2+2m\alpha+2m+\alpha}{N}\right). \qquad \text{(B.6)}
\end{aligned}
$$

□

Proof of (17) *(sample likelihood estimator for case study of geological fault data).* Given the model (16), the sample likelihood is

$$
\begin{aligned}
L_s &= \prod_{i\in s} \frac{\mathrm{pr}\,(I_i = 1 \mid y_i)\, f\,(y_i)}{\mathrm{pr}\,(I_i = 1)} \\
&\propto \left(\prod_{i\in S, y_i<h_1} h_1\theta e^{-\theta y_i}\right)\left(\prod_{i\in S, h_1<y_i\le h_2} y_i\theta e^{-\theta y_i}\right)\left(\prod_{i\in S, y_i\ge h_2} h_2\theta e^{-\theta y_i}\right)\times \\
&\qquad \left\{\prod_{i\in s}\left(\int_0^{h_1} h_1\theta e^{-\theta y_i}\, dy + \int_{h_1}^{h_2} y\theta e^{-\theta y_i}\, dy + \int_{h_2}^{\infty} h_2\theta e^{-\theta y_i}\, dy\right)\right\}^{-1} \\
&= \left(\prod_{i\in S, y_i<h_1} h_1\theta e^{-\theta y_i}\right)\left(\prod_{i\in S, h_1<y_i\le h_2} y_i\theta e^{-\theta y_i}\right)\left(\prod_{i\in S, y_i\ge h_2} h_2\theta e^{-\theta y_i}\right)\times \\
&\qquad \left\{\prod_{i\in s}\left(h_1 - \frac{e^{-h_2\theta}}{\theta} + \frac{e^{-h_1\theta}}{\theta}\right)\right\}^{-1} .
\end{aligned}
$$

Therefore

$$
\frac{\partial}{\partial\theta}\log\,(L_s) = \sum_{i\in S}\left(\frac{1}{\theta} - y_i\right) - n\left(\frac{-h_1\theta e^{-h_1\theta} - e^{-h_1\theta} + h_2\theta e^{-h_2\theta} + e^{-h_2\theta}}{\theta^2 h_1 + \theta e^{-h_1\theta} - \theta e^{-h_2\theta}}\right)
$$

and

$$
\bar{y}_s = \frac{1}{\hat{\theta}} - \frac{1}{\hat{\theta}}\left(\frac{-h_1\hat{\theta} e^{-h_1\hat{\theta}} - e^{-h_1\hat{\theta}} + h_2\hat{\theta} e^{-h_2\hat{\theta}} + e^{-h_2\hat{\theta}}}{\hat{\theta} h_1 + e^{-h_1\hat{\theta}} - e^{-h_2\hat{\theta}}}\right).
$$

□

Appendix C. Pseudo-likelihood proofs

A Taylor series approximation for the variance of a ratio of two random variables A and B is

$$
\mathrm{var}\,(A/B) \approx \frac{\{E(A)\}^2}{\{E(B)\}^2}\left[\frac{\mathrm{var}\,(A)}{\{E(A)\}^2} - 2\frac{\mathrm{cov}\,(A,B)}{E(A)E(B)} + \frac{\mathrm{var}\,(B)}{\{(E(B)\}^2}\right]. \tag{C.1}
$$

The following four lemmas will be used in the derivation of the anticipated variance of the pseudo-likelihood estimator.

Lemma 3. *If $X \sim Gamma\,(\alpha,\theta)$, then by properties of the gamma distribution,*

$$
E\,(X^q) = \frac{\Gamma\,(q+\alpha)}{\theta^q\Gamma\,(\alpha)}
$$

for any q such that $q + \alpha > 0$.

Lemma 4. *Suppose s is selected from U with probability proportional to a power m of size with replacement, and $\alpha + m + q > 0$ and $\alpha + m > 0$. Then the following result applies.*

$$E\left(\sum_{i\in S} y_i^q\right) = E_m\left\{E_p\left(\sum_{i\in S} y_i^q\right)\right\} = E_m\left\{n\sum_{k\in U}\left(\frac{Y_k^m}{t_{Y^m}} Y_k^q\right)\right\} \approx \frac{E_m\left(Y_k^{m+q}\right)}{E_m\left(t_{Y^m}\right)}$$
$$= n\theta^{-q}\frac{\Gamma(\alpha+m+q)}{\Gamma(\alpha+m)} \tag{C.2}$$

from Lemma 3.

Lemma 5. *Suppose s is selected from U with probability proportional to a power m of size with replacement, and $\alpha + m + q_1 > 0$, $\alpha + m + q_2 > 0$, $\alpha + m + q_1 + q_2 > 0$ and $\alpha + m > 0$. Then the following result applies.*

$$\text{acov}\left(\sum_{i\in S} y_i^{q_1}, \sum_{i\in S} y_i^{q_2}\right) = E_m\left\{E_p\left(\sum_{i\in S} y_i^{q_1}\sum_{i\in S} y_i^{q_2}\right)\right\} - E_m\left\{E_p\left(\sum_{i\in S} y_i^{q_1}\right)\right\} E_m\left\{E_p\left(\sum_{i\in S} y_i^{q_2}\right)\right\}$$
$$= E_m\left\{E_p\left(\sum_{i\in S, j\in S} y_i^{q_1} y_j^{q_2}\right)\right\} - E_m\left(n\sum_{k\in U}\frac{Y_k^m}{t_{Y^m}} Y_k^{q_1}\right) E_m\left(n\sum_{k\in U}\frac{Y_k^m}{t_{Y^m}} Y_k^{q_2}\right)$$
$$\approx E_m\left\{E_p\left(\sum_{i\in S} y_i^{q_1} y_i^{q_2}\right) + E_p\left(\sum_{i,j\in S, i\neq j} y_i^{q_1} y_j^{q_2}\right)\right\}$$
$$- n\frac{\Gamma(\alpha+m+q_1)}{\theta^{q_1}\Gamma(\alpha+m)} n\frac{\Gamma(\alpha+m+q_2)}{\theta^{q_2}\Gamma(\alpha+m)}$$
$$= E_m\left\{n\sum_{k\in U}\frac{Y_k^m}{t_{Y^m}} Y_k^{q_1+q_2} + \left(n^2 - n\right)\sum_{k\in U, l\in U}\frac{Y_k^m Y_l^m}{(t_{Y^m})^2} Y_k^{q_1} Y_l^{q_2}\right\}$$
$$- n^2\frac{\Gamma(\alpha+m+q_1)\Gamma(\alpha+m+q_2)}{\theta^{q_1+q_2}\Gamma(\alpha+m)^2}$$

$$= nE_m\left(\sum_{k\in U}\frac{Y_k^{q_1+q_2+m}}{t_{Y^m}}\right)+\left(n^2-n\right)\left[E_m\left\{\sum_{k\in U}\frac{Y_k^{q_1+q_2+2m}}{(t_{Y^m})^2}\right\}+\right.$$

$$\left.E_m\left\{\sum_{k,l\in U,k\neq l}\frac{Y_k^{q_1+m}Y_l^{q_2+m}}{(t_{Y^m})^2}\right\}\right]$$

$$-n^2\frac{\Gamma(\alpha+m+q_1)\Gamma(\alpha+m+q_2)}{\theta^{q_1+q_2}\Gamma(\alpha+m)^2}$$

$$\approx \frac{n\theta^m\Gamma(\alpha)}{\Gamma(m+\alpha)}\frac{\Gamma(q_1+q_2+m+\alpha)}{\theta^{q_1+q_2+m}\Gamma(\alpha)}+\frac{\left(n^2-n\right)}{N^2}\frac{\theta^{2m}\Gamma(\alpha)^2}{\Gamma(m+\alpha)^2}\left\{\right.$$

$$\frac{N\Gamma(q_1+q_2+2m+\alpha)}{\theta^{q_1+q_2+2m}\Gamma(\alpha)}+$$

$$\left.\frac{N(N-1)\Gamma(q_1+m+\alpha)\Gamma(q_2+m+\alpha)}{\theta^{q_1+m}\theta^{q_2+m}\Gamma(\alpha)^2}\right\}$$

$$-n^2\frac{\Gamma(q_1+m+\alpha)\Gamma(q_2+m+\alpha)}{\theta^{q_1+q_2}\Gamma(m+\alpha)^2}$$

$$\approx \frac{n}{\theta^{q1+q2}}\left\{\frac{\Gamma(q_1+q_2+m+\alpha)}{\Gamma(m+\alpha)}+\frac{(n-1)\Gamma(\alpha)\Gamma(q_1+q_2+2m+\alpha)}{N\Gamma(m+\alpha)^2}+\right.$$

$$\left.\frac{(1-n-N)\Gamma(q_1+m+\alpha)\Gamma(q_2+m+\alpha)}{N\Gamma(m+\alpha)^2}\right\}$$

using a Taylor series approximation and Lemma 3.

Lemma 6. *For y_i selected with probability proportional to a power m of y_i from a population drawn from a gamma distribution with parameters (α, θ), the anticipated variance*

$$AV\left(\sum_{i\in S}y_i^q\right)\approx\frac{n}{\theta^{2q}}\left\{\frac{\Gamma(2q+m+\alpha)}{\Gamma(m+\alpha)}+\frac{(n-1)\Gamma(\alpha)\Gamma(2q+2m+\alpha)}{N\Gamma(m+\alpha)^2}+\right.$$

$$\left.\frac{(1-n-N)\Gamma(q+m+\alpha)^2}{N\Gamma(m+\alpha)^2}\right\}$$

for any q, as long as $\alpha+m+q>0$, $\alpha+m+2q>0$, and $\alpha+m>0$, from Lemma 5.

Proof of (15). Combining (C.1), Lemma 4, Lemma 5 and Lemma 6 gives

$$
\begin{aligned}
&AV\left(\hat{\theta}_{PL}\right) \\
&\approx \alpha^2 \frac{\left[E_m\left\{E_p\left(\sum_{i\in S} y_i^{-m}\right)\right\}\right]^2}{\left[E_m\left\{E_p\left(\sum_{i\in S} y_i^{1-m}\right)\right\}\right]^2}\left(\frac{AV\left(\sum_{i\in S} y_i^{-m}\right)}{\left[E_m\left\{E_p\left(\sum_{i\in S} y_i^{-m}\right)\right\}\right]^2}\right. \\
&\qquad \left. -\,2\,\frac{\operatorname{acov}\left(\sum_{i\in S} y_i^{-m}, \sum_{i\in S} y_i^{1-m}\right)}{E_m\left\{E_p\left(\sum_{i\in S} y_i^{-m}\right)\right\} E_m\left\{E_p\left(\sum_{i\in S} y_i^{1-m}\right)\right\}} + \frac{AV\left(\sum_{i\in S} y_i^{1-m}\right)}{\left[E_m\left\{E_p\left(\sum_{i\in S} y_i^{1-m}\right)\right\}\right]^2}\right) \\
&= \frac{\theta^2}{n}\left[1 + \frac{\Gamma(\alpha-m)\,\Gamma(\alpha+m)}{\Gamma(\alpha)^2} - 2\frac{\Gamma(\alpha+m)\,\Gamma(\alpha-m+1)}{\Gamma(\alpha)\,\Gamma(\alpha+1)}\right. \\
&\qquad\qquad \left. + \frac{\Gamma(\alpha-m+2)\,\Gamma(\alpha+m) + \left\{(n-1)\,N^{-1}\alpha - \alpha^2\right\}\Gamma(\alpha)^2}{\Gamma(\alpha+1)^2}\right].
\end{aligned}
$$

□

Corollary 7 (Special case: $m = 1$)**.** *When $m = 1$, the pseudo-likelihood estimator becomes*

$$\hat{\theta}_{PL} = \frac{\alpha}{n}\sum_{i\in S} y_i^{-1} \tag{C.3}$$

and the anticipated variance becomes

$$AV\left(\hat{\theta}_{PL}\right) \approx \frac{\theta^2}{n(\alpha-1)}\left\{1 + \frac{(n-1)(\alpha-1)}{N\alpha}\right\} = \frac{\theta^2}{n(\alpha-1)} + \frac{\theta^2(n-1)}{Nn\alpha} \tag{C.4}$$

for $\alpha > 1$, from (15).

Lemma 8. *If $X \sim Gamma(\alpha, \theta)$ and $m \geq \alpha$, then $E\left[X^{-m}\right]$ is infinite.*

Proof.

$$E\left[X^{-m}\right] = \int_0^\infty x^{-m}\frac{\theta^\alpha x^{\alpha-1}e^{-\theta x}}{\Gamma(\alpha)}\,dx = \frac{\theta^\alpha}{\Gamma(\alpha)}\int_0^\infty x^{-m+\alpha-1}e^{-\theta x}\,dx$$

by definition of the gamma function. For $x > 0$, $x^{-m+\alpha-1}e^{-\theta x}$ is always non-negative, so if $\int_0^1 x^{-m+\alpha-1}e^{-\theta x}\,dx$ is infinite, then $\int_0^\infty x^{-m+\alpha-1}e^{-\theta x}\,dx$ is also infinite. For $0 \leq x \leq 1$,

$e^{-\theta x} \geq e^{-\theta}$. Also, if $0 \leq x \leq 1$ and $m \geq \alpha$, then $x^{-m+\alpha-1}e^{-\theta x} \geq x^{-1}e^{-\theta x}$. If $m \geq \alpha$, then

$$\int_0^\infty x^{-m+\alpha-1}e^{-\theta x}\,dx \geq \int_0^1 x^{-m+\alpha-1}e^{-\theta x}\,dx \geq \int_0^1 x^{-1}e^{-\theta x}\,dx$$
$$\geq \int_0^1 x^{-1}e^{-\theta}\,dx = \log(x)\,e^{-\theta}\,\big]_0^1 = \infty\,.$$

□

Proof of (18) *(pseudo-likelihood estimator for case study of geological fault data).* Given the model (16), the probability weighted score function is

$$\mathrm{sc}_w(\theta) \propto \sum_{i\in S, y_i<h_1} \frac{\frac{1}{\theta} - y_i}{h_1} + \sum_{i\in S, h_1<y_i\leq h_2} \frac{\frac{1}{\theta} - y_i}{y_i} + \sum_{i\in S, y_i\geq h_2} \frac{\frac{1}{\theta} - y_i}{h_2}$$

and so

$$\hat{\theta} = \frac{\frac{n_1}{h_1} + \sum_{i\in S, h_1<y_i\leq h_2} \frac{1}{y_i} + \frac{n_3}{h_2}}{\sum_{i\in S, y_i<h_1} \frac{y_i}{h_1} + n_2 + \sum_{i\in S, y_i\geq h_2} \frac{y_i}{h_2}}$$

where n_1 is the number of sample units y_i such that $0 \leq y_i < h_1$, n_2 is the number of sample units y_i such that $h_1 \leq y_i < h_2$, and n_3 is the number of sample units y_i such that $h_2 \leq y_i$. □

References

Beaumont, J. F., 2008. A new approach to weighting and inference in sample surveys. Biometrika 95, 539–53.

Binder, D. A., December 1983. On the variances of asymptotically normal estimators from complex surveys. International Statistical Review 51 (3), 279–292.

Borgos, H. G., Omre, H., Townsend, C., 2002. Size distribution of geological faults: Model choice and parameter estimation. Statistical Modelling 2, 217–234.

Chambers, R. L., Steel, D. G., Wang, S., Welsh, A. H., 2012. Maximum Likelihood Estimation for Sample Surveys. CRC Press, Boca Raton.

Clark, R. G., 2020. Maximum likelihood estimation for outcome-dependent samples. Australian & New Zealand Journal of Statistics 62 (1), 49–70.

Gilbert, P., Varadhan, R., 2016. numDeriv: Accurate Numerical Derivatives. R package version 2016.8-1.
URL `https://CRAN.R-project.org/package=numDeriv`

Godambe, V., Thompson, M., August 1986. Parameters of superpopulation and survey population: Their relationships and estimation. International Statistical Review 54 (2), 127–138.

Gray, L., Gorman, E., White, I. R., Katikireddi, S. V., McCartney, G., Rutherford, L., Leyland, A. H., 2020. Correcting for non-participation bias in health surveys using record-linkage, synthetic observations and pattern mixture modelling. Statistical Methods in Medical Research 29 (4), 1212–1226.

Kopra, J., Härkänen, T., Tolonen, H., Karvanen, J., 2015. Correcting for non-ignorable missingness in smoking trends. STAT 4 (1).

Krieger, A. M., Pfeffermann, D., December 1992. Maximum likelihood estimation from complex sample surveys. Survey Methodology 18 (2), 225–239.

Lawless, J. F., 1997. Likelihood and pseudo likelihood estimation based on response-biased observation. In: Selected Proceedings of the Symposium on Estimating Functions. Vol. 32. Institute of Mathematical Statistics, pp. 43–55.

Levy, P. S., Lemeshow, S., 2008. Sampling of Populations: Methods and Application. John Wiley and Sons, Inc.

Lohr, S. L., 2021. Sampling: Design and Analysis, 3rd Edition. Chapman and Hall/CRC.

Lumley, T., 2011. Complex Surveys: A Guide to Analysis Using R. John Wiley & Sons.

Orchard, T., Woodbury, M. A., 1972. A missing information principle: Theory and applications. Proceedings of the 6th Berkeley Symposium on Mathematical Statistics 1, 697–715.

Patil, G. P., Ord, J. K., February 1976. On size-biased sampling and related form-invariant weighted distributions. Sankhyā: The Indian Journal of Statistics, Series B 38 (1), 48–61.

Pfeffermann, D., December 2011. Modelling of complex survey data: Why model? Why is it a problem? How can we approach it? Survey Methodology 37 (2), 115–136.

Pfeffermann, D., Krieger, A. M., Rinott, Y., October 1998. Parametric distributions of complex survey data under informative probability sampling. Statistica Sinica 8 (4), 1087–1114.

Pfeffermann, D., Moura, F. A. D. S., Silva, P. L. D. N., 2006. Multi-level modelling under informative sampling. Biometrika 93 (4), 943–59.

Rao, J., Wu, C., 2009. Handbook of Statistics. Vol. 29. Elsevier, Ch. 30 - Empirical Likelihood Methods, pp. 189–207.

Särndal, C., Swensson, B., Wretman, J., 1992. Model Assisted Survey Sampling. Springer series in statistics. Springer-Verlag.

Solow, A. R., Smith, W., September 1997. On fossil preservation and the stratigraphic ranges of taxa. Paleobiology 23 (3), 271–277.

Sugden, R. A., Smith, T. M. F., 1984. Ignorable and informative designs in survey sampling inference. Biometrika 71 (3), 495–506.