CENTRE FOR **STOCHASTIC GEOMETRY**
AND ADVANCED **BIOIMAGING**

Ina Trolle Andersen, Ute Hahn and Eva B. Vedel Jensen

Vanishing auxiliary variables in PPS sampling – with applications in microscopy

No. 01, February 2014

Vanishing auxiliary variables in PPS sampling – with applications in microscopy

Ina Trolle Andersen[1,2], Ute Hahn[1] and Eva B. Vedel Jensen[1]

[1]Department of Mathematics, Aarhus University
[2]Stereological Research Laboratory, Aarhus University

February 3, 2014

Abstract

Recently, non-uniform sampling has been suggested in microscopy to increase efficiency. More precisely, sampling proportional to size (PPS) has been introduced where the probability of sampling a unit in the population is proportional to the value of an auxiliary variable. Unfortunately, vanishing auxiliary variables are a common phenomenon in microscopy and, accordingly, part of the population is not accessible, using PPS sampling. We propose a modification of the design, for which an optimal solution can be found, using a model assisted approach. The optimal design has independent interest in sampling theory. We verify robustness of the new approach by numerical results, and we use real data to illustrate the applicability.

Keywords: microscopy, model assisted sampling, optimal allocation, proportional regression models, systematic PPS sampling, vanishing auxiliary variables

1 Introduction

Non-uniform sampling has considerable practical interest in microscopy, as the structures under study often show pronounced inhomogeneity. In these cases, when using uniform sampling, most of the sampled fields of view (FOV) will contain no or only little information of the feature of interest, and, as a consequence, the sampling becomes highly inefficient. Alternatively, one can use automatic computerized image analysis to provide measurements of auxiliary variables, which are expected to give information about the feature of interest. Combining this information with non-uniform sampling may then lead to a considerable reduction in estimator variance compared to the traditional systematic uniform sampling, see Gardi et al. (2008a,b).

This idea of empirical importance sampling has been given a stochastic formulation in Hansen et al. (2011), using point process theory. In the paper by Hansen et al.

Corresponding author: Ina Trolle Andersen, email: ita@imf.au.dk

(2011), statistical tools are developed for assessing the efficiency and constructing optimal model-based estimators of intensities in the class of generalized proportional regression models. These estimators can be used in practice, but several problems arise, which motivates further research.

One of the problems is that if the proportionality assumption is not met, the model-based estimator may be biased, which is unacceptable for the majority of researchers working in microscopy. Therefore it may be preferable to keep the original design-based Horvitz-Thompson estimator, which preserves unbiasedness regardless of proportionality or not, and focus on modifying the sampling design to improve efficiency of the estimator.

Another important problem which is not addressed in Hansen et al. (2011) are vanishing auxiliary variables, which occur in practical applications. The term refers to cases where there exist FOVs with the auxiliary variable equal to zero but with positive cell count. In a study of Keller et al. (2013), 10 % of the cells were in fact found in such FOVs. Unbiasedness of the original Horvitz-Thompson estimator requires positive inclusion probabilities for all FOVs with a positive cell count. In sampling proportional to size (PPS sampling), where size is measured by the auxiliary variables, one therefore has to change the sampling, if vanishing auxiliary variables can occur.

The workaround in microscopy, as suggested in Gardi et al. (2008a,b), is to add a small constant $\varepsilon > 0$ to all auxiliary variables before sampling. Current software (used in Keller et al. (2013)), uses by default an unrealistically small constant. In cases as the one described in Keller et al. (2013), the unbiasedness would therefore be paid for by an extremely high variance, if the default had been used. This is caused by the rare cases, where the sample includes the problematic FOVs mentioned above. On the other hand, large values of ε may decrease the efficiency one hopes to gain from PPS sampling, compared to uniform sampling. Therefore optimal ways of choosing such ε is important in practical application.

This problem is addressed in the present paper. We consider a sampling design for a finite population of units, numbered $\{1, \ldots, N\}$. The sample is a random subset $S \subseteq \{1, \ldots, N\}$ of the population with n elements, say. Some of the sampling units i have zero inclusion probabilities $\pi_i = P(i \in S)$, for instance,

$$\pi_1 = \cdots = \pi_{N_0} = 0,$$

where $N_0 < N$. We modify the design, such that the resulting sample still has size n, and such that it retains a constant *positive* inclusion probability π_0 for the units $1, \ldots, N_0$ and inclusion probabilities proportional to the original inclusion probabilities for the remaining units. Under mild regularity conditions, we find the optimal design of this type. This result, which is of independent interest in sampling theory, can be used to determine an optimal value of ε in the original problem described above.

The composition of the paper is as follows. The sampling set-up is presented in Section 2, while the optimal design is derived in Section 3, where it is also shown, that under a proportional regression model, the optimality result simplifies. This framework, where both design and model play a role, is often referred to as a model-assisted approach (Särndal et al., 2003). The robustness of the optimal design against

parameter misspecification and departures from proportionality is investigated in Section 4. An analysis of data from microscopy, using the developed methods, is presented in Section 5. Conclusions may be found in Section 6. A proof concerning an equivalence between systematic PPS sampling and stratified sampling is deferred to an appendix.

2 Set-up

We consider a finite population of N units and assume that a realization of a random variable Y_i (the variable of interest) is available for each unit i. Additionally, we assume that $Y_1, \ldots, Y_N$ are uncorrelated. The aim is to predict the population total, $T = \sum_{i=1}^{N} Y_i$, for a realization of $Y = \{Y_1, \ldots, Y_N\}$.

2.1 The Horvitz-Thompson predictor

Let $S \subseteq \{1, \ldots, N\}$ be a random sample of size n, independent of Y. A predictor of the population total, well-known from survey sampling theory, cf. Horvitz and Thompson (1952) and Särndal et al. (2003, p. 42), is the Horvitz-Thompson predictor

$$\widehat{T} = \sum_{i \in S} \frac{Y_i}{\pi_i}, \tag{2.1}$$

where $\pi_i = P(i \in S)$ is the probability that the ith unit is included in the sample. The sampling design is called non-uniform if the inclusion probabilities π_i are non-constant. The predictor $\widehat{T}$ is design-unbiased, i.e.

$$\mathbb{E}[\widehat{T}|Y] = T,$$

if the inclusion probabilities $\pi_i > 0$ are all positive. Under the assumption that $\pi_i > 0$ for all i, the design variance takes the form

$$\mathrm{Var}[\widehat{T}|Y] = \sum_{i=1}^{N} \sum_{j=i+1}^{N} (\pi_i \pi_j - \pi_{ij}) \left(\frac{Y_i}{\pi_i} - \frac{Y_j}{\pi_j} \right)^2, \tag{2.2}$$

where π_{ij} is the (i,j)th joint inclusion probability $\pi_{ij} = P(i \in S, j \in S)$.

The prediction error is defined as $\mathbb{E}[(\widehat{T} - T)^2]$. A sampling design is called optimal under a model for Y, if it minimizes the prediction error. Since $\widehat{T}$ is design-unbiased, the prediction error is equal to the mean design variance

$$\mathbb{E}[(\widehat{T} - T)^2] = \mathrm{Var}(\widehat{T}) - \mathrm{Var}(T) = \mathbb{E}[\mathrm{Var}[\widehat{T}|Y]].$$

If the inclusion probabilities are proportional to the mean values of the Y_is, i.e.

$$\pi_i \propto \mathbb{E}(Y_i),$$

then we have the following result for the mean variance

$$\mathbb{E}[\mathrm{Var}[\widehat{T}|Y]] = \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1 \right) \mathrm{Var}(Y_i). \tag{2.3}$$

Note that the mean variance only depends on the inclusion probabilities π_i and the variances of the Y_is. More detailed properties of the sampling design such as the joint inclusion probabilities do not appear in the formula.

2.2 Systematic PPS sampling

In order to obtain an efficient predictor of T, the information from a non-random auxiliary variable x_i associated with Y_i, $i = 1, \ldots, N$, may be used. We let $x = \{x_1, \ldots, x_N\}$.

Often, the inclusion probability π_i is chosen proportional to x_i (sampling proportional to size, PPS sampling), as one expects x_i to be roughly proportional to Y_i. For a PPS sample S of size n, we have

$$\pi_i = n\frac{x_i}{x_\cdot},$$

where $x_\cdot = \sum_{i=1}^{n} x_i$.

A PPS sampling scheme, which is widely used in sampling due to its simplicity and efficiency, is systematic PPS sampling. This design was originally introduced in Madow (1949), see also Murthy et al. (1967), Iachan (1982) and Särndal et al. (2003, Section 3.6) for more details and references. It can be implemented as follows. Let $1, \ldots, N$ refer to an ordering of the units. Sampling is performed on cumulative weights with a random starting point in $[0, \frac{x_\cdot}{n}]$, followed by equidistant selections of the units. More precisely, let $W_i = \sum_{j=1}^{i} x_j$, $i = 1, \ldots, N$, denote the cumulated weights with $W_0 = 0$. Let $V_1 \sim \text{unif}([0, \frac{x_\cdot}{n}])$, independent of Y and x, and let $V_j = V_1 + (j-1)\frac{x_\cdot}{n}, j = 2, \ldots, n$. Then, the sample S consists of those units i for which $[W_{i-1}, W_i]$ contains at least one V_j. Under the assumption that each interval $[W_{i-1}, W_i]$ can contain at most one V_j, we have

$$\sum_{j=1}^{n} \mathbf{1}\{V_j \in [W_{i-1}, W_i]\} \leq 1$$

for all i, and the inclusion probabilities take the intended form

$$\begin{aligned}
\pi_i &= P(\sum_{j=1}^{n} \mathbf{1}\{V_j \in [W_{i-1}, W_i]\} = 1) \\
&= \mathbb{E}(\sum_{j=1}^{n} \mathbf{1}\{V_j \in [W_{i-1}, W_i]\}) \\
&= n\frac{x_i}{x_\cdot}.
\end{aligned}$$

Figure 1 illustrates systematic PPS sampling with an ordering according to the size of the auxiliary variable x. A specific variant of systematic PPS sampling, called the proportionator, was suggested in Gardi et al. (2008a,b) for analysis of microscopy images. The design uses the principles of the so-called smooth fractionator (Gundersen, 2002) to order the sampling units, which corresponds to the balanced systematic sampling described in Murthy et al. (1967, Section 5.9d). If we let $1, \ldots, N$ denote

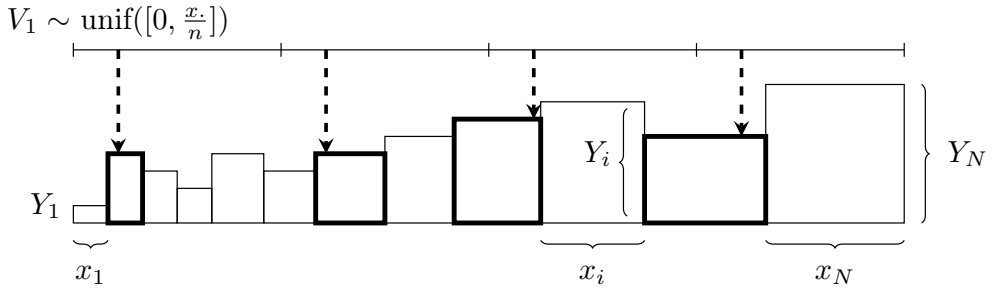

Figure 1: Illustration of systematic PPS sampling. Each box represents a sampling unit i with height given by the variable of interest Y_i, and width given by the auxiliary variable x_i. The sampling units have been ordered according to the sizes of the auxiliary variables, and sampling is then performed on the cumulative weights from a random starting point in $[0, \frac{x_\cdot}{n}]$, followed by equidistant selections of the units.

the ordering such that $x_1 \leq x_2 \leq \ldots \leq x_N$, $M = \frac{N}{2}$ if N is even and $M = \frac{N+1}{2}$ if N is odd, the smooth ordering becomes $[1], \ldots, [N]$, where

$$[i] = \begin{cases} 2i - 1, & i \leq M, \\ 2(N + 1 - i), & i > M. \end{cases} \tag{2.4}$$

This alternative ordering has been proven to be superior to the ordinary uniform systematic sampling design, e.g. when linear trend is present (Bellhouse and Rao, 1975), and the efficiency is illustrated in Gundersen (2002).

In the sampling literature, there also exists a with-replacement version of PPS sampling. Here, the predictor takes the form (Hansen and Hurwitz, 1943)

$$\widehat{T}^{WR} = \frac{1}{n} \sum_{i=1}^{N} \#\{i \in S\} \frac{Y_i}{p_i},$$

where $\#\{i \in S\}$ denotes the number of times unit i is sampled and the draw-by-draw inclusion probability of unit i is given by $p_i = x_i/x_\cdot$.

2.3 ε-corrected PPS sampling

In the present paper, we will address the problem of vanishing auxiliary variables, where there exist $i \in \{1, \ldots, N\}$, such that $x_i = 0$ and $Y_i > 0$. As a consequence, if PPS sampling is used there exist units i with $Y_i > 0$, but $\pi_i = 0$ and the Horvitz-Thompson predictor (2.1) will be biased. To adjust for this, one can add a small constant $\varepsilon > 0$ to the auxiliary variables which are zero. The resulting PPS sampling design will be called ε-corrected.

Let $N_0 = \#\{i | x_i = 0\}$, and suppose that the units are ordered such that

$$x_1 = \cdots = x_{N_0} = 0.$$

Then, the inclusion probabilities of the ε-corrected PPS sampling design with sample

size n become

$$\pi_i = \begin{cases} n\dfrac{\varepsilon}{x. + N_0\varepsilon}, & i = 1, \ldots, N_0, \\ n\dfrac{x_i}{x. + N_0\varepsilon}, & i = N_0 + 1, \ldots, N. \end{cases} \tag{2.5}$$

It is important that ε is not chosen too small. When ε is chosen unrealistically small, like it was done in microscopy until recently, the result is an extremely large variance. In fact, with inclusion probabilities as specified in (2.5), $\operatorname{Var}\widehat{T} \to \infty$, when $\varepsilon \to 0$, if $Y_i > 0$ for just one $i \in \{1, \ldots, N_0\}$. To see this, note that

$$\operatorname{Var}\widehat{T} \geq \sum_{i=1}^{N} \frac{1}{\pi_i} Y_i^2 - T^2 \to \infty, \text{ as } \varepsilon \to 0. \tag{2.6}$$

On the other hand, ε should not be chosen too large, because then the sampling is directed towards the first N_0 units and a possible proportionality between x_i and Y_i among units with $x_i > 0$ is not utilized in the sampling.

The ε-corrected systematic PPS sampling can be considered as a kind of stratification, based on the auxiliary variables, cf. Figure 2. This observation opens up for the possibility of finding an optimal ε, using optimal allocation in stratified sampling. Early references on optimal allocation are Neyman (1934), Stuart (1954) and Rao (1968), see also Murthy et al. (1967, Section 7) and Särndal et al. (2003, Section 3.7). Stratification is a standard variance reduction technique in sampling,

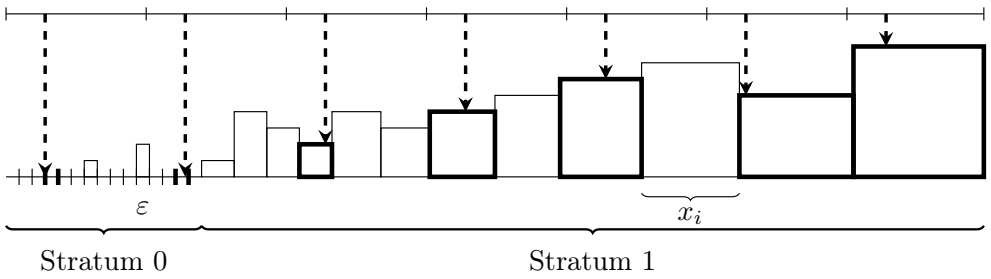

Figure 2: Illustration of ε-corrected systematic PPS sampling, where a small constant ε has been added to each unit with $x_i = 0$ to ensure an unbiased predictor. Due to the systematic sampling, the sampling scheme has a build-in stratification mechanism, such that an almost fixed fraction of the sampled units will be in each stratum.

where the population is divided into strata and independent samples are taken from each stratum. In ε-corrected PPS sampling, we can regard the population as divided into two strata, Stratum 0 consisting of the sampling units with $x_i = 0$ and Stratum 1 consisting of the sampling units with $x_i > 0$. If we let $U_0 = \{1, \ldots, N_0\}$ and $U_1 = \{N_0 + 1, \ldots, N\}$ be the notation used for the two strata, we have for the Horvitz-Thompson predictor based on ε-corrected PPS sampling

$$\widehat{T} = \sum_{i \in S} \frac{Y_i}{\pi_i} = \sum_{i \in S_0} \frac{Y_i}{\pi_i} + \sum_{i \in S_1} \frac{Y_i}{\pi_i} = \widehat{T}_0 + \widehat{T}_1,$$

say, where $S_h = S \cap U_h$, $h = 0, 1$. In the case where the expected number

$$n_0 = n\frac{N_0\varepsilon}{x. + N_0\varepsilon} \tag{2.7}$$

of units sampled from Stratum 0 is an integer, ε-corrected systematic PPS sampling as illustrated in Figure 2 yields same mean variance as a stratified scheme with two independent systematic samples, one in each stratum. In the general case where n_0 is not necessarily an integer, it is possible to derive an expression for the difference between the mean variances under the two designs, see Theorem 1 below. The proof of Theorem 1 is deferred to the Appendix.

Theorem 1. *For an ε-corrected systematic PPS sample S, let*

$$\widehat{T} = \widehat{T}_0 + \widehat{T}_1 \quad \textit{with} \quad \widehat{T}_h := \sum_{i \in S \cap U_h} \frac{Y_i}{\pi_i}.$$

Let $\widehat{T}^{st}$ be a random variable, distributed as $\widehat{T}_0^{st} + \widehat{T}_1^{st}$, where $\widehat{T}_0^{st}$ and $\widehat{T}_1^{st}$ are independent random variables, and $\widehat{T}_h^{st}$ is distributed as $\widehat{T}_h$, $h = 0, 1$. Under the assumption $\mathbb{E}\, Y_i = \mathbb{E}\, Y_j, j, i \in U_0$,

$$\mathbb{E}[\mathrm{Var}[\widehat{T} \mid Y]] = \mathbb{E}[\mathrm{Var}[\widehat{T}^{st} \mid Y]] - 2\nu(1-\nu)\bar{\tau}_0\bar{\tau}_1,$$

where $\nu = n_0 - \lfloor n_0 \rfloor$, $\bar{\tau}_0 := \mathbb{E}\, T_0/n_0$ and

$$\bar{\tau}_1 := \begin{cases} \mathbb{E}\, T_1/n_1 \textit{ (or some arbitrary constant)}, & \nu = 0 \\ \dfrac{1}{\nu}\, \mathbb{E}\left[\mathbb{E}(\widehat{T}_1 - T_1 \mid S) \,\middle|\, \#(S \cap U_0) = \lfloor n_0 \rfloor \right], & \nu > 0, \end{cases}$$

where $n_1 = n - n_0$.

For a model with proportionality in Stratum 1 between $\mathbb{E}\, Y_j$ and π_j such that $\mathbb{E}\, Y_j/\pi_j = \mathbb{E}\, T_1/n_1$ for all $N_0 + 1 \le j \le N$, we get

$$\mathbb{E}(\widehat{T}_1 - T_1 \mid S) = \sum_{j \in S \cap U_1} \frac{\mathbb{E}\, Y_j}{\pi_j} - \mathbb{E}\, T_1 = (\#(S \cap U_1) - n_1)\frac{\mathbb{E}\, T_1}{n_1}.$$

Thus, the definition of $\bar{\tau}_1$ reduces to $\bar{\tau}_1 = \mathbb{E}\, T_1/n_1$, and the difference between the mean variances of the two predictors becomes

$$\mathbb{E}[\mathrm{Var}[\widehat{T} \mid Y]] - \mathbb{E}[\mathrm{Var}[\widehat{T}^{st} \mid Y]] = -2\nu(1-\nu)\frac{\mathbb{E}\, T_0}{n_0}\frac{\mathbb{E}\, T_1}{n_1}.$$

In that case, the original design leads to a slightly smaller mean variance than the stratified design, but, with growing sample size n, the difference decreases of order n^{-2}.

Since the difference between the variance of the stratified and the original design is small, only stratified sampling will from this point be considered. We find in the next section an optimal ε, based on optimal allocation in stratified sampling.

3 An optimal stratified design

Consider a sampling design with sample size n and inclusion probabilities $\tilde{\pi}_i$ such that $\sum_{i=1}^{N} \tilde{\pi}_i = n$. Let us suppose that

$$\begin{aligned} \tilde{\pi}_i &= 0, \quad i = 1, \ldots, N_0, \\ \tilde{\pi}_i &> 0, \quad i = N_0 + 1, \ldots, N. \end{aligned}$$

We want to modify the design such that the first N_0 units are assigned a constant positive inclusion probability and the remaining units have inclusion probabilities proportional to the original ones. If we let n_0 be the expected sample size among the first N_0 units, the modified sampling design will have the following inclusion probabilities

$$\pi_i = \begin{cases} \dfrac{n_0}{N_0}, & i = 1, \ldots, N_0, \\ \left(1 - \dfrac{n_0}{n}\right) \tilde{\pi}_i, & i = N_0 + 1, \ldots, N. \end{cases} \tag{3.1}$$

The theorem below gives the optimal stratified design of this type. Here, as in the previous section, stratification refers to a division of the population into two strata, Stratum 0 $(1, \ldots, N_0)$ and Stratum 1 $(N_0+1, \ldots, N)$. The result holds under the following model assumptions

$$\mathbb{E}(Y_i) = \beta_0, \quad V(Y_i) = \sigma_0^2, \quad i = 1, \ldots, N_0, \tag{3.2}$$

$$\mathbb{E}(Y_i) \propto \tilde{\pi}_i, \quad V(Y_i) = \sigma_i^2, \quad i = N_0 + 1, \ldots, N, \tag{3.3}$$

where $\sigma_0^2 > 0$ and $\sigma_i^2 > 0, i = N_0 + 1, \ldots, N$.

Theorem 2. *Let S be a sampling design with positive inclusion probabilities of the form* (3.1) *and suppose that* (3.2) *and* (3.3) *hold. Then, under stratified sampling, the expected variance $\mathbb{E}[\mathrm{Var}[\widehat{T}^{st}|Y]]$ of the Horvitz-Thompson predictor $\widehat{T}^{st} = \sum_{i \in S} Y_i/\pi_i = \widehat{T}_0^{st} + \widehat{T}_1^{st}$, where $\widehat{T}_0^{st}$ and $\widehat{T}_1^{st}$ are based on independent samples S_0 and S_1 in Stratum 0 and Stratum 1, respectively, is minimized if the sample size in Stratum 0 is chosen as*

$$n_0 = \min\left(n \frac{N_0}{N_0 + \sqrt{n \sum_{i=N_0+1}^{N} \sigma_i^2/(\sigma_0^2 \tilde{\pi}_i)}}, N_0 \right). \tag{3.4}$$

Proof. The expected variance of $\widehat{T}^{st}$ is

$$\mathbb{E}[\mathrm{Var}[\widehat{T}^{st}|Y]] = \mathbb{E}[\mathrm{Var}[\widehat{T}_0^{st}|Y]] + \mathbb{E}[\mathrm{Var}[\widehat{T}_1^{st}|Y]],$$

since, conditionally on Y, $\widehat{T}_0^{st}$ and $\widehat{T}_1^{st}$ are independent. Within each stratum the mean values of Y_i is proportional to the inclusion probabilities π_i and, using (2.3) on each stratum separately, we find

$$\mathbb{E}[\mathrm{Var}[\widehat{T}^{st}|Y]] = N_0 \left(\frac{N_0}{n_0} - 1 \right) \sigma_0^2 + \sum_{i=N_0+1}^{N} \left(\frac{1}{\left(1 - \frac{n_0}{n}\right) \tilde{\pi}_i} - 1 \right) \sigma_i^2 = f(n_0),$$

say. We find

$$f'(n_0) = -\frac{N_0^2 \sigma_0^2}{n_0^2} + \frac{V}{n \left(1 - \frac{n_0}{n}\right)^2},$$

where

$$V = \sum_{i=N_0+1}^{N} \frac{\sigma_i^2}{\tilde{\pi}_i}.$$

The equation $f'(n_0) = 0$ is equivalent to the following equation

$$\left(\frac{N_0^2}{n}\sigma_0^2 - V\right) n_0^2 - 2\sigma_0^2 N_0^2 n_0 + n\sigma_0^2 N_0^2 = 0,$$

which has the following two solutions

$$n_0 = \frac{nN_0}{N_0 - \sqrt{nV/\sigma_0^2}},\ \frac{nN_0}{N_0 + \sqrt{nV/\sigma_0^2}}.$$

Only the second solution will result in a minimum of $f(n_0)$. Using that $n_0 \leq N_0$, we get (3.4). □

In the case of PPS sampling with probabilities according to an auxiliary variable, as described in the previous section, we have $\tilde{\pi}_i = nx_i/x.$ with $x_1 = \cdots = x_{N_0} = 0$. It follows from Theorem 2 that under PPS sampling with (3.2) and (3.3) fulfilled, the optimal allocation becomes

$$n_0 = \min\left(n\frac{N_0}{N_0 + \sqrt{x. \sum_{i=N_0+1}^{N} \sigma_i^2/(\sigma_0^2 x_i)}}, N_0\right). \tag{3.5}$$

We will from now on assume that $n_0 < N_0$ such that n_0 is equal to the first of the two terms inside the minimum sign. Using the relation (2.7) between n_0 and ε, the ε minimizing $\mathbb{E}[\mathrm{Var}[\widehat{T}^{st}|Y]]$ then becomes

$$\varepsilon = \frac{\sqrt{\sigma_0^2 x.}}{\sqrt{\sum_{i=N_0+1}^{N} \sigma_i^2/x_i}}. \tag{3.6}$$

Further simplifications are possible if we assume that the Y_is in Stratum 1 fulfil a proportional regression model with $1 \leq g \leq 2$,

$$\mathbb{E}(Y_i) = \beta_1 x_i,\ \mathrm{Var}(Y_i) = \sigma_1^2 x_i^g, \quad i = N_0 + 1, \ldots, N \tag{3.7}$$

and, in addition, the mean-variance relationship is the same in the two strata, i.e.

$$\frac{\sigma_0^2}{\beta_0^g} = \frac{\sigma_1^2}{\beta_1^g}. \tag{3.8}$$

Then, the optimal allocation becomes

$$n_0 = n\frac{(\beta_0/\beta_1)^{g/2} N_0}{(\beta_0/\beta_1)^{g/2} N_0 + \sqrt{x.(x^{g-1}).}}, \tag{3.9}$$

where $(x^{g-1}). = \sum_{i=1}^{N} x_i^{g-1} = \sum_{i=N_0+1}^{N} x_i^{g-1}$. The optimal allocation in (3.9) can alternatively be expressed, using the natural parameter $q = \mathbb{E}(T_0)/\mathbb{E}(T_1)$, where $T_0 = \sum_{i=1}^{N_0} Y_i$ and $T_1 = \sum_{i=N_0+1}^{N} Y_i$. We find

$$n_0 = n\frac{\sqrt{N_0^{2-g}(qx.)^g}}{\sqrt{N_0^{2-g}(qx.)^g} + \sqrt{x.(x^{g-1}).}}. \tag{3.10}$$

In the special cases with $g = 1$ and $g = 2$, we find

$$n_0 = \begin{cases} n\dfrac{\sqrt{qk}}{1+\sqrt{qk}}, & g = 1, \\ n\dfrac{q}{1+q}, & g = 2, \end{cases} \tag{3.11}$$

where $k = N_0/(N-N_0)$. Under the model specified in (3.2), (3.7) and (3.8), $q = \frac{\beta_0 N_0}{\beta_1 x_\cdot}$.

Notice that under the assumptions of Theorem 2, the optimal choice of n_0 (or ε) does not depend on joint inclusion probabilities within the strata.

4 Robustness

In this section, we investigate the robustness of the optimal allocation under the extended proportional regression model against departures from this model and parameter misspecification. We study the relative inflation in mean variance

$$R = \frac{\mathbb{E}[\mathrm{Var}[\widehat{T}(n_0')|Y]]}{\mathbb{E}[\mathrm{Var}[\widehat{T}(n_0)|Y]]},$$

where n_0 and n_0' are calculated according to (3.10), n_0 with the true values of the parameters and the correct model, and n_0' with alternative parameter values or the wrong model, and $\widehat{T}(.)$ is the resulting estimator.

4.1 Robustness against parameter misspecification

Consider the case that the extended proportional regression model given by (3.2), (3.7) and (3.8) holds, but that the parameter g is misspecified. This parameter controls the mean-variance relation, viz.

$$\mathrm{Var}\, Y_i \propto (\mathbb{E}\, Y_i)^g.$$

In Table 1, R is shown for the case $g = 1$. For n_0', the true value of q is used, but g is wrongly assumed to be 2. Using (3.11), we find

$$R = \frac{(1+q)(1+k)(1-\frac{n}{N})}{(1+\sqrt{qk})^2 - (1+q)(1+k)\frac{n}{N}}. \tag{4.1}$$

Table 1 shows the value of R for different combinations of k and q, with $n/N = 0.1$. The range of k and q has been chosen such that it represents what is expected in the application we have in mind, see Section 5. It is seen, that the expected variance is increased quite markedly if we wrongly assume that $g = 2$. The same conclusion holds for other choices of n/N. We also considered the 'opposite' case, where the true value of g is 2, but g is wrongly assumed to be 1. In this case, R also depends on the realized values of the auxiliary variable. With $C_x = N(x^2)_\cdot/((1+k)x_\cdot^2)$, we get

$$R = \frac{(1+\sqrt{qk})(1+\frac{q^2}{\sqrt{qk}}) - \frac{n}{N}(1+k)(\frac{q^2}{k}+C_x)}{(1+q)^2 - \frac{n}{N}(1+k)(\frac{q^2}{k}+C_x)}. \tag{4.2}$$

$k \setminus q$	0.025	0.05	0.10	0.15
0.25	1.113	1.069	1.028	1.010
0.50	1.278	1.197	1.115	1.071
1.00	1.624	1.468	1.309	1.222
2.00	2.326	2.000	1.683	1.514
4.00	3.781	3.011	2.341	2.010

Table 1: The relative inflation R in mean variance, given by (4.1), by wrongly using $g = 2$, when the true value is $g = 1$. The results are for varying values of q and k, when $n/N = 0.1$.

If the auxiliary variables $x_{N_0+1}, \ldots, x_N$ are i.i.d. realizations of a random variable X, we have $C_x \approx \mathbb{E}[X^2]/\,\mathbb{E}[X]^2$. The assumption $\pi_i = (n - n_0)x_i/x. \leq 1$ excludes very skew distributions, thus it prevents large values of C_x. In the case illustrated in Table 2, where we let $C_x = 1.3$, R was much closer to one than in the previous case where $g = 1$, but $g = 2$ is assumed. The same conclusion holds for other choices of n/N. Table 3 shows values of R, when n_0' is calculated, using the true values of $g = 1$ and $k = 2$, and varying values of a guessed value $\hat{q}$ and a true value of q. The formula is here

$$R = \frac{(1 + \sqrt{\hat{q}k})(1 + \sqrt{\frac{q^2}{\hat{q}}k}) - (1 + q)(1 + k)\frac{n}{N}}{(1 + \sqrt{qk})^2 - (1 + q)(1 + k)\frac{n}{N}}. \tag{4.3}$$

These results indicate that R is robust against misspecification of q.

4.2 Robustness against departures from proportionality

Let us now investigate the robustness of the optimal allocation based on the proportional regression model against departures from proportionality between x_i and $\mathbb{E}\,Y_i$. The model used in the robustness investigations is inspired by the applications we have in mind. We suppose that

$$\begin{array}{ll} Y_i \sim \text{pois}(\beta_0), & i = 1, \ldots, N_0, \\ Y_i | X_i = x_i \sim \text{pois}(x_i^\delta), & i = N_0 + 1, \ldots, N. \end{array} \tag{4.4}$$

For all choices of δ, this model represents a mean-variance relation with $g = 1$, i.e., $\text{Var}\,Y_i \propto (\mathbb{E}\,Y_i)^1$. If $\delta = 1$, the extended proportional regression model, specified in (3.2), (3.7) and (3.8), holds with $\sigma_0^2 = \beta_0$ and $\sigma_1^2 = \beta_1$. If instead $\delta \neq 1$, the proportional regression relationship between x_i and Y_i does not hold for $i > N_0$. We study the consequences of using n_0 as given in (3.11) with $g = 1$, even though the underlying assumptions are not fulfilled. In contrast to the case that $\delta = 1$, the (true) mean variance is now influenced by the specific design. We focus on systematic PPS sampling in Stratum 1, with inclusion probabilities proportional to the x_is.

Various distributions of the auxiliary variable have been tested. Here, we present the results for the case where x_i is a realization of

$$X_i \sim \text{beta}(\gamma_1, \gamma_2)\rho + \tau, \tag{4.5}$$

$i = N_0 + 1, \ldots, N$. Figure 3 shows scatterplots of (X, Y) for realizations of the model given by (4.4) and (4.5) with δ=0.5, 1 and 2, respectively. The parameter β_0 is chosen such that the parameter $q = \mathbb{E}(T_0)/\mathbb{E}(T_1)$ is the same in all three cases. The relative inflation R in mean variance due to allocation following (3.11) with

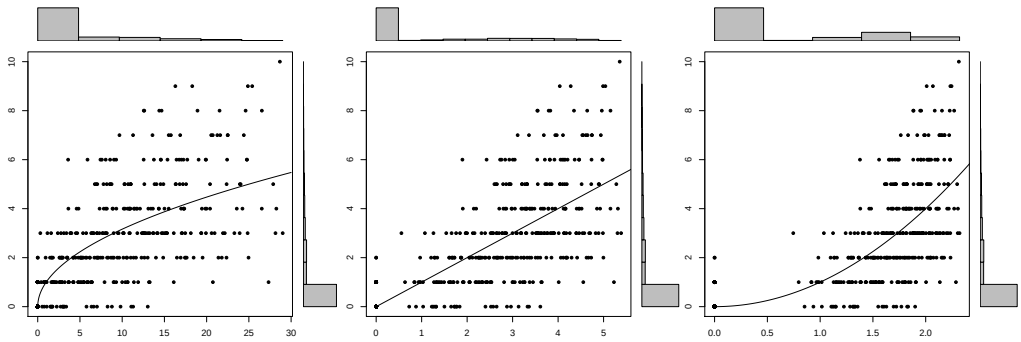

Figure 3: Scatterplots of (X, Y) for realizations of the model given by (4.5) and (4.4) with (left to right) δ=0.5, 1 and 2, respectively, together with the mean relation $Y = X^{\delta}$. The parameters in the distribution (4.5) are $\gamma_1 = \gamma_2 = 2$, $\rho = 5$, $\tau = 0.5$ while the parameter q is equal to 0.05 in all three cases. The values of N and N_0 are $N = 1000$ and $N_0 = 2/3N$. The empirical marginal distributions of X and Y is shown on the upper and right side of the graphs. For more details, see the text.

$g = 1$ was calculated for various values of q and δ. The correct value of q was used in the allocation. For each pair of parameters q and δ, one realization of the X_is was considered, and the true optimal value of n_0, which is needed for calculation of the denominator of R, was determined. In all the cases considered, $0.025 \leq q \leq 0.15$ and $0.5 \leq \delta \leq 2$, optimal allocation assuming proportionality showed robustness against departures from proportionality ($R \in [1; 1.03]$). In Figure 4, the mean variance is shown for n fixed ($n = 100$) as a function of the sample proportion n_0/n in Stratum 0. The variances are shown for simple random sampling (SRS), PPS sampling with replacement (WR), and systematic PPS sampling, all under stratification. Note that the variances are only shown for a range of the values of n_0, as the variance becomes very large for extreme choices of n_0. This emphasizes the importance of good choices of n_0. Although the variances differ, the optimal allocations are almost identical for PPS with replacement and systematic PPS sampling. In the case $\delta = 0.5$, shown to the left of Figure 4, we gain much from using systematic sampling, as PPS WR and systematic PPS sampling, differ the most in this case. Here, SRS actually performs better than PPS WR. This can be mainly ascribed to stratification, as without stratification, the variance of SRS is approximately 35.000 in the case of $\delta = 0.5$.

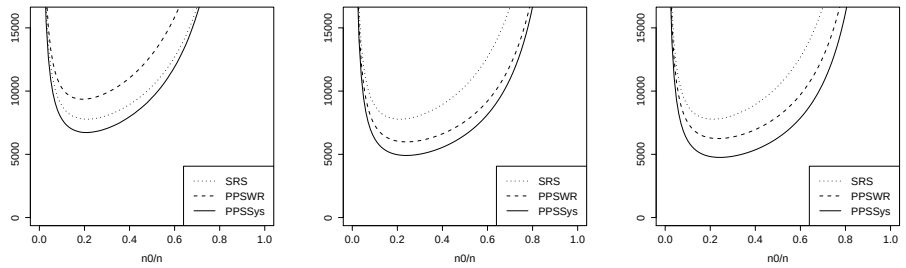

Figure 4: The mean variance under simple random sampling (SRS), PPS sampling with replacement (WR) and systematic PPS sampling, all under stratification, is shown for $q = 0.05$ and (from left to right) δ=0.5, 1 and 2, respectively, as a function of the sample proportion in Stratum 0. The remaining parameter values are specified in Figure 3. For more details, see the text.

$k \setminus q$	0.025	0.05	0.10	0.15
0.25	1.042	1.036	1.020	1.009
0.50	1.079	1.082	1.068	1.050
1.00	1.142	1.160	1.156	1.137
2.00	1.267	1.315	1.329	1.311
4.00	1.670	1.780	1.802	1.752

Table 2: The relative inflation R in the mean variance, given by (4.2), by wrongly using $g = 1$, when the true value is $g = 2$. The results are for varying values of q and k, when $n/N = 0.1$ and $C_x = 1.3$.

$\hat{q} \setminus q$	0.025	0.05	0.10	0.15
0.025	1.000	1.030	1.133	1.185
0.05	1.023	1.000	1.031	1.055
0.10	1.096	1.028	1.000	1.002
0.15	1.155	1.065	1.009	1.000

Table 3: The relative inflation R in mean variance, given by (4.3), for varying values of the guessed value $\hat{q}$ and the true value q, when $n/N = 0.1$, $g = 1$ and $k = 2$.

5 Analyzing data from microscopy

In this section, we use the developed methods in the analysis of a data set from microscopy (Keller et al., 2013), collected with the purpose of estimating osteoclast cell numbers in paws from mice with experimental arthritis. The tissue sections analyzed were divided by a grid into N small fields of view (FOVs) or observation windows. The random variable Y_i is the number of cells in FOV i, while x_i indicates the amount of a pre-chosen colour in FOV i associated with the staining of the cells. The x_i values are easily determined by automatic image analysis at a low magnification. This is in contrast to the cell counts Y_i which are time-consuming to determine, as they have to be done at high magnification by an expert-user.

The data set from Keller et al. (2013) is unique in the sense that it is exhaustive comprising 100 % of FOVs and covering the total section areas. Figure 5 shows a scatterplot of the auxiliary variable x and the number of cells y in each of the $N = 2703$ FOVs. As the population analyzed in Keller et al. (2013) is completely

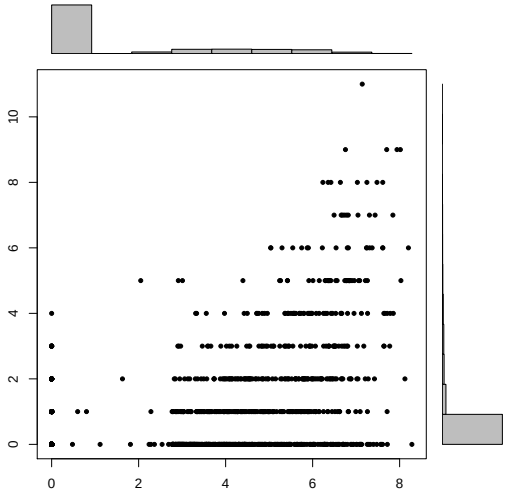

Figure 5: Scatterplot of the auxiliary variable x and the number of cells y in each of the $N = 2703$ FOVs. There are $N_0 = 1915$ FOVs with $x = 0$.

known, containing 10 % cells in FOVs with $x = 0$ ($q = 1/9$), these data are suitable for testing how far the allocation provided by the new approach using models in combination with optimal allocation is from the actual optimum. We first check the proportionality assumption and choose model parameters, and then study the variance as a function of allocation.

5.1 Proportionality

In Figure 6, it is investigated whether a linear relationship between x and y is a satisfactory description of the data in Stratum 1, consisting of $N_1 = 788$ FOVs. The data was partitioned into bins of size 35 (± 1), resulting in a total number of 22 bins.

The mean proportionality in Stratum 1 is not fulfilled, see the left panel in Figure 6. A linear regression on the log-transformed and binned x and y gave a

much more satisfactory description, see Figure 6 right panel. The estimated relation is $\log y = -3.93 + 2.40 \log x$, which corresponds to a model with $\delta = 2.4$.

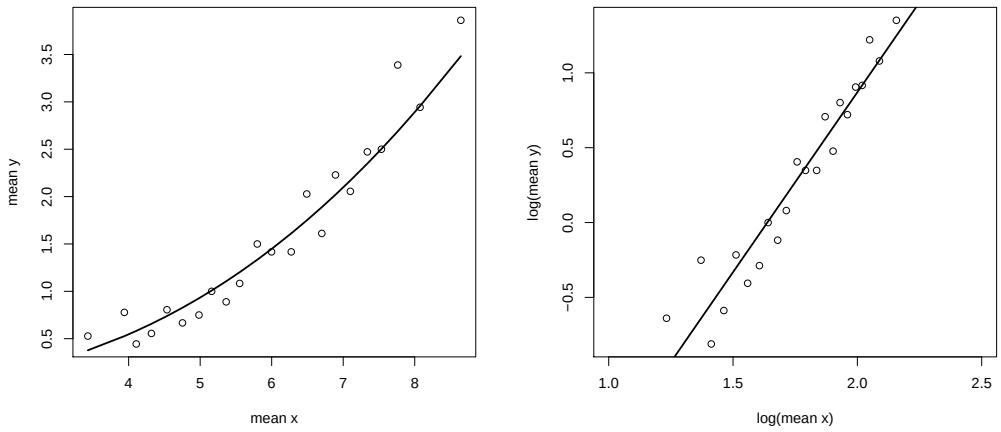

Figure 6: Means (left) and log means (right) of x and y-values in 22 bins in Stratum 1. The curve in the left panel was obtained by transforming the regression line from the right panel.

5.2 Relation between mean and variance

Under the extended proportional regression model, it is assumed that $\operatorname{Var} Y_i \propto \mathbb{E}\, Y_i^g$ holds in Stratum 1, cf. (3.7). To choose an appropriate value for g for use in (3.11), we compare the empirical means $\bar{y}$ and variances s^2 estimated in the bins in Stratum 1 as specified in Subsection 5.1.

Linear regression of the log transformed s^2 and $\bar{y}$ gives a relation $\log s^2 = 0.44 + 1.22 \log \bar{y}$, which is shown in the left panel of Figure 7 (full drawn line). Although the slope 1.22 is significantly different from 1 ($p = 0.021$), we will use $g = 1$ in the further investigations. For completeness, a line with slope one is also shown in the left panel (dotted line). Figure 7, right panel, shows the same estimated relations in a mean-variance plot by transformation of the line in the left panel (full drawn curve), together with a fitted line through the origin (stippled line).

An additional assumption of the model is given in (3.8), which means that both strata have the same mean-variance relation, i.e., the ratio $\operatorname{Var} Y_i / \mathbb{E}\, Y_i^g$ must be the same. Here we assume $g = 1$. While $s^2/\bar{y} = 1.69$ in Stratum 0, the ratio was found to be 2.03 in Stratum 1, thus (3.8) with $g = 1$ appears only approximately fulfilled.

5.3 Optimal allocation

Following the results of Section 5.1 and 5.2, we will describe the data by a proportional regression model as used in the simulations in Section 4, with $g = 1$, $\sigma^2/\beta = 2$, $q = 1/9$ and $\delta = 2.4$. The simulations in Section 4 indicated that moderate departures from proportionality ($\delta \neq 1$) are not critical for optimal allocation. It thus seems reasonable to use the optimal allocation given in (3.11) for $g = 1$. We

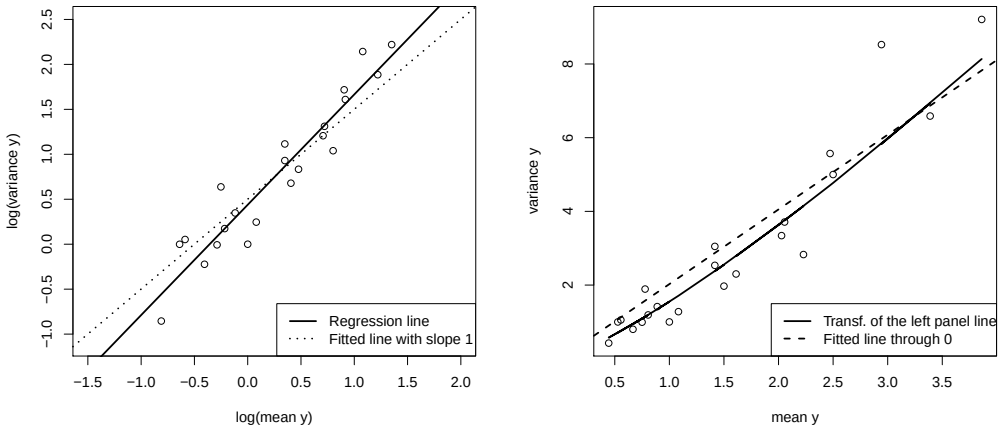

Figure 7: The left panel shows log-variances plotted against log-means of y in each of the 22 bins from Stratum 1, together with a regression line (full drawn) and a fitted line with slope 1 (dotted). The right panel shows variances plotted against means of y in each bin, together with a transformation of the left panel regression line (full drawn) and a fitted line through 0 (stippled).

investigate how well this fits the actual optimum, and compare with the allocation given by (3.11) for $g = 2$.

Figure 8 shows the variance of stratified SRS, stratified PPS WR and stratified systematic PPS (with a balanced ordering of the auxiliary variable, see (2.4)) as functions of the proportion of the sample of size $n = 0.10N$, allocated in Stratum 0. The variances of PPS WR and SRS are smooth functions, whereas the variance of systematic PPS shows a more complicated behaviour, due to the systematic sampling. It is however clear that the variance of systematic PPS becomes very large if n_0 is chosen too small, and in most cases the variance is smaller than the one for PPS WR and SRS with the same allocation. An overall impression of the variance of systematic PPS is obtained by binning of size 20 except for the small (and large) values of n_0, which removes the huge fluctuations, see Figure 8, right panel. Using the binned data, optimal allocation based on (3.10) for $g = 1$, corresponding to $n_0 = 0.34n$, very well fits the true optimum. If instead $g = 2$ is used, corresponding to $n_0 = 0.10n$, the variance becomes a factor 2 larger. From (2.7) we get that $n_0 = 0.34n$ corresponds to $\varepsilon = 1.05$ and $n_0 = 0.10n$ corresponds to $\varepsilon = 0.23$, hence $\varepsilon = 1$ used in Keller et al. (2013) was in fact remarkable close to the optimal choice.

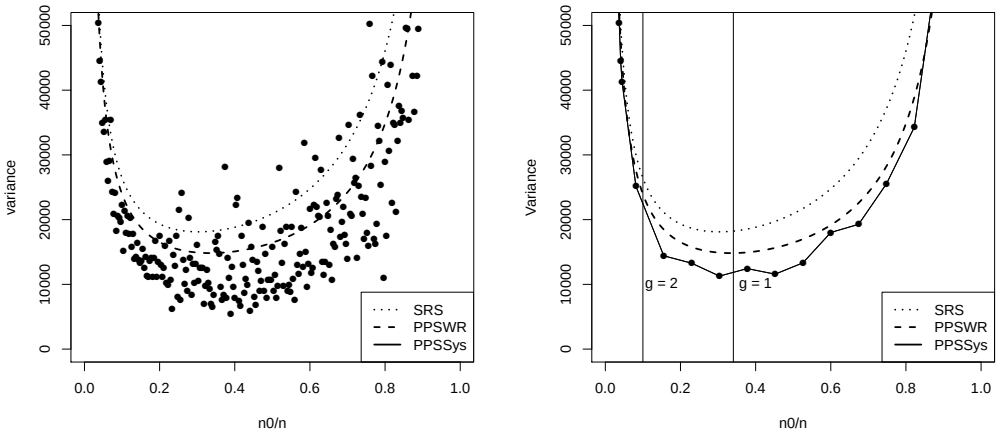

Figure 8: Left: The variance of stratified SRS (dotted), stratified PPS WR (stippled) and systematic PPS with a balanced ordering of the auxiliary variable (open circles) as a function of n_0/n. Right: The same plot, except that the variance for the proportionator is obtained from a partition of the original ones into bins of size 20, where data begins to fluctuate. The sample allocations obtained by (3.11) with $g = 1$ and $g = 2$ are marked with vertical lines.

6 Conclusion

In microscopy, vanishing auxiliary variables are dealt with by adding a small positive constant. In this paper it has been shown, both theoretically and by simulations, that it is of great importance to choose this constant wisely, in order to obtain an efficient predictor. To solve the problem of choosing such constant in an optimal manner, a model-assisted approach has been suggested, where the mean variance is minimized. The optimization depends on the choice of just a few parameters. Investigations based on numerical calculation as well as simulations suggest that the optimum is robust against departures of proportionality in the regression model and misspecification of the parameter q, determining the part of the population total stemming from sampling units with vanishing auxiliary variables. Numerical calculation also showed that the parameter g, controlling the variance of the variables of interest, given the auxiliary variable, must be chosen with care. Under the assumed Poisson model with $g = 1$, choosing $g = 2$ caused a substantial loss in efficiency, whereas the opposite case was less pronounced.

To see how well the approach works in practice, data from microscopy was investigated. Proportionality was not fulfilled, but the simulations suggested that the lack of proportionality was not critical for the optimum derived for the proportional regression model with $g = 1$. Compared to using the optimum for $g = 2$, the variance was almost halved.

7 Acknowledgement

This research was supported by Centre for Stochastic Geometry and Advanced Bioimaging, funded by the Villum Foundation. The authors are grateful for the expertise from Jens R. Nyengaard concerning problems of practical applications and to Kresten K. Keller and co-authors of Keller et al. (2013) for providing data.

References

Bellhouse, D. R. and Rao, J. (1975). Systematic sampling in the presence of a trend. *Biometrika*, 62(3):694–697.

Gardi, J. E., Nyengaard, J. R., and Gundersen, H. J. G. (2008a). Automatic sampling for unbiased and efficient stereological estimation using the proportionator in biological studies. *Journal of Microscopy*, 230(1):108–120.

Gardi, J. E., Nyengaard, J. R., and Gundersen, H. J. G. (2008b). The proportionator: unbiased stereological estimation using biased automatic image analysis and non-uniform probability proportional to size sampling. *Computers in Biology and Medicine*, 38(3):313–328.

Gundersen, H. J. G. (2002). The smooth fractionator. *Journal of Microscopy*, 207(3):191–210.

Hansen, L. V., Kiderlen, M., and Jensen, E. B. V. (2011). Image-based empirical importance sampling: An efficient way of estimating intensities. *Scandinavian Journal of Statistics*, 38(3):393–408.

Hansen, M. H. and Hurwitz, W. N. (1943). On the theory of sampling from finite populations. *The Annals of Mathematical Statistics*, 14(4):333–362.

Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47(260):663–685.

Iachan, R. (1982). Systematic sampling: a critical review. *International Statistical Review*, 50(3):293–303.

Keller, K. K., Andersen, I. T., Andersen, J. B., Hahn, U., Steengaard-Pedersen, K., Hauge, E.-M., and Nyengaard, J. R. (2013). Improving efficiency in stereology: a study applying the proportionator and the autodisector on virtual slides. *Journal of Microscopy*, 251(1):68–76.

Madow, W. G. (1949). On the theory of systematic sampling, II. *The Annals of Mathematical Statistics*, 20(3):333–354.

Murthy, M. N. et al. (1967). *Sampling Theory and Methods*. Calcutta-35: Statistical Publishing Society, 204/1, Barrackpore Trunk Road, India.

Neyman, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4):558–625.

Rao, T. J. (1968). On the allocation of sample size in stratified sampling. *Annals of the Institute of Statistical Mathematics*, 20(1):159–166.

Särndal, C. E., Swensson, B., and Wretman, J. H. (2003). *Model assisted survey sampling.* Springer Verlag.

Stuart, A. (1954). A simple presentation of optimum sampling results. *Journal of the Royal Statistical Society, Series B (Methodological)*, 239–241.

Appendix

Proof of Theorem 1. Due to the independence assumptions, we have

$$\begin{aligned}\operatorname{Var}[\widehat{T}^{st} \mid Y] &= \operatorname{Var}[\widehat{T}_0^{st} \mid Y] + \operatorname{Var}[\widehat{T}_1^{st} \mid Y] \\ &= \operatorname{Var}[\widehat{T}_0 \mid Y] + \operatorname{Var}[\widehat{T}_1 \mid Y].\end{aligned}$$

However, whilst $\widehat{T}_0^{st}$ and $\widehat{T}_1^{st}$ are independent, their counterparts $\widehat{T}_0$ and $\widehat{T}_1$ are not, as they are based on the same sample S. Therefore,

$$\mathbb{E}[\operatorname{Var}[\widehat{T} \mid Y]] = \mathbb{E}[\operatorname{Var}[\widehat{T}^{st} \mid Y]] + 2\,\mathbb{E}[\operatorname{Cov}[\widehat{T}_0, \widehat{T}_1 \mid Y]].$$

To find $\mathbb{E}[\operatorname{Cov}[\widehat{T}_0, \widehat{T}_1 \mid Y]]$, write

$$\begin{aligned}\mathbb{E}[\operatorname{Cov}[\widehat{T}_0, \widehat{T}_1 \mid Y]] &= \operatorname{Cov}[\widehat{T}_0, \widehat{T}_1] - \operatorname{Cov}[\mathbb{E}[\widehat{T}_0 \mid Y], \mathbb{E}[\widehat{T}_1 \mid Y]] \\ &= \operatorname{Cov}[\widehat{T}_0, \widehat{T}_1],\end{aligned}$$

by unbiasedness of the parts $\widehat{T}_h$, $h = 0, 1$. Due to the fact that Y_i and Y_j are uncorrelated for $i \neq j$, $\widehat{T}_0$ and $\widehat{T}_1$ are also uncorrelated, given the sample S. Thus,

$$\begin{aligned}\operatorname{Cov}[\widehat{T}_0, \widehat{T}_1] &= \mathbb{E}[\operatorname{Cov}[\widehat{T}_0, \widehat{T}_1 \mid S]] + \operatorname{Cov}[\mathbb{E}[\widehat{T}_0 \mid Y], \mathbb{E}[\widehat{T}_1 \mid S]] \\ &= \operatorname{Cov}[\mathbb{E}[\widehat{T}_0 \mid Y], \mathbb{E}[\widehat{T}_1 \mid S]] \\ &= \mathbb{E}\left[\mathbb{E}[\widehat{T}_0 \mid S] - \mathbb{E}\,\widehat{T}_0\right]\left[\mathbb{E}[\widehat{T}_1 \mid S] - \mathbb{E}\,\widehat{T}_1\right].\end{aligned}$$

Since the Y_is in Stratum 0 all have the same mean,

$$\mathbb{E}[\widehat{T}_0 \mid S] = \sum_{i \in S \cap U_0} \frac{\mathbb{E}\,Y_i}{\pi_i} = \#(S \cap U_0)\frac{\mathbb{E}\,Y_1}{\pi_1}$$

depends only on the cardinality $\#(S \cap U_0)$. With probability ν, we have that $\#(S \cap U_0) = \lfloor n_0 \rfloor + 1$, and with probability $1 - \nu$, $\#(S \cap U_0) = \lfloor n_0 \rfloor$. Writing

$$A := \{S \mid \#(S \cap U_0) = \lfloor n_0 \rfloor + 1\},$$

and using the fact that $\mathbb{E}\,Y_1/\pi_1 = \mathbb{E}\,T_0/n_0 = \bar{\tau}_0$, we get

$$\mathbb{E}[\widehat{T}_0 \mid S] - \mathbb{E}\,\widehat{T}_0 = \#(S \cap U_0)\bar{\tau}_0 - \mathbb{E}\,T_0 = \begin{cases}(1-\nu)\bar{\tau}_0, & S \in A, \\ -\nu\bar{\tau}_0, & S \notin A.\end{cases}$$

Thus,

$$\begin{aligned}&\big[\mathbb{E}[\widehat{T}_0 \mid S] - \mathbb{E}\,\widehat{T}_0\big]\big[\mathbb{E}[\widehat{T}_1 \mid S] - \mathbb{E}\,\widehat{T}_0\big] \\ &\qquad = \begin{cases} (1-\nu)\bar{\tau}_0\big[\mathbb{E}[\widehat{T}_1 \mid S] - \mathbb{E}\,T_1\big], & S \in A, \\ -\nu\bar{\tau}_0\big[\mathbb{E}[\widehat{T}_1 \mid S] - \mathbb{E}\,T_1\big], & S \notin A. \end{cases}\end{aligned} \tag{7.1}$$

Let $f_1(S) = \mathbb{E}[\widehat{T}_1 \mid S] - \mathbb{E}\,T_1$. Taking the expectation of (7.1), we obtain for $\nu > 0$

$$\begin{aligned}\operatorname{Cov}[\widehat{T}_0, \widehat{T}_1] &= (1-\nu)\bar{\tau}_0\,\mathbb{E}[f_1(S) \mid S \in A]P(S \in A) - \nu\bar{\tau}_0\,\mathbb{E}[f_1(S) \mid S \notin A]P(S \notin A) \\ &= [-(1-\nu)\bar{\tau}_0 - \nu\bar{\tau}_0]\,\mathbb{E}[f_1(S) \mid S \notin A]P(S \notin A),\end{aligned}$$

where we in the second equality have used that

$$\begin{aligned}0 &= \mathbb{E}[f_1(S)] \\ &= \mathbb{E}[f_1(S) \mid S \in A]P(S \in A) + \mathbb{E}[f_1(S) \mid S \notin A]P(S \notin A).\end{aligned}$$ □