MODEL OPTIMUM ALLOCATION OF SAMPLE SIZE TO STRATA IN PROBABILITY PROPORTIONAL TO SIZE SAMPLING WITHOUT REPLACEMENT

Survey & Health Policy Research Center
Technical Report

March 2024

MODEL OPTIMUM ALLOCATION OF SAMPLE SIZE TO STRATA IN PROBABILITY PROPORTIONAL TO SIZE SAMPLING WITHOUT REPLACEMENT

SUNWOONG KIM[1]*, STEVEN G. HEERINGA[2] AND SOHYUNG PARK[1]

Dongguk University and University of Michigan

Summary

It is well-known that probability proportional to size (PPS) sampling methods without replacement, simply called πPS sampling, frequently provide more efficient sample estimates than simple random sampling or PPS sampling with replacement. We investigate methods of allocating sample size to strata using super-population regression models that may be beneficial to πPS sampling methods. This study focuses on Sampford's method, which is one of the more popular πPS sampling methods among practitioners. We present the true optimal allocation for his method under the assumption that the values of the characteristic under study are known. Based on general super-population regression models with the intercept term, overlooked in the previous studies, we derive new alternatives to the true optimal allocation that may be easily solved by convex mathematical programming algorithms. We illustrate this model allocation for finite populations generated from a hypothetical population.

Key words: convex mathematical programming algorithms; PPS sampling; Sampford's method; sample allocation; stratified sampling; super-population regression model

---

*Author to whom correspondence should be addressed.
[1]Survey and Health Policy Research Center and Department of Statistics, Dongguk University, 30 Pildong-ro, 1-gil, Jung-gu, Seoul, South Korea 04620
e-mail: sunwk@dongguk.edu
[2]Survey Research Center, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, MI 48106, USA.

1. Introduction

In stratified sampling the finite population of N units is divided into h strata of sizes N_h, $h = 1, 2, \cdots, H$, and a sample of a chosen size n_h is selected within each stratum. The selections are made independently in distinct strata. Let y_{hi} be the value of the characteristic $\boldsymbol{Y}$ under study for unit i in stratum h. One of the important roles of the survey sampler is to determine the values of n_h in the respective strata, that is, sample allocation, which will result in the greatest precision for sample estimates of true parameter such as the population total $Y = \sum_{h=1}^{H} \sum_{i=1}^{N_h} y_{hi}$ or the population mean $\overline{Y} = Y / N$.

Under stratified simple random sampling (SSRS) without replacement the following sample allocations are appeared in the introductory texts: (i) proportional allocation, suggested by Bowley (1926), and (ii) Neyman (1934) allocation. Proportional allocation simply assigns n_h in proportion to N_h, while Neyman allocation is given by the formula

$$n_h = nN_h S_{yh} / \sum_{h=1}^{H} N_h S_{yh} , \tag{1}$$

where n is the total sample size,

$$S_{yh}^2 = \sum_{i=1}^{N_h} (y_{hi} - \overline{Y}_h)^2 / (N_h - 1) , \tag{2}$$

and

$$\overline{Y}_h = \sum_{i=1}^{N_h} y_{hi} / N_h . \tag{3}$$

For SSRS with replacement S_{yh} in (1) is replaced by σ_{yh} , where $\sigma_{yh}^2 = \sum_{i=1}^{N_h} (y_{hi} - \overline{Y}_h)^2 / N_h$. It is noted that Neyman allocation (1) is a *true optimal allocation* for minimizing the variance of sample estimates of Y or $\overline{Y}$. But this true optimal allocation is not available in practice, since the values of S_{yh}^2 (or those of y_{hi}) are often unknown.

In such cases what we call x-optimal allocation is an alternative to Neyman allocation. This method uses the values x_{hi} of $\boldsymbol{X}$, a positive auxiliary characteristic assumed to be highly correlated with the characteristic $\boldsymbol{Y}$ under study. It substitutes S_{xh} for S_{yh} in (1), namely,

$$n_h = nN_h S_{xh} / \sum_{h=1}^{H} N_h S_{xh} \tag{4}$$

where S_{xh}^2 and $\overline{X}_h$ are calculated by the values of x_{hi} instead of those of y_{hi} in (2) and (3), respectively.

However, if the correlation between $\boldsymbol{X}$ and $\boldsymbol{Y}$ is not almost perfect, this allocation is not ‘optimal’, and furthermore, if there are substantial differences between S_{xh} and S_{yh} , it might result in lower precision for sample estimates compared to proportional allocation. Thus, the substitution of S_{xh} for S_{yh} without any valid justification should be avoided.

As an alternative, model-assisted methods with practical advantages over x-optimal allocation have been studied. Hanurav (1965), Rao (1968), Reddy (1976), Rao (1977), Dayal (1985), and Gupt (2003) showed that a super-population regression model with respect to $\boldsymbol{X}$ and $\boldsymbol{Y}$ can be appropriately used for the sample allocation in SSRS. This technique using a model is what we simply call *model allocation*, which may be applied to other sampling methods with efficiency better than simple random sampling (SRS).

It is well-known that under many situations sampling strategies with varying probabilities such as probability proportional to size (PPS) sampling with replacement or without replacement provide more efficient sample estimates than SRS.

There are a few studies on model allocation in stratified PPS sampling with replacement. For example, see Rao (1977) and Gupt & Rao (1997).

PPS sampling without replacement, simply called πPS sampling, is often more efficient than PPS sampling with replacement, as described in Rao & Bayless (1969) and Bayless & Rao (1970). But there are very few studies on model allocation in stratified πPS sampling. Rao's (1968) study, followed by Rao (1977), remains valuable. He suggested a model allocation approach using a super-population regression model without the intercept. The primary objective of his approach is to minimize the expected variance of the Horvitz & Thompson (H-T) (1952) estimator under the model. An interesting result is that his approach always gives the same sample allocation for all πPS sampling methods, as shown in Section 2.

However, his result raises a question: It may be desirable to introduce an intercept term into the super-population regression model. If the intercept is included in the model, is sample allocation still the same for any πPS sampling?

Though it is proved in Section 3, the presence of the intercept in the model leads to sample allocation problems that differ according to the chosen πPS sampling method. Thus one would like to pay attention to a specific allocation strategy appropriate for a given πPS sampling method, in particular, methods that are popular with samplers.

In fact, a host of πPS sampling methods have been developed to select samples of size equal to or greater than two. See Brewer & Hanif (1983). Most methods for the sample size greater than two are not easily applied in practice. Some of them may construct a good design for reducing the variance of sample estimates compared to alternative methods and achieve unbiased variance estimation. Among suggested methods, Sampford's (1967) method, which is the extension of Brewer's (1963) method and was discussed by Rao & Bayless (1969), Bayless & Rao (1970), Cochran (1977), Särndal (1996), Smith (2001), Rao (2005), Tillé (2006), Bondesson *et al*. (2006), and Haziza *et al*. (2008), is the better known to the samplers. His method is also called the Rao-Sampford method, since Rao (1965) developed the same procedure. Gabler (1981) proved that Sampford's method is always more efficient than PPS sampling with replacement. His method has not been widely used in the past due to its computational complexity, but it can be easily implemented with modern computing power. For example, it is available in the recent version of SAS or SPSS or R package "sampfling" (http://cran.r-project.org/).

Accordingly, we may add a further question: If we use Sampford's (1967) πPS sampling method, what sample allocation strategy under the super-population regression model with the intercept would be followed?

In this paper, we attempt to answer the above two questions. We first begin by revisiting Rao's (1968) method in Section 2. In section 3, we show that under Sampford's sampling method the introduction of the intercept term into the model results in allocation problems looking complicated, but those that can be easily solved by optimization approaches. In section 4, we illustrate this model allocation for the finite population generated from a hypothetical population.

2. Rao's method

Let s be a sample of size n_h drawn from each stratum and let $P(\cdot)$ denote a sampling design such that $P(s)$ gives the probability of selecting s under the given sampling method. Let S be the set of all possible samples from each stratum. The total sample size n is:

$$n = \sum_{h=1}^{H} n_h \ . \qquad (5)$$

Then the probability that the unit i in the stratum h will be in a sample, denoted π_{hi}, is given by

$$\pi_{hi} = \sum_{i \in s,\ s \in S} P(s)\,,\ i = 1, 2, \cdots, N_h\,,\ h = 1, 2, \cdots, H\ , \qquad (6)$$

which are called the first-order inclusion probabilities.

Also, the probability that both of the units i and j in the stratum h will be included in a sample, denoted π_{hij}, is obtained by

$$\pi_{hij} = \sum_{i,j \in s,\ s \in S} P(s)\ ,\ i \neq j = 1, 2, \cdots, N_h\ ,\ h = 1, 2, \cdots, H\ . \tag{7}$$

These are termed the joint probabilities or the second-order inclusion probabilities.

As an estimator of the population total Y, consider the H-T estimator

$$\hat{Y}_{HT} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} \frac{y_{hi}}{\pi_{hi}}\ , \tag{8}$$

where $\pi_{hi} = n_h p_{hi}$, $p_{hi} = x_{hi} / X_h$, $X_h = \sum_{i=1}^{N_h} x_{hi}$, and $0 < \pi_{hi} < 1$.

This estimator is an unbiased estimator of Y, with variance:

$$Var\left(\hat{Y}_{HT}\right) = \sum_{h=1}^{H} \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left(\pi_{hi}\pi_{hj} - \pi_{hij}\right) \left(\frac{y_{hi}}{\pi_{hi}} - \frac{y_{hj}}{\pi_{hj}}\right)^2 . \tag{9}$$

Rao (1968) considered the following super-population regression model without the intercept:

$$y_{hi} = \beta x_{hi} + \varepsilon_{hi}\ , \tag{10}$$

where $E_\xi\left(y_{hi} \middle| x_{hi}\right) = \beta x_{hi}$, $V_\xi\left(y_{hi} \middle| x_{hi}\right) = \sigma^2 x_{hi}^g$ and $Cov_\xi\left(y_{hi}, y_{hj} \middle| x_{hi}, x_{hj}\right) = 0$. Here E_ξ, V_ξ and Cov_ξ denote the model expectation, variance and covariance given x_{hi}'s respectively over all the finite populations that can be drawn from the super-population. β, σ^2 and g are super-population parameters with $\sigma^2 > 0$ and $1 \le g \le 2$.

Then we have the following expected variance under the model (10):

$$E_{\xi} Var\left(\hat{Y}_{HT}\right) = \sum_{h=1}^{H} \sum_{i=1}^{N_h} \left(\frac{1}{\pi_{hi}} - 1 \right) \sigma^2 x_{hi}^g . \tag{11}$$

To minimize (11) subject to the condition (5), using the Lagrange multiplier λ, consider

$$\sum_{h=1}^{H} \sum_{i=1}^{N_h} \left(\frac{1}{n_h p_{hi}} - 1 \right) \sigma^2 x_{hi}^g + \lambda \left(\sum_{h=1}^{H} n_h - n \right). \tag{12}$$

Equating (12) to zero and differentiating with respect to n_h, we have

$$n_h = \frac{1}{\sqrt{\lambda}} \sqrt{\sum_{i=1}^{N_h} \frac{\sigma^2 x_{hi}^g}{p_{hi}}} . \tag{13}$$

Substituting n_h in (5), we have

$$\frac{1}{\sqrt{\lambda}} = n \Big/ \sum_{h=1}^{H} \sqrt{\sum_{i=1}^{N_h} \frac{\sigma^2 x_{hi}^g}{p_{hi}}} . \tag{14}$$

Replacing $1/\sqrt{\lambda}$ in (13) with (14), eventually we have the following model allocation in each stratum:

$$n_h = n \frac{\sqrt{X_h \sum_{i=1}^{N_h} x_{hi}^{g-1}}}{\sum_{h=1}^{H} \sqrt{X_h \sum_{i=1}^{N_h} x_{hi}^{g-1}}} . \tag{15}$$

With the assumption $V_{\xi}\left(y_{hi} \middle| x_{hi}\right) = v(x_{hi})$, where $v(\cdot)$ is a given function, Rao (1977) obtained a form different from (15). Note that if $g = 2$, (15) reduces to:

$$n_h = n \frac{X_h}{\sum_{h=1}^{H} X_h} , \tag{16}$$

which is called x-proportional allocation to the stratum.

Looking at the expected variance in (11) and the model allocation in (15), it does not involve the joint probabilities π_{hij} in each stratum. It indicates that under the model without the intercept (10) the specific sampling design properties of a given πPS sampling method that determine the π_{hij} are not reflected in the sample allocation, resulting in the same sample allocation for any πPS sampling. Hence the following issues, as mentioned in the Introduction, are of interest:

(a) The super-population regression model which one may wish to employ in many surveys will be:

$$y_{hi} = \alpha + \beta x_{hi} + \varepsilon_{hi}\,, \tag{17}$$

which is a general form and (10) is a special form of (17) when $\alpha = 0$. Considering the intercept term α, we need to reexamine the most appropriate sample allocation strategy for stratified πPS sampling.

(b) It will be shown in the following section that using (17) gives a sample allocation involving the joint probabilities π_{hij} depending on the chosen πPS sampling. If we focus on Sampford's (1967) method for πPS sampling, what sample allocation strategy would be appropriate?

Section 3 will address these issues of sample allocation.

3. Alternative sample allocations under stratified πPS sampling

As mentioned above, Neyman allocation (1) is the *true optimal allocation* for SSRS. Since Rao (1968) and Rao (1977) did not deal with what the *true optimal allocation* is for stratified πPS sampling, we describe it first.

3.1. True optimal allocation

Assume that the values of y_{hi} are known. Instead of (9) we consider the following form of the variance of the H-T estimator

$$Var\left(\hat{Y}_{HT}\right)=\sum_{h=1}^{H}\sum_{i=1}^{N_h}\frac{y_{hi}^2}{\pi_{hi}}-\sum_{h=1}^{H}\sum_{i=1}^{N_h}y_{hi}^2+2\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\frac{\pi_{hij}}{\pi_{hi}\pi_{hj}}y_{hi}y_{hj}-2\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}y_{hi}y_{hj} \qquad (18)$$

Since the second and fourth terms in (18) are independent of n_h , the minimization of the variance of the H-T estimator in terms of n_h reduces to the minimization of

$$\sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{N_h}\frac{y_{hi}^2}{p_{hi}}+2\sum_{h=1}^{H}\frac{1}{n_h^2}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\frac{\pi_{hij}}{p_{hi}p_{hj}}y_{hi}y_{hj} \ . \qquad (19)$$

Unfortunately, with the Lagrange's multiplier method or other simple methods, one cannot derive an allocation formula with respect to n_h for minimizing (19) under the condition (5), due to n_h^{-2} of the second term in (19). Moreover, the joint probabilities π_{hij} of the second term in (19) should be evaluated according to the chosen πPS sampling method.

We focus on Sampford's (1967) method. Under this method n_h units are selected with replacement in each stratum. The first unit in stratum h is selected with probability p_{hi} and all subsequent units with probabilities

$\lambda_{hi} / \sum_{i=1}^{N_h} \lambda_{hi}$, where $\lambda_{hi} = p_{hi} / (1 - n_h p_{hi})$. Any sample that does not contain n_h distinct units is rejected. It is noted that $\pi_{hi} = n_h p_{hi}$ for his method.

Because the exact calculation of all π_{hij} for his method is complicated and computationally prohibitive, the following approximate expression correct to $O(N_h^{-4})$ under the assumptions that (i) n_h is small relative to N_h and (ii) p_{hi} is of $O(N_h^{-1})$ is useful:

$$\begin{aligned}\pi_{hij} = n_h(n_h - 1) p_{hi} p_{hj} [1 + \{(p_{hi} + p_{hj}) - \sum_{k=1}^{N_h} p_{hk}^2\} \\ + \{2(p_{hi}^2 + p_{hj}^2) - 2\sum_{k=1}^{N_h} p_{hk}^3 - (n_h - 2) p_{hi} p_{hj} \\ + (n_h - 3)(p_{hi} + p_{hj}) \sum_{k=1}^{N_h} p_{hk}^2 - (n_h - 3)(\sum_{k=1}^{N_h} p_{hk}^2)^2\}] .\end{aligned} \tag{20}$$

This approximation was derived by Asok & Sukhatme (1976) based on an asymptotic theory.

When substituting (20) for π_{hij} in (19), we have

$$\sum_{h=1}^{H} \frac{1}{n_h} \sum_{i=1}^{N_h} \frac{y_{hi}^2}{p_{hi}} + 2 \sum_{h=1}^{H} \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left(1 - \frac{1}{n_h}\right) (n_h \pi_{hij1} + \pi_{hij2}) y_{hi} y_{hj} \ , \tag{21}$$

where

$$\pi_{hij1} = (p_{hi} + p_{hj}) \sum_{k=1}^{N_h} p_{hk}^2 - p_{hi} p_{hj} - (\sum_{k=1}^{N_h} p_{hk}^2)^2 \tag{22}$$

and

$$\pi_{hij2} = 1 + \{(p_{hi} + p_{hj}) - \sum_{k=1}^{N_h} p_{hk}^2\} + 2(p_{hi}^2 + p_{hj}^2) - 2\sum_{k=1}^{N_h} p_{hk}^3$$

$$+2p_{hi}p_{hj} - 3(p_{hi} + p_{hj})\sum_{k=1}^{N_h} p_{hk}^2 + 3(\sum_{k=1}^{N_h} p_{hk}^2)^2$$

(23)

From (21), we can derive the form (24) in terms of n_h, which is the objective function of the optimization problem (or nonlinear programming problem):

Minimize

$$\sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{N_h}\frac{y_{hi}^2}{p_{hi}} + 2\sum_{h=1}^{H} n_h \sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\pi_{hij1} y_{hi} y_{hj} - 2\sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\pi_{hij2} y_{hi} y_{hj} \qquad (24)$$

subject to $\sum_{h=1}^{H} n_h = n$ in (5).

The solution n_h, $h = 1, 2, \cdots, H$, to this optimization problem will be the *true optimal allocation* for minimizing the variance of the H-T estimator under Sampford's method for stratified sampling.

3.2. Model allocations

Because the values of y_{hi} are often unknown, the optimization problem defined above may not be applied in practice. Instead we assume two different super-population regression models involving an intercept term:

Model I:

$$y_{hi} = \alpha + \beta x_{hi} + \varepsilon_{hi}, \quad h = 1, 2, \cdots, H, \ i = 1, \cdots, N_h, \qquad (25)$$

where ε_{hi} is numerically negligible, that is, x perfectly explains y.

Model II:

$$y_{hi} = \alpha + \beta x_{hi} + \varepsilon_{hi}\,, \quad h = 1, 2, \cdots, H\ ,\ i = 1, 2, \cdots, N_h \tag{26}$$

where $E_{\xi}\left(y_{hi}\left|x_{hi}\right.\right) = \alpha + \beta x_{hi}$, $V_{\xi}\left(y_{hi}\left|x_{hi}\right.\right) = \sigma^2 x_{hi}^{g}$, and $Cov_{\xi}\left(y_{hi}, y_{hj}\left|x_{hi}, x_{hj}\right.\right) = 0$.

Model I was used by Des Raj (1956). The model (10) is the special case of the *Model II*, where $\alpha = 0$. *Model II* was assumed by Reddy (1976), Rao (1977) and Dayal (1985) for the sample allocation under SSRS. Assume that the values of x_{hi} and the super-population parameters are known for the two models.

Theorem 1. *Under the Model I, the sample allocation problem for the minimization of the expected variance of the H-T estimator under any πPS sampling is equivalent to minimizing*

$$\sum_{h=1}^{H} \frac{A_h}{n_h^2} + \sum_{h=1}^{H} \frac{B_h}{n_h}\ , \tag{27}$$

where

$$A_h = 2X_h^2 \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \frac{\alpha^2 + \alpha\beta(x_{hi} + x_{hj})}{x_{hi}x_{hj}} \pi_{hij} \tag{28}$$

and

$$B_h = X_h \left(\sum_{i=1}^{N_h} \frac{(\alpha + \beta x_{hi})^2}{x_{hi}} - \beta^2 X_h \right). \tag{29}$$

Proof. Consider the four terms in the right-hand side of expression (18) for the variance of the H-T estimator. The expected variance $E_{\xi}Var\left(\hat{Y}_{HT}\right)$ under the *Model I* is the sum of the expected values for those four terms. The expected values for the second and fourth terms in (18) are known values that do not involve n_h, and those for the other terms in (18) do depend on n_h and are given by:

$$\sum_{h=1}^{H}\frac{X_h}{n_h}\sum_{i=1}^{N_h}\frac{(\alpha+\beta x_{hi})^2}{x_{hi}}+\left[2\sum_{h=1}^{H}\frac{X_h^2}{n_h^2}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\frac{\alpha^2+\alpha\beta(x_{hi}+x_{hj})}{x_{hi}x_{hj}}\pi_{hij}+\beta^2\sum_{h=1}^{H}X_h^2-\beta^2\sum_{h=1}^{H}\frac{X_h^2}{n_h}\right], \tag{30}$$

which is derived from the fact that $\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\pi_{hij}=n_h(n_h-1)/2$.

Since $\beta^2\sum_{h=1}^{H}X_h^2$ is also known, the quantity to be minimized in (30) is:

$$\left[\sum_{h=1}^{H}\frac{X_h}{n_h}\sum_{i=1}^{N_h}\frac{(\alpha+\beta x_{hi})^2}{x_{hi}}\right]+\left[2\sum_{h=1}^{H}\frac{X_h^2}{n_h^2}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\frac{\alpha^2+\alpha\beta(x_{hi}+x_{hj})}{x_{hi}x_{hj}}\pi_{hij}-\beta^2\sum_{h=1}^{H}\frac{X_h^2}{n_h}\right] \tag{31}$$

The proof follows from substitution of (28) and (29) in (31).

Remark 1. The minimization of the expected variance in terms of n_h under the *Model I* with the intercept term reduces to minimization of the function (27), which involves the joint probabilities π_{hij} in each stratum that in turn depend on the chosen πPS sampling method.

Remark 2. The minimization of (27) in terms of n_h under the condition (5) is a simple optimization problem because the A_h in (28) and the B_h in (29) are known values.

From (20) and (27) we obtain the following theorem.

Theorem 2. *Under the Model I, the sample allocation problem to minimize the expected variance of the H-T estimator under Sampford's method when using the joint probabilities, correct to* $O(N_h^{-4})$ *given in (20) is equivalent to minimizing*

$$\sum_{h=1}^{H} C_h n_h + \sum_{h=1}^{H} \frac{D_h}{n_h}, \tag{32}$$

where

$$C_h = 2\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2 + \alpha\beta(x_{hi} + x_{hj})\right\}\pi_{hij1} \tag{33}$$

and

$$D_h = B_h - 2\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2 + \alpha\beta(x_{hi} + x_{hj})\right\}\pi_{hij2}. \tag{34}$$

Proof. Substituting π_{hij} from (20) in (28) for the first term of (27), we have

$$\sum_{h=1}^{H}\frac{A_h}{n_h^2} = 2\sum_{h=1}^{H}\left(1 - \frac{1}{n_h}\right)\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2 + \alpha\beta(x_{hi} + x_{hj})\right\}(n_h\pi_{hij1} + \pi_{hij2})$$

$$= 2\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h} n_h\left\{\alpha^2 + \alpha\beta(x_{hi} + x_{hj})\right\}\pi_{hij1}$$

$$+2\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2 + \alpha\beta(x_{hi} + x_{hj})\right\}\pi_{hij2}$$

$$-2\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2+\alpha\beta(x_{hi}+x_{hj})\right\}\pi_{hij1}$$

$$-2\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\frac{1}{n_h}\left\{\alpha^2+\alpha\beta(x_{hi}+x_{hj})\right\}\pi_{hij2}\ . \tag{35}$$

Since the second and third terms in (35) are the known values, the minimization of (35) reduces to minimizing:

$$2\sum_{h=1}^{H}n_h\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2+\alpha\beta(x_{hi}+x_{hj})\right\}\pi_{hij1}-2\sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2+\alpha\beta(x_{hi}+x_{hj})\right\}\pi_{hij2}\ . \tag{36}$$

Adding $\sum_{h=1}^{H}\frac{B_h}{n_h}$ in (27) to (36), we have the following minimization problem corresponding to the minimization of (27):

$$2\sum_{h=1}^{H}n_h\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2+\alpha\beta(x_{hi}+x_{hj})\right\}\pi_{hij1}$$

$$+\sum_{h=1}^{H}\frac{1}{n_h}\left[B_h-2\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\alpha^2+\alpha\beta(x_{hi}+x_{hj})\right\}\pi_{hij2}\right]. \tag{37}$$

This completes the proof.

Remark 3. The minimization of (32) under (5) is a simple allocation problem in terms of n_h because the C_h in (33) and the D_h in (34) are the known values.

Remark 4. We can define the following optimization problem with respect to n_h:

Minimize

$$\sum_{h=1}^{H}C_h n_h+\sum_{h=1}^{H}\frac{D_h}{n_h} \tag{38}$$

subject to $\sum_{h=1}^{H} n_h = n$ in (5).

In addition to (5), the following conditions can be added:

$$n_h \le N_h \,,\ \ h = 1, 2, \cdots, H \tag{39}$$

and

$$n_h \ge 2 \,,\ \ h = 1, 2, \cdots, H \ . \tag{40}$$

Also, other possible conditions will be:

$$n_h p_{hi} < 1 \,,\ \ i = 1, 2, \cdots, N_h \,,\ \ h = 1, 2, \cdots, H \ . \tag{41}$$

The optimization problem in Remark 4 may be easily handled by convex mathematical programming algorithms and the solution to the problem would provide an efficient sample allocation strategy under the *Model I* when using Sampford's sampling procedure.

We obtain the following theorems regarding the minimization of the variance of the H-T estimator in πPS sampling under the assumption of *Model II*, which is more practical than *Model I*.

Theorem 3. *Under Model II, the sample allocation problem for minimizing the expected variance of the H-T estimator under any* πPS *sampling amounts to minimizing*:

$$\sum_{h=1}^{H} \frac{A_h^*}{n_h^2} + \sum_{h=1}^{H} \frac{B_h^*}{n_h} \ , \tag{42}$$

where

$$A_h^* = 2\alpha X_h^2 \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left(x_{hj}^{-1} - x_{hi}^{-1} \right) \left(\alpha x_{hi}^{-1} + \beta \right) \pi_{hij} \tag{43}$$

and

$$B_h^* = \sigma^2 X_h \sum_{i=1}^{N_h} x_{hi}^{g-1} \ . \tag{44}$$

Proof. Consider a different form of (9) using $\pi_{hi} = n_h p_{hi}$:

$$Var\left(\hat{Y}_{HT}\right) = \sum_{h=1}^{H} \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left(p_{hi} p_{hj} - \frac{\pi_{hij}}{n_h^2} \right) \left(\frac{y_{hi}}{p_{hi}} - \frac{y_{hj}}{p_{hj}} \right)^2 . \tag{45}$$

By using

$$E_\xi y_{hi}^2 = \sigma^2 x_{hi}^g + \alpha^2 + \beta^2 x_{hi}^2 + 2\alpha\beta x_{hi} \tag{46}$$

and

$$E_\xi (y_{hi} y_{hj}) = \alpha^2 + \alpha\beta(x_{hi} + x_{hj}) + \beta^2 x_{hi} x_{hj} , \tag{47}$$

we obtain

$$E_\xi \left(\frac{y_{hi}}{p_{hi}} - \frac{y_{hj}}{p_{hj}} \right)^2 = 2\sigma^2 X_h^g p_{hi}^{g-2} + 2\alpha X_h^2 \frac{x_{hj} - x_{hi}}{x_{hi} x_{hj}} \left(\alpha x_{hi}^{-1} + \beta \right) . \tag{48}$$

Then we get:

$$\begin{aligned} E_\xi Var\left(\hat{Y}_{HT}\right) &= \Delta_\xi + 2\alpha \sum_{h=1}^{H} X_h^2 \left(\sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left(p_{hi} p_{hj} - \frac{\pi_{hij}}{n_h^2} \right) \frac{x_{hj} - x_{hi}}{x_{hi} x_{hj}} \left(\alpha x_{hi}^{-1} + \beta \right) \right) \\ &= \Delta_\xi + 2\alpha \sum_{h=1}^{H} \left(\sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left(x_{hj} - x_{hi} \right) \left(\alpha x_{hi}^{-1} + \beta \right) \right) \end{aligned} \tag{49}$$

$$+2\alpha\sum_{h=1}^{H}\frac{X_h^2}{n_h^2}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left(x_{hj}^{-1}-x_{hi}^{-1}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij},$$

where

$$\Delta_\xi = 2\sigma^2\sum_{h=1}^{H}X_h^g\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}p_{hi}^{g-2}\left(p_{hi}p_{hj}-\frac{\pi_{hij}}{n_h^2}\right)$$

$$=\sigma^2\sum_{h=1}^{H}\sum_{i=1}^{N_h}\frac{X_h^g}{n_h}(1-n_h p_{hi})p_{hi}^{g-1}$$

$$=\sigma^2\sum_{h=1}^{H}\sum_{i=1}^{N_h}\left(\frac{1}{n_h p_{hi}}-1\right)x_{hi}^g$$

$$=\sigma^2\sum_{h=1}^{H}\frac{X_h}{n_h}\sum_{i=1}^{N_h}x_{hi}^{g-1}-\sigma^2\sum_{h=1}^{H}\sum_{i=1}^{N_h}x_{hi}^g. \tag{50}$$

Since the second term in (49) and the second term in (50) are independent of n_h, the minimization of the expected variance of (45) reduces to minimizing:

$$2\alpha\sum_{h=1}^{H}\frac{X_h^2}{n_h^2}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left(x_{hj}^{-1}-x_{hi}^{-1}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij}+\sigma^2\sum_{h=1}^{H}\frac{X_h}{n_h}\sum_{i=1}^{N_h}x_{hi}^{g-1}. \tag{51}$$

Since (51) equals (42), the proof is completed.

Remark 5. Under *Model II* with the intercept term the minimization of the expected variance in terms of n_h amounts to the minimization of the function (42) involving the joint probabilities, which differ according to the chosen πPS sampling method.

Remark 6. Since the A_h^* in (43) and the B_h^* in (44) are the known values, minimizing (42) in terms of n_h subject to the condition (5) is a simple optimization problem.

Remark 7. Δ_ξ in (50) is an alternative form to (11), which is the expected variance of the H-T estimator under the model (10) without the intercept term. Hence the expected variance of the H-T estimator under *Model II* with the intercept term consists of (11) plus the additional terms, as shown in (49).

Corollary 1. *Under the Model II with* $\alpha = 0$*, the sample allocation problem for minimizing the expected variance of the H-T estimator under any* πPS *sampling is equivalent to minimizing:*

$$\sum_{h=1}^{H} \frac{X_h}{n_h} \sum_{i=1}^{N_h} x_{hi}^{g-1} \,. \tag{52}$$

Proof. When $\alpha = 0$, (49) in Theorem 3 reduces to simply Δ_ξ, which is expressed as (50). The second term in (50) does not depend on n_h, and the minimization of (50) reduces to the one of (52). Hence, we have the corollary.

Remark 8. (52) does not depend on the joint probabilities.

Remark 9. It is interesting to note that when solving for n_h by using the Lagrange multiplier λ to minimize (52) subject to the condition (5), it gives

(15), which is the sample allocation under the model (10). This is because the model (10) is *Model II* with $\alpha = 0$.

Theorem 4. *Under Model II, the sample allocation problem for Sampford's sampling method in minimizing the expected variance of the H-T estimator, when using the joint probabilities (20) correct to* $O(N_h^{-4})$, *is equivalent to minimizing:*

$$\sum_{h=1}^{H} C_h^* n_h + \sum_{h=1}^{H} \frac{D_h^*}{n_h}, \tag{53}$$

where

$$C_h^* = 2\alpha \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left\{ \left(x_{hi} - x_{hj} \right) \left(\alpha x_{hi}^{-1} + \beta \right) \pi_{hij1} \right\}, \tag{54}$$

and

$$D_h^* = B_h^* - 2\alpha \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left\{ \left(x_{hi} - x_{hj} \right) \left(\alpha x_{hi}^{-1} + \beta \right) \pi_{hij2} \right\}. \tag{55}$$

Proof. Substituting (20) in the first term of (42) and using (22) and (23), we obtain

$$\begin{aligned}
\sum_{h=1}^{H} \frac{A_h^*}{n_h^2} &= 2\alpha \sum_{h=1}^{H} \frac{X_h^2}{n_h^2} n_h (n_h - 1) \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left\{ \left(x_{hj}^{-1} - x_{hi}^{-1} \right) \left(\alpha x_{hi}^{-1} + \beta \right) p_{hi} p_{hj} \left(n_h \pi_{hij1} + \pi_{hij2} \right) \right\} \\
&= 2\alpha \sum_{h=1}^{H} \left(1 - \frac{1}{n_h} \right) \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left\{ \left(x_{hi} - x_{hj} \right) \left(\alpha x_{hi}^{-1} + \beta \right) \left(n_h \pi_{hij1} + \pi_{hij2} \right) \right\} \\
&= 2\alpha \sum_{h=1}^{H} n_h \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left\{ \left(x_{hi} - x_{hj} \right) \left(\alpha x_{hi}^{-1} + \beta \right) \pi_{hij1} \right\} \\
&\qquad + 2\alpha \sum_{h=1}^{H} \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} \left\{ \left(x_{hi} - x_{hj} \right) \left(\alpha x_{hi}^{-1} + \beta \right) \pi_{hij2} \right\}
\end{aligned} \tag{56}$$

$$-2\alpha\sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\left(x_{hi}-x_{hj}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij1}\right\}$$
$$-2\alpha\sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\left(x_{hi}-x_{hj}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij2}\right\}.$$

Since the second and third terms in (56) are independent of n_h , the minimization of (56) reduces to minimizing the other terms:

$$2\alpha\sum_{h=1}^{H}n_h\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\left(x_{hi}-x_{hj}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij1}\right\}$$
$$-2\alpha\sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\left(x_{hi}-x_{hj}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij2}\right\}. \tag{57}$$

Thus, the minimization of (42) with (43) and (44) amounts to minimizing

$$2\alpha\sum_{h=1}^{H}n_h\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\left(x_{hi}-x_{hj}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij1}\right\}$$
$$-2\alpha\sum_{h=1}^{H}\frac{1}{n_h}\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\left(x_{hi}-x_{hj}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij2}\right\} \tag{58}$$
$$+\sum_{h=1}^{H}\frac{B_h^*}{n_h}.$$

Accordingly, the following reduced form from (58) can be obtained.

$$2\alpha\sum_{h=1}^{H}n_h\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\left(x_{hi}-x_{hj}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij1}\right\}$$
$$+\sum_{h=1}^{H}\frac{1}{n_h}\left[B_h^*-2\alpha\sum_{i=1}^{N_h}\sum_{j>i}^{N_h}\left\{\left(x_{hi}-x_{hj}\right)\left(\alpha x_{hi}^{-1}+\beta\right)\pi_{hij2}\right\}\right] \tag{59}$$

Hence, we have proved the theorem.

Remark 10. Since the C_h^* in (54) and the D_h^* in (55) are the known values, the minimization of (53) subject to (5) is a simple allocation problem in terms of n_h.

Remark 11. In order to find a solution for n_h, we may define the following optimization problem:

Minimize

$$\sum_{h=1}^{H} C_h^* n_h + \sum_{h=1}^{H} \frac{D_h^*}{n_h} \tag{60}$$

subject to $\sum_{h=1}^{H} n_h = n$ in (5).

The solution to this optimization problem easily solved by convex mathematical programming algorithms will provide an optimum sample allocation under *Model II* in using Sampford's method. As discussed in Remark 4, the conditions (39), (40) and (41) can be also used with preferences.

4. Simulations

To examine sample allocation under Sampford's method, we considered the super-population *Model II* given in Hansen *et al*. (1983, p. 781): the positive auxiliary characteristic $\boldsymbol{X}$ has a gamma distribution with density function $0.4x_{hi}\exp(-x_{hi}/5)$ and the characteristic $\boldsymbol{Y}$ under study, conditional on $\boldsymbol{X}$, has a gamma distribution with density function $1/(b^c\Gamma(c))y_{hi}^{c-1}\exp(-y_{hi}/b)$ where $b = 1.25x_{hi}^{3/2}(8+5x_{hi})^{-1}$ and $c = 0.04x_{hi}^{-3/2}(8+5x_{hi})^2$. Accordingly,

$$E_{\xi}\left(y_{hi}\middle|x_{hi}\right)=0.4+0.25x_{hi} \tag{61}$$

and

$$V_{\xi}\left(y_{hi}\middle|x_{hi}\right)=0.0625x_{hi}^{3/2}\,. \tag{62}$$

Finite populations with sizes 30, 60, 90, 120, 150, 180, 210, 240, 270, and 300 were generated from the super-population. The value of the characteristic $\boldsymbol{X}$ is assumed to be known for each unit in each finite population. Each finite population was divided into 3 strata and each stratum has the same size (e.g., 10, 10, and 10 for size 30). We considered two types of stratification: (A) The units in the finite population are arranged in increasing order of the value of $\boldsymbol{X}$ and the first N_1 are considered as stratum 1 and the second N_2 as stratum 2 and the remaining N_3 as stratum 3; (B) The units in the finite population are randomly assigned to each stratum.

Before selection, the units in each stratum were arranged in increasing order of the value of $\boldsymbol{X}$, so that $x_{h1}\le x_{h2}\le\cdots\le x_{hN_h}$. The total sample size n for each population is 10 percent of each population size, but for the three population sizes 30, 60 and 90, $n=10$, which is to be allocated so that at least two units are to be chosen from each stratum.

The comparison between the true optimal allocation and the model allocation according to the type of stratification is given in the following tables. The true optimal allocation is the solution to the optimization problem given by the minimization of (24) subject to (5), while the model allocation is the solution to that given by the one of (60) under (5) in Remark 11. Those solutions satisfying $n_h p_{hi} < 1$ were obtained using 'nonlinear programming (NLP) Procedure' of SAS/OR running convex mathematical programming algorithms. See SAS/OR (2018).

TABLE 1

Comparison of true optimal allocation and model allocation for Stratification Type A

N	n	True Optimal Allocation			Model Allocation			R.E.
		n_1	n_2	n_3	n_1	n_2	n_3	
30	10	2	3	5	2	3	5	1.0000
		(1.64)	(3.29)	(5.07)	(2.13)	(3.12)	(4.75)	
60	10	2	3	5	2	3	5	1.0000
		(1.57)	(3.36)	(5.07)	(2.06)	(3.05)	(4.89)	
90	10	2	3	5	2	3	5	1.0000
		(1.63)	(3.19)	(5.18)	(1.85)	(2.87)	(5.28)	
120	12	2	4	6	2	4	6	1.0000
		(1.97)	(3.79)	(6.24)	(2.28)	(3.40)	(6.32)	
150	15	2	5	8	3	4	8	1.0011
		(2.41)	(4.43)	(8.16)	(2.90)	(4.26)	(7.84)	
180	18	3	5	10	4	5	9	1.0232
		(2.93)	(5.20)	(9.87)	(3.42)	(5.16)	(9.42)	
210	21	3	6	12	4	6	11	1.0102
		(3.29)	(5.71)	(12.00)	(3.95)	(5.97)	(11.08)	
240	24	4	7	13	5	7	12	1.0189
		(3.85)	(6.83)	(13.32)	(4.70)	(6.91)	(12.39)	
270	27	4	8	15	5	8	14	1.0026
		(4.36)	(8.05)	(14.59)	(5.27)	(7.76)	(13.97)	
300	30	5	9	16	6	9	15	1.0092
		(4.98)	(9.01)	(16.01)	(5.94)	(8.60)	(15.46)	

Note: Figures in parentheses are unrounded values and R.E. is the relative efficiency calculated as the variance (9) for model allocation divided by that for true optimal allocation.

Table 1 for stratification type A shows that for the first four populations, the true optimal allocation and model allocation coincide, and for the other populations, the two allocations are very similar, resulting in the relative efficiency (R.E.), defined as the variance of the H-T estimator for model allocation over that for true optimal allocation, being equal to 1 or very close to 1. Table 2 shows that for stratification type B the model allocation for the first two populations is slightly different from the true optimal allocation, and for the others the two allocations are equal. Accordingly, it shows that the model allocation is a good alternative to the true optimal allocation.

TABLE 2

Comparison of true optimal allocation and model allocation for Stratification Type B

N	n	True Optimal Allocation			Model Allocation			R.E.
		n_1	n_2	n_3	n_1	n_2	n_3	
30	10	3	4	3	3	3	4	1.1082
		(2.52)	(4.11)	(3.37)	(2.82)	(3.51)	(3.67)	
60	10	3	3	4	3	4	3	1.0386
		(3.50)	(2.97)	(3.53)	(3.27)	(3.50)	(3.23)	
90	10	3	3	4	3	3	4	1.0000
		(3.13)	(3.24)	(3.63)	(3.26)	(3.16)	(3.58)	
120	12	4	4	4	4	4	4	1.0000
		(4.15)	(3.56)	(4.29)	(4.18)	(3.86)	(3.96)	
150	15	5	5	5	5	5	5	1.0000
		(5.16)	(4.48)	(5.36)	(5.23)	(4.81)	(4.96)	
180	18	6	6	6	6	6	6	1.0000
		(5.59)	(5.99)	(6.42)	(5.96)	(5.99)	(6.05)	
210	21	7	7	7	7	7	7	1.0000
		(7.26)	(6.67)	(7.07)	(6.84)	(7.29)	(6.87)	
240	24	8	8	8	8	8	8	1.0000
		(7.57)	(8.02)	(8.41)	(8.29)	(7.70)	(8.01)	
270	27	9	9	9	9	9	9	1.0000
		(8.42)	(9.37)	(9.21)	(9.11)	(8.96)	(8.93)	
300	30	10	10	10	10	10	10	1.0000
		(9.65)	(10.08)	(10.27)	(10.45)	(9.68)	(9.87)	

Note: See Table 1

5. Concluding remarks

We have addressed the topic of efficient sample allocation in stratified samples using more general super-population regression models than those investigated by Rao (1968). Under more general models that include an intercept term, we have developed several theorems that are useful for deciding sample allocation in πPS sampling designs. Also, through the theorems we have shown how to apply this sample allocation theory for Sampford's (1967) sampling method, one of the more common πPS sampling designs used in survey practice.

Based on the theorems developed in this paper, the optimization problem with respect to the stratum sample sizes can be solved by using software involving convex mathematical programming algorithms. This is a straightforward approach for sample allocation when using more efficient πPS sampling methods.

Also, although we assumed that the super-population parameters are known for the two models, they can be estimated in practice. Including Harvey (1976), Godfrey *et al*. (1984), and Särndal & Wright (1984), there would be many useful references for estimation of model parameters.

In addition to Sampford' sampling, the approach can be applied to a variety of πPS sampling without replacement designs. In future work, it will be important to extend the theory and methods described here to allocation problems under more complicated super-population models and situations where the super-population model can vary across strata. The approach may also be useful in implementing more sophisticated survey designs such as responsive designs, suggested by Groves and Heeringa (2006), to achieve higher quality statistics.

References

ASOK, C. & SUKHATME, B.V. (1976). On Sampford's procedure of unequal probability sampling without replacement. *J. Amer. Statist. Assoc*. **71**, 912-918.

BAYLESS, D.L. & RAO, J.N.K. (1970). An empirical study of stabilities of estimators and variance estimators in unequal probability sampling (n=3 or 4). *J. Amer. Statist. Assoc*. **65**, 1645-1667.

BONDESSON, L., TRAAT, I. & LUNDQVIST, A. (2006). Pareto sampling versus Sampford and conditional Poisson sampling. *Scand. J. Statist.* **33**, 699-720.

BOWLEY, A.L. (1926). Measurement of the precision attained in sampling. *Bulletin of the International Statistical Institute*. **22**, Supplement to Liv. **1**, 6-62.

BREWER, K.R.W. (1963). A model of systematic sampling with unequal probabilities. *Aust. J. Statist.* 5, 5-13.

BREWER, K.R.W. & HANIF, M. (1983). *Sampling with Unequal Probabilities*. Lecture notes in statistics, No. 15. New York: Springer-Verlag.

COCHRAN, W.G. (1977). *Sampling Techniques*. Third edition. New York: John Wiley & Sons.

DAYAL, S. (1985). Allocation of sample using values of auxiliary characteristic. *J. Statist. Plann. Inference* **11**, 321-328.

DES RAJ (1956). A note on the determination of optimum probabilities in sampling without replacement. *Sankhyā* **17**, 197-200.

GABLER, S. (1981). A comparison of Sampford's sampling procedure versus unequal probability sampling with replacement. *Biometrika* **68**, 725-727.

GODFREY, J., ROSHWALB, A. & WRIGHT, R.L. (1984). Model-based stratification in inventory cost estimation. *J. Bus. Econom. Statist.* **2**, 1-9.

GROVES, R.M. & HEERINGA, S.G. (2006). Responsive design for household surveys: tools for actively controlling survey errors and costs. *J. Roy. Statist. Soc. Ser. A* 169, 439-457.

GUPT, B.K. (2003). Sample size allocation for stratified sampling under a correlated superpopulation model. *Metron* **61**, 35-52.

GUPT, B.K. & RAO, T.J. (1997). Stratified PPS sampling and allocation of sample size. *J. Indian. Soc. Agricultural Statist.* **50**, 199-208.

HANSEN, M.H., MADOW, W.G. & TEPPING, B.J. (1983). An evaluation of model-dependent and probability-sampling inferences in sample surveys. *J. Amer. Statist. Assoc*. **78**, 776-793.

HANURAV, T.V. (1965). Optimum sampling strategies and related problems. Ph. D. thesis, Indian Statistical Institute, Calcutta.

HARVEY, A.C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica* **44**, 461-465.

HAZIZA, D., MECATTI, F. & RAO, J.N.K. (2008). Evaluation of some approximate variance estimators under the Rao-Sampford unequal probability sampling design. *Metron* **66**, 91-108.

HORVITZ, D.G., & THOMPSON, D.J. (1952). A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc*. **47**, 663-685.

NEYMAN, J. (1934). On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. *J. Roy. Statist. Soc.* **97**, 558-625.

RAO, J. N.K. (1965). On two simple schemes of unequal probability sampling without replacement. *J. Indian Statist. Assoc.* **3**, 173-180.

RAO, J. N. K. (2005). Interplay between sample survey theory and practice: an appraisal. *Survey Methodology* **31**, 117-138.

RAO, J. N. K. & BAYLESS, D.L. (1969). An empirical study of the stabilities of estimators and variance estimators in unequal probability sampling of two units per stratum. *J. Amer. Statist. Assoc*. **64**, 540-559.

RAO, T. J. (1968). On the allocation of sample size in stratified sampling. *Ann. Inst. Statist. Math.* **20**, 159-166.

RAO, T. J. (1977). Optimum allocation of sample size and prior distributions: a review. *Internat. Statist. Rev.* **45**, 173-179.

REDDY, V.N. (1976). Stratified simple random sampling and prior distributions. *Ann. Inst. Statist. Math.* **28**, 445-459.

SAMPFORD, M.R. (1967). On sampling without replacement with unequal probabilities of selection. *Biometrika* **54**, 499-513.

SÄRNDAL, C.E. (1996). Efficient estimators with simple variance in unequal probability sampling. *J. Amer. Statist. Assoc*. **91**, 1289-1300.

SÄRNDAL, C.E. & WRIGHT, R.L. (1984). Cosmetic form of estimators in survey sampling. *Scand. J. Statist.* **11**, 146-156.

SAS/OR (2018). *User's Guide: Mathematical Programming*, Version 15.1, Cary: SAS Institute Inc.

SMITH, T.M.F. (2001). Biometrika centenary: sample surveys. *Biometrika* **88**, 167-194.

TILLÉ, Y. (2006). *Sampling Algorithms*. Springer series in statistics, New York: Springer Science+Business Media, Inc.