Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

# Statistical methods for sampling cross-classified populations under constraints


by Louis-Paul Rivest

Release date: January 3, 2024

Statistics Canada Statistique Canada

Canada

## How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

**Email at** infostats@statcan.gc.ca

**Telephone,** from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

## Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

## Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada



**An HTML version is also available.**

*Cette publication est aussi disponible en français*.

Survey Methodology, December 2023
Vol. 49, No. 2, pp. 347-362
Statistics Canada, Catalogue No. 12-001-X


# Statistical methods for sampling cross-classified populations under constraints

**Louis-Paul Rivest**[1]

Abstract

The article considers sampling designs for populations that can be represented as a $N \times M$ matrix. For instance when investigating tourist activities, the rows could be locations visited by tourists and the columns days in the tourist season. The goal is to sample cells $(i, j)$ of the matrix when the number of selections within each row and each column is fixed *a priori*. The $i^{\text{th}}$ row sample size represents the number of selected cells within row $i$; the $j^{\text{th}}$ column sample size is the number of selected cells within column $j$. A matrix sampling design gives an $N \times M$ matrix of sample indicators, with entry 1 at position $(i, j)$ if cell $(i, j)$ is sampled and 0 otherwise. The first matrix sampling design investigated has one level of sampling, row and column sample sizes are set in advance: the row sample sizes can vary while the column sample sizes are all equal. The fixed margins can be seen as balancing constraints and algorithms available for selecting such samples are reviewed. A new estimator for the variance of the Horvitz-Thompson estimator for the mean of survey variable $y$ is then presented. Several levels of sampling might be necessary to account for all the constraints; this involves multi-level matrix sampling designs that are also investigated.

**Key Words:** Balanced sampling; Creel surveys; Cube method; Multi-level sampling; Monte Carlo simulation; Variance estimation.

## 1. Introduction

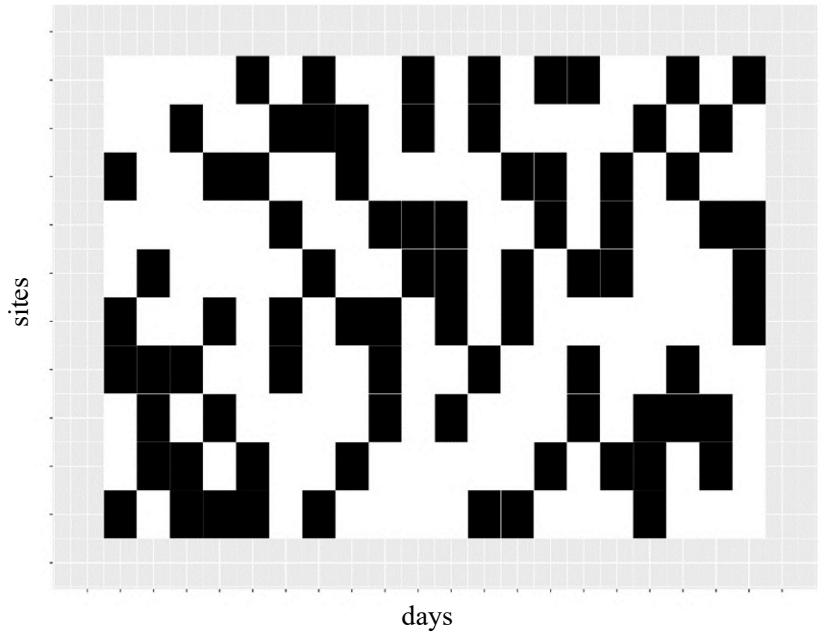
Sampling from cross-classified populations raises interesting statistical issues, see Juillard, Chauvet and Ruiz-Gazen (2017) for a recent discussion. When each cell of the cross-classification contains a single unit, the population to sample has size $NM$ as it can then be viewed as a $N \times M$ matrix. The sample consists of cells $(i, j)$ of the population matrix; this defines the $N \times M$ matrix $\mathbf{Z}$ of sample indicators, with $Z_{ij} = 1$ if cell $(i, j)$ is selected and $Z_{ij} = 0$ otherwise. We focus on designs where the number of selections within each row and each column is fixed *a priori*. We define the $i^{\text{th}}$ row sample size as the number of selected cells within row $i$ while the $j^{\text{th}}$ column sample size is the number of selected cells within column $j$. This work studies matrix sampling designs for which the row and the column sample sizes are predetermined; the row sample sizes vary from one row to the next while all the column sample sizes are equal. This generalizes a stratified sampling design that would apply if the sample sizes for only one dimension, either row or column, were fixed. Multi-level generalizations of the proposed design are also introduced.

Populations having a matrix format occur, for instance, when pooling tourists (Deville and Maumy-Bertrand, 2006) and in creel surveys (Kozfkay and Dillon, 2010); time, that is days, is one dimension of the matrix and the other one is location, such as venues frequented by tourists and fishing sites. Figure 1.1 presents a sample selected in a $10 \times 20$ population matrix with row sample sizes equal to $m = 8$ and column sample sizes equal to $n = 4$. The black entries identify the cells $(i, j)$ that are selected, for which $Z_{ij} = 1$. Having constraints on the row and the column sample sizes is useful in many contexts. In Ida, Rivest and

1. Louis-Paul Rivest, Département de mathématiques et de statistique, Université Laval, 1045 rue de la médecine, Québec, Canada, G1V 0A6. E-mail: louis-paul.rivest@mat.ulaval.ca.

Daigle (2018)'s population matrix the rows are sites and the columns are days. The column sample size $n$ is the number of sites that can be visited in one day while the row sample sizes $\{m_i : i = 1, \ldots, N\}$ are related to sites' importance. The sample matrix presented in Figure1.1 also applies for planning $N$ repeated surveys in a population of size $M$. The $i^{\text{th}}$ row of $\mathbf{Z}$ identifies the units selected in survey $i$. The row sample sizes are determined by the objectives of the individual surveys. The fixed column sample size creates a dependency between the row samples that ensures their coordination (Matei and Tillé, 2005): the response burden is shared equally among all population units. This work focusses mostly on the first application as it investigates the sampling properties of the Horvitz-Thompson estimator of the mean of the survey variable $y$ over the $NM$ population units: $y_{ij}$ is the value of the survey variable for cell $(i, j), i = 1, \ldots, N;$ $j = 1, \ldots, M$. For instance, $y_{ij}$ is the total number of hours of fishing at site $i$ on day $j$ in Ida et al. (2018). The goal is to estimate the population mean, $\bar{y} = \sum_{i,j} y_{ij} \big/ (NM)$ using the Horvitz-Thompson estimator $\hat{\bar{y}} = \sum_{i=1}^{N} \hat{\bar{y}}_{i\bullet} \big/ N$, where $\hat{\bar{y}}_{i\bullet}$ is the sample mean for row $i$.

**Figure 1.1 Sample units, in black, drawn out of a $10 \times 20$ population matrix.**




The set of $N \times M$ 0-1 matrices $\mathbf{Z}$ with row totals given by $Z_{i\bullet} = m_i, i = 1, \ldots, N$ and column totals $Z_{\bullet j} = n, j = 1, \ldots, M$ is fairly large (Barvinok, 2010). The goal of the matrix sampling design is to select uniformly among that set, thus all the $N \times M$ matrices $\mathbf{Z}$ fulfilling the constraints on the row and the column sample sizes are equally likely to be selected. This can be achieved using the cube algorithm (Deville and Tillé, 2004); the constraints on the margins are then interpreted as balancing constraints. Rivest and Ebouele (2020) also discuss other sampling algorithms such as hypergeometric sampling and sampling through a Markov chain defined on the set of acceptable matrices $\mathbf{Z}$.

The next section summarizes the findings of Rivest and Ebouele (2020) for the sampling design illustrated in Figure 1.1. A new approach, called the conditional approach, for the evaluation of the sampling properties of the Horvitz-Thompson estimator is proposed in Section 3. Section 4 suggests a new estimator for the between row covariances. Section 5 considers a hierarchical sampling design to select the rows of the population matrix. It also investigates the properties of the Horvitz-Thompson estimator of the population mean of the survey variable for this new design.

# 2. The properties of the matrix sampling design with fixed row and column totals

Consider a sampling design that selects the $N \times M$ matrix of sample indicators $\{Z_{ij}\}$ uniformly among all the 0-1 matrix with row totals given by $\{m_i\}$ and fixed column total $n$ for $j = 1, \ldots, M$ where $\sum_{i=1}^{N} m_i = Mn$. Given a matrix $\mathbf{Z}$ that meets these constraints, any permutation of the columns of $\mathbf{Z}$ is an acceptable matrix of sample indicators. This implies that the sampling design for selecting the units in row $i$ is without replacement simple random sampling of $m_i$ units among $M$ as all possible samples have the same probability of being chosen. This also entails that the sampling designs for selecting $n$ units among $M$ in a column are identical. It is a without replacement design with variable selection probabilities and a fixed sample size.

The unit inclusion probabilities for cell $(i, j)$ of the proposed design are $\pi_{ij} = \gamma_i = m_i / M$ for $i = 1, \ldots, N$ and $j = 1, \ldots, M$. The joint inclusion probabilities of two cells $(i, j)$ and $(k, \ell)$ is $\pi_{ij,k\ell}$. They can be expressed in terms of $\gamma_{ik}$ the joint inclusion probabilities when sampling units within a column. They are given by

$$\pi_{ij,k\ell} = \begin{cases} \gamma_{ik} & i \neq k,\ j = \ell, \\ m_i(m_i - 1)/\{M(M-1)\} & i = k,\ j \neq \ell, \\ \dfrac{m_i m_k}{M(M-1)} - \dfrac{\gamma_{ik}}{M-1} & i \neq k,\ j \neq \ell. \end{cases} \tag{2.1}$$

The joint selection probabilities in the same row or in the same column of a matrix are deduced from the row and the column sampling designs discussed above. When $i \neq k, j \neq \ell$ the joint selection probabilities only depend on $(i, k)$ because of the column exchangeability. One has

$$m_i m_k = \sum_{j \neq \ell} E(Z_{ij} Z_{k\ell}) + \sum_{j} E(Z_{ij} Z_{kj}) = M(M-1)\pi_{ij,k\ell} + M\gamma_{ik}.$$

Solving this equation gives the general formula for $\pi_{ij,k\ell}$.

The joint probability for sampling rows $i \neq k$ in two different columns, $j \neq \ell$, is larger than that for sampling $i$ and $k$ in the same column as

$$\pi_{ij,k\ell} = \frac{\gamma_{ik}}{M-1}\left( M\frac{\gamma_i \gamma_k}{\gamma_{ik}} - 1 \right) > \gamma_{ik},$$

provided that

$$\gamma_i\gamma_k \; > \; \gamma_{ik}. \tag{2.2}$$

This condition ensures that the Sen-Yates-Grundy variance estimator for the Horvitz-Thompson estimator of a column total is positive. The condition $\gamma_i\gamma_k > \gamma_{ik}$ is satisfied by the conditional Poisson sampling design (Chen and Dempster, 1994). It is conjectured that (2.2) is true for the design for sampling units within a column as this design converges to a conditional Poisson sampling design when $M$ goes to $\infty$ and $N$ is fixed (Rivest, 2021). If true, (2.2) would also mean that the probability $\gamma_{ik}$ that a column is sampled in both row $i$ and row $k$ is less than $\gamma_i\gamma_k$, the inclusion probability if the rows were sampled independently. Thus the fixed row and column totals create a negative coordination between row samples, see Grafström and Matei (2015) for a discussion of positive and negative coordination between samples.

Another interesting result is that, for $i \neq k$ and $j \neq \ell$,

$$\mathrm{Cov}(Z_{ij}, Z_{k\ell}) \; = \; -\frac{1}{M-1}\mathrm{Cov}(Z_{ij}, Z_{kj}) \; = \; -\frac{1}{M-1}(\gamma_{ik} - \gamma_i\gamma_k).$$

This result remains true when $i = k$ provided that $\gamma_{ii}$ is defined as being equal to $\gamma_i = m_i / M$. Indeed, one has $\mathrm{Cov}(Z_{ij}, Z_{i\ell}) = -\gamma_i(1-\gamma_i)/(M-1)$ which is minus the variance of $Z_{ij}$ divided by $M-1$. This is used to prove that the covariance between $\hat{\bar{y}}_{i\bullet}$ and $\hat{\bar{y}}_{k\bullet}$, the sample means for survey variable $y$ in rows $i$ and $k$ is

$$\mathrm{Cov}(\hat{\bar{y}}_{i\bullet}, \hat{\bar{y}}_{k\bullet}) \; = \; \Delta_{ik}S_{ik}, \quad i,k = 1,\ldots,N$$

where $\Delta_{ik} = \{\gamma_{ik} / (\gamma_i\gamma_k) - 1\} / M$ is the $(i,k)$ entry of a $N \times N$ matrix $\mathbf{\Delta}$ and $S_{ik}$ is the covariance between rows $i$ and $k$, $S_{ik} = \sum_{j=1}^{M}(y_{ij} - \bar{y}_{i\bullet})(y_{kj} - \bar{y}_{k\bullet})/(M-1)$, $i,k = 1,\ldots,N$ and $\bar{y}_{i\bullet} = \sum_{j=1}^{M} y_{ij}/M$. This result is used to evaluate the variance of $\hat{\bar{y}} = \sum_{i=1}^{N}\hat{\bar{y}}_{i\bullet}/N$, the Horvitz-Thompson estimator of the mean of $y$:

$$\mathrm{Var}(\hat{\bar{y}}) \; = \; \frac{\mathrm{tr}(\mathbf{S\Delta})}{N^2}, \tag{2.3}$$

where $\mathbf{S}$ is the $N \times N$ covariance matrix of the $y$ column vectors and tr is the trace operator. If (2.2) holds, then the off-diagonal entries of $\mathbf{\Delta}$ are negative and (2.3) is smaller than $\sum_i \Delta_{ii}S_{ii} / N^2$, the variance of the stratified estimator obtained when sampling the rows independently when the between row covariances, $S_{ik}$, are positive. Note also that to evaluate (2.3) one needs numerical values for the joint selection probabilities $\gamma_{ik}$. These can be obtained through simulations or using a numerical algorithm to evaluate the joint selection probabilities for the conditional Poisson sampling design, see Tillé (2006) and Rivest (2021), that can be used as an approximation. A conditional variance formula, that does not use the joint selection probabilities $\gamma_{ik}$ is proposed in the next section.

An obvious choice for an estimator of (2.3) would be the Sen-Yates-Grundy variance estimator. Unfortunately, the condition for it to be positive fails. Indeed the joint selection probability in two different rows and two different columns, $\pi_{ij,k\ell}$ satisfies

$$\pi_{ij,k\ell} - \pi_{ij}\pi_{k\ell} \; = \; -\frac{1}{M-1}(\gamma_{ik} - \gamma_i\gamma_k),$$

which as argued in the discussion of (2.2) should be positive. Thus the Sen-Yates-Grundy variance estimator can be negative and an alternative estimator is needed. A plug-in estimator for (2.3) is given in Section 4. It demands the estimation of the $N(N-1)/2$ between row covariances. The negative coordination between row samples renders the construction of estimators difficult. The proposals in Rivest and Ebouele (2020) are not really satisfactory as they give biased estimators when the column sample sizes are small. Alternative estimators are considered in Section 4.

## 2.1 Extensions to unequal column sample sizes

This section assumes that both the row and the column sample sizes of the matrix $\mathbf{Z}$ vary. They are given by $\{m_i : i = 1,\ldots,N\}$ and $\{n_j : j = 1,\ldots,M\}$. The set of possible samples consists of $N \times M$ 0-1 matrices with fixed row and column totals. All the algorithms reviewed in Rivest and Ebouele (2020) can be used to select the sample uniformly in that set. The resulting design is however rather complex when $n_j$ takes several positive values. There are no closed form expressions for the unit inclusion probability of cell $(i, j)$ and for the joint inclusion of cells $(i, j)$ and $(j, \ell)$ and there does not seem to be a manageable expression for the variance of the Horvitz-Thompson estimator. The limiting sampling design within a column converges, as $M$ goes to infinity, to a generalization of the conditional Poisson sampling design with untractable single inclusion probabilities, see Rivest (2021). Thus, to implement a design with varying column sample sizes, a simple solution is to stratify by column sample size. Independent matrix designs are then used to select the matrix sample in each stratum.

The matrix sampling design of the previous section can be extended to situations where the two possible column sample sizes are either 0 or $n$. Suppose that out of $M$ columns, $M_0 < M$ have a non-null sample. The row sample sizes $\{m_i : i = 1,\ldots,N\}$ satisfy $\sum_i m_i = M_0 n$. The selection of a matrix $\mathbf{Z}$ of sample indicators proceeds in two steps. Step 1 uses without replacement simple random sampling to select the $M_0$ columns with non-null samples and a matrix design is used at step 2 to select the sampled cells in the $M_0$ columns chosen at step 1.

Let $\gamma_{ik}$ be the conditional joint selection probability for rows $i$ and $k$ given that the column has been selected at step 1. The findings of the previous section are easily generalized to this new design. For instance

$$\mathrm{Cov}(Z_{ij}, Z_{k\ell}) = -\frac{1}{M-1}\mathrm{Cov}(Z_{ij}, Z_{kj}) = -\frac{1}{M-1}\left(M_0\gamma_{ik}/M - m_i m_k/M^2\right).$$

In addition variance formula (2.3) holds with a matrix $\Delta^g$ defined by $\Delta^g_{ik} = \{MM_0\gamma_{ik}/(m_i m_k) - 1\}/M$ for $i, k = 1,\ldots,n,$ provided that one sets $\gamma_{ii} = m_i/M_0$.

# 3. A conditional matrix sampling design

This section discusses a conditional sampling design for which the matrix of sample indicators $\mathbf{Z}$ is fixed, up to a random permutation of its columns. It derives a conditional alternative to (2.3), the variance of the Horvitz-Thompson estimator.

Let $\mathbf{Z}_0$ be a 0-1 matrix with row sums equal to $\{m_i\colon i=1,\ldots,N\}$ and column sums all equal to $n$. Suppose that the random matrix of sample indicators $\mathbf{Z}$ is obtained by randomly permuting the columns of $\mathbf{Z}_0$. This conditional sampling design shares many of the properties discussed in Section 2. The design for sampling units within row $i$ is simple random sampling of $m_i$ units in a population of size $M$. The design for sampling units within a column is the same for each column. This design gives a probability of $1/M$ to each of the columns of $\mathbf{Z}_0$ and the probability for selecting unit $i$ and $k$ within a column is

$$\gamma_{ik}^c = \sum_{j=1}^{M} Z_{0ij} Z_{0kj} \Big/ M, \tag{3.1}$$

where $Z_{0ij}, Z_{0kj}$ are the entries $(i, j)$ and $(k, j)$ of $\mathbf{Z}_0$ and exponent $c$ stands for conditional. For this design (2.1) holds with $\gamma_{ik}$ replaced by the conditional joint selection probability $\gamma_{ik}^c$. In some instances, the sampling design proposed in Section 2 is a conditional sampling design. This occurs when the column sample size is either $n=1$ or $n=N-1$ since all the possible sample indicator matrices $\mathbf{Z}$ are then equal up to a permutation of their columns.

The conditional variance of the Horvitz-Thompson estimator $\hat{\bar{y}}$ is a function of the $N\times N$ matrix $\mathbf{\Delta}^c$ whose $(i,k)$ entry is equal to $\Delta_{ik}^c = \left\{\gamma_{ik}^c/(\gamma_i\gamma_k)-1\right\}/M$, $\gamma_{ik}^c$ is defined in (3.1) and $\gamma_i = m_i/M$, as defined in Section 2. It is given by

$$\mathrm{Var}_c\left(\hat{\bar{y}}\right) = \frac{\mathrm{tr}\left(\mathbf{S}\mathbf{\Delta}^c\right)}{N^2}. \tag{3.2}$$

Given a random matrix $\mathbf{Z}$ of sample indicators obtained with one of the algorithms presented in Rivest and Ebouele (2020), one can evaluate the conditional joint inclusion probabilities using (3.1), applied to the matrix $\mathbf{Z}$. Then the conditional variance formula (3.2) is simpler than (2.3) as it does not require the evaluation of the unconditional joint selection probabilities $\gamma_{ik}$. The derivation of a simple variance formula is the main application of the conditional approach.

The conditional matrix sampling design is a low entropy design and it may happen that (3.2) cannot be estimated. This occurs when one of the conditional joint selection probabilities (3.1) is equal to 0. Consider, for instance, the design in the Monte Carlo simulations of Section 4.1. It has $N=9$, $M=36$, $n=2$ and row sample sizes $m_i$ varying between 6 and 11. This design involves $N(N-1)/2=36$ joint selection probabilities and it not possible to find a matrix $\mathbf{Z}_0$ with row totals varying between 6 and 11 for which the 36 values of (3.1) are positive. For this design, only (2.3) can be estimated.

# 4. A new estimator for the between row covariance

This section suggests new estimators for the $N\times N$ covariance matrix $\mathbf{S}$ for $y$. The diagonal elements of $\mathbf{S}$ are easily estimated using the row sample variances. Rivest and Ebouele (2020) use the columns that are sampled in both rows, $i$ and $k$, to estimate the covariance $S_{ik}$. The joint sample size for rows $i$ and $k$ is often less than 2, considering the negative coordination between row samples noted in Section 2. Thus many covariances cannot be estimated using this approach and an alternative estimation strategy is proposed

in this section. It gives nearly unbiased estimators of the variances (2.3) and (3.2) of the Horvitz-Thompson estimator

The proposed covariance estimator relies on the following expression for the covariance as a U-statistic,

$$\begin{aligned} S_{ik} &= \frac{1}{M-1}\sum_{j=1}^{M}(y_{ij}-\bar{y}_{i\bullet})(y_{kj}-\bar{y}_{k\bullet}) \\ &= \frac{1}{2M(M-1)}\left\{\sum_{j\neq\ell}^{M}(y_{ij}-y_{k\ell})^2-(M-1)\sum_{j=1}^{M}(y_{ij}-y_{kj})^2\right\}, \end{aligned} \tag{4.1}$$

where $i \neq k$. See Appendix for a derivation of (4.1).

The new covariance estimator uses the joint selection probabilities $\gamma_{ik}$ of Section 2 that are assumed to be strictly positive, to construct estimators of the two terms in (4.1). This yields

$$\begin{aligned} \hat{S}_{ik} &= \frac{1}{2M(M-1)}\left\{\sum_{j\neq\ell}^{M}\frac{Z_{ij}Z_{k\ell}(y_{ij}-y_{k\ell})^2}{M\gamma_i\gamma_k/(M-1)-\gamma_{ik}/(M-1)}-(M-1)\sum_{j=1}^{M}\frac{Z_{ij}Z_{kj}(y_{ij}-y_{kj})^2}{\gamma_{ik}}\right\} \\ &= \frac{1}{2M}\left\{\sum_{j\neq\ell}^{M}\frac{Z_{ij}Z_{k\ell}(y_{ij}-y_{k\ell})^2}{M\gamma_i\gamma_k-\gamma_{ik}}-\sum_{j=1}^{M}\frac{Z_{ij}Z_{kj}(y_{ij}-y_{kj})^2}{\gamma_{ik}}\right\}. \end{aligned} \tag{4.2}$$

The plug-in unbiased variance estimator for (2.3) is simply $v_{\text{PI}} = \text{tr}(\hat{\mathbf{S}}\Delta)/N^2$, where $\hat{\mathbf{S}}$ has entries given by (4.2). Observe that this covariance estimator can be constructed for the unconditional and the conditional sampling designs presented in Sections 2 and 3. For a conditional sampling design, one replaces $\gamma_{ik}$ by $\gamma_{ik}^c$, see (3.1), in (4.2) as long as $\gamma_{ik}^c > 0$.

Covariance estimator (4.2) is very variable as its second term involves a division by $\gamma_{ik}$ that can be very small. This leads to a covariance matrix estimator $\hat{\mathbf{S}}$ which is not positive definite and to an estimator for (2.3) that can, on some rare occasions, be negative. A more stable estimator can be obtained by assuming that all the between row correlations are equal. An estimator of the common correlation is then

$$\hat{\rho} = \frac{\sum_{i>k}\hat{S}_{ik}}{\sum_{i>k}\sqrt{\hat{S}_{ii}\hat{S}_{kk}}}, \tag{4.3}$$

where $\hat{S}_{ii}$ and $\hat{S}_{kk}$ are the sample variances for row $i$ and row $k$ respectively. An alternative covariance estimator is $\hat{S}_{ik}^{(\text{ec})} = \hat{\rho}\sqrt{\hat{S}_{ii}\hat{S}_{kk}}$. It leads to an equal correlation estimator, $v_{\text{ec}}$, for (2.3). A stratified variance estimator, valid if the rows were sampled independently, $v_{\text{str}} = \sum_i \Delta_{ii}\hat{S}_{ii}/N^2$, is also included as a benchmark in the Monte Carlo simulations that are presented next.

## 4.1 A Monte Carlo investigation of the sampling properties of the new variance estimators

The sampling properties of variance estimator (4.2) and its equal correlation alternative are investigated in two population matrices. The first one has $N = 9$, $M = 36$, column sample size is $n = 2$ and row sample sizes $m_i$ are (11,10,7,10,11,5,6,6,6). In the second one, $M$ and the row sample sizes are doubled while

$N = 9$ and $n = 2$ are unchanged. The $y$ variable in cell $(i, j)$ has a log-normal distribution with expectation $m_i$ and variance $1.72 \times m_i^2$. It is given by

$$y_{ij} = m_i \exp(a_j + e_{ij}) \quad i = 1,\ldots N,\ j = 1,\ldots M,$$

where the column effect $a_j$ and the errors $e_{ij}$ are independent variables respectively distributed according to a $N\left(-\sigma_a^2/2, \sigma_a^2\right)$ and a $N\left(-\sigma_e^2/2, \sigma_e^2\right)$ distributions where $\sigma_a^2 + \sigma_e^2 = 0.8$. Simulations with $\sigma_a^2 = 0$, 0.2 and 0.4 are reported. This corresponds to a between row correlation $\rho = \left\{\exp(\sigma_a^2) - 1\right\} \Big/ \left\{\exp(\sigma_a^2 + \sigma^2) - 1\right\}$ of respectively 0, 0.18 and 0.40. Table 4.1 uses 6 simulated populations and the moments of variance estimators are calcuated using $10^4$ Monte Carlo samples. These Monte Carlo samples are drawn using the MCMC swap algorithm of Oksanen, Blanchet, Friendly, Kindt, Legendre, McGlinn, Minchin, O'Hara, Simpson, Solymos, Stevens, Szoecs and Wagner (2020) discussed in Rivest and Ebouele (2020) to which the reader is referred for more details on the simulations. The results are reported in Table 4.1.

When $\rho = 0$ the three variance estimators are nearly unbiased in Table 4.1. The simple stratified variance estimator does not capture the positive correlation between rows and over-estimates the variance when $\rho > 0$. The plug-in and equal correlation estimators are unbiased for the 6 populations considered in Table 4.1. The standard deviations are also revealing. That for $v_{\text{str}}$ is small as this estimator does not depend on the covariance estimators. The plug-in estimator is the most variable while the equal correlation that is based on the average covariance has a smaller standard deviation.

**Table 4.1**
**The variance of $\hat{\bar{y}}$ evaluated using (2.3) and the expectations, and standard deviations between parenthesis, of three variance estimators, the stratified estimator $(\text{str})$, the plug-in estimator $(\text{PI})$, the equal correlation estimator $(\text{ec})$.**

| $M$ | $\rho$ | $\text{var}\left(\hat{\bar{y}}\right)$ | $v_{\text{str}}$ | $v_{\text{PI}}$ | $v_{\text{ec}}$ |
|---|---|---|---|---|---|
| 36 | 0 | 0.932 | 0.913 (0.684) | 0.930 (0.757) | 0.928 (0.746) |
| 36 | 0.18 | 0.653 | 0.800 (0.258) | 0.654 (0.288) | 0.657 (0.280) |
| 36 | 0.40 | 0.544 | 0.730 (0.306) | 0.542 (0.338) | 0.543 (0.329) |
| 72 | 0 | 0.486 | 0.506 (0.225) | 0.485 (0.252) | 0.486 (0.245) |
| 72 | 0.18 | 0.350 | 0.424 (0.136) | 0.350 (0.151) | 0.349 (0.145) |
| 72 | 0.40 | 0.334 | 0.564 (0.182) | 0.334 (0.190) | 0.332 (0.182) |

To explain the near unbiasedness of $v_{\text{ec}}$ in Table 4.1, one easily checks that this estimator is indeed unbiased when all the row sample sizes are constant and equal to $m$. The matrix $\boldsymbol{\Delta}$ can then be expressed in terms of the $N \times N$ identity matrix $\mathbf{I}_N$ and a $N \times 1$ vector of 1's, $\mathbf{1}_N$, as $\boldsymbol{\Delta} = (N - n)(\mathbf{I}_N - \mathbf{1}_N \mathbf{1}_N^\top / N) \Big/ \left\{m(N-1)\right\}$ and (2.3) only involves $\sum_{i \neq k} S_{ik}$. It is easily checked that the equal correlation estimator of that sum is unbiased as (4.2) is unbiased.

# 5. A multi-level matrix design

This section suggests multi-level generalizations of the matrix sampling designs of Sections 2 and 3. It involves clusters of rows and level 1 selects a sample of clusters for each column. This is done using a level 1 random 0-1 matrix with fixed margins. A population matrix is then created for each level 1 cluster of rows;

the number of rows in the matrix is the size of the cluster and the number of columns is equal to the column total for that cluster in the level 1 sample indicator matrix. Level 2 selection is done independently in the population matrices of each level 1 cluster. Level three sampling is done in a similar way. This section focusses on two-level designs and uses the following notation:

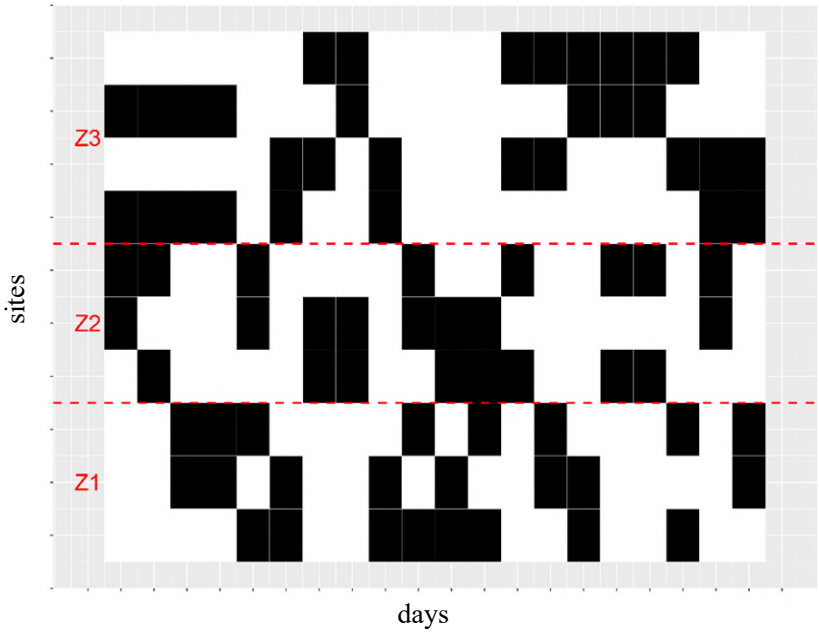
- $N^{(1)}$ is the number row clusters. The size of the level 1 matrix of sample indicators is $N^{(1)} \times M$; the row and column sample sizes are $\{m_i^{(1)}: i = 1, \ldots, N^{(1)}\}$ and $n^{(1)}$;
- $N_i^{(2)}, i = 1, \ldots, N^{(1)}$ are the sizes of the level 1 row clusters. The level 2 matrix of sample indicators for cluster $i$ has dimension $N_i^{(2)} \times m_i^{(1)}$; the row and column sample sizes are $\{m_{k(i)}^{(2)}: i = 1, \ldots, N_i^{(2)}\}$ and $n^{(2)}$;
- The matrices of sample indicators can be combined in a matrix with $N = \sum N_i^{(2)}$ rows and $M$ columns. The column sample size is $n = n^{(1)} \times n^{(2)}$ and the sample size for the $k^{\text{th}}$ row of cluster $i$ is $m_{k(i)}^{(2)}$.

Note $\sum_i m_i^{(1)} \Big/ n_{(1)} = M$ while for $i = 1, \ldots, N^{(1)}$, the sum of the row totals of cluster $i$, has to be a multiple of $n^{(2)}$ as

$$\sum_k m_{k(i)}^{(2)} \Big/ n^{(2)} = m_i^{(1)}. \tag{5.1}$$

Figure 5.1 gives a sample obtained with two levels of sampling in a matrix similar to that of Figure 1.1. One has $N = 10$, $M = 20$, while the row sample sizes are $m = 8$ and column sample sizes are $n = 4$. The first level of sampling involves three clusters of rows, of respective size 3, 3, and 4. The 4 $\mathbf{Z}$ matrices used to construct Figure 5.1 are given in Appendix.

**Figure 5.1 A sample drawn from a $10 \times 20$ population matrix using a two level matrix sampling design.**



The hierarchical design proposed in this section is useful to accommodate constraints. For instance, when the rows are sites, a cluster is a set of neighboring sites than can be visited on the same day. When planning repeated, say monthly, surveys for the same population over one year, clusters could be used to ensure that a unit is surveyed only either in the first six months or in the last six months of the year.

The hierarchical design shares a key property with the single level sampling designs presented in Sections 2 and 3. Given a matrix of sample indicators that meets the row and the column constraints, such as that given in Figure 5.1, any permutation of the columns of that matrix gives an acceptable sample matrix. Thus, under this hierarchical sampling scheme, the sampling design for selecting units within a row is without replacement simple random sampling. The selection probability for row $k$ of cluster $i$ is $m^{(2)}_{k(i)} / M$. This is the product of the level 1 selection probability for cluster $i$, $m^{(1)}_i / M$ times the level 2 selection probability $m^{(2)}_{k(i)} / m^{(1)}_i$ in row $k$ of cluster $i$. All the designs for sampling units within a column are identical. It has two levels and its joint selection probabilities $\{\gamma^{(2)}_{k(i),\ell(j)}: i, j = 1, \ldots, N^{(1)}, k = 1, \ldots, N^{(2)}_i, \ell = 1, \ldots, N^{(2)}_j\}$, are calculated as follows:

- For rows in the same cluster, $i = j$ then $\gamma^{(2)}_{k(i),\ell(i)} = \left(m^{(1)}_i / M\right)\gamma^{(2)}_{i,k,\ell}$ where $\gamma^{(2)}_{i,k,\ell}$ is the joint selection probabilities for rows $k$ and $\ell$ of cluster $i$ at the second level of sampling.
- For rows in different clusters, $i \neq j$, then $\gamma^{(2)}_{k(i),\ell(j)} = \gamma^{(1)}_{i,j}\left(m^{(2)}_{k(i)} / m^{(1)}_i\right)\left(m^{(2)}_{\ell(j)} / m^{(1)}_j\right)$ is the product of the joint selection probability for these two clusters at level 1, times the level 2 single selection probabilities for rows $k$ and $\ell$ in clusters $i$ and $j$.

In this construction the joint selection probabilities, for levels 1 and 2, can be approximated by those of conditional Poisson sampling as discussed in Section 2.

It is now convenient to change the notation and to let $i, k = 1, \ldots, N$ denote rows of the design matrix, in agreement with Sections 2 and 3. The selection probability in row $i$ is $\gamma_i = m_i / M$ and the joint selection probabilities are $\{\gamma^h_{ik}\}$, where exponent $h$ means hierarchical. All the findings of Sections 2 and 3 apply to the hierarchical design provided that the joint selection probabilities $\{\gamma_{ik}\}$ are replaced by their multi-level alternatives, $\{\gamma^h_{ik}\}$. For instance (2.1) and (2.3) hold for this new design, when written is terms of $\{\gamma^h_{ik}\}$ and the matrix $\mathbf{\Delta}^h$ whose $(i,k)$ entry is $\Delta^h_{ik} = \left\{\gamma^h_{ik} / (\gamma_i\gamma_k) - 1\right\} / M$. The variance estimators proposed in Section 4 applies to the hierarchical design proposed in this section provided that all the joint selection probabilities $\{\gamma^h_{ik}\}$ are strictly positive. The estimates of (2.3) and (3.2) are evaluated using the matrix $\mathbf{\Delta}^h$ for the hierarchical design and the covariance estimator of Section 4.

## 5.1 A Monte Carlo investigation

This section revisits the Monte Carlo simulations of Section 4.1. The populations are sampled using a two level design: within each column the two rows sampled need to belong to the same cluster. The clusters of rows for level 1 consist of $\{1, 2, 3\}$, $\{4, 5, 6\}$, and $\{7, 8, 9\}$ and the row sample sizes are (11, 10, 7, 10, 11, 5, 6, 6, 6), the same as those used in Section 4.1. At level 1, $\mathbf{Z}$ is a $3 \times 36$ matrix. Its column totals are

$n=1$ while from (5.1), its row totals are (14, 13, 9). The three $\mathbf{Z}$ matrices for level 2 have respectively 14, 13, and 9 columns and three rows; their column total is $n=2.$ For this problem the designs of Sections 2 and 3 are the same as the column totals of all the $\mathbf{Z}$ matrices are either 1 or $N-1.$

Since only one cluster is sampled in each column the joint selection probability $\gamma_{ik}^{h}$ is 0 for two rows $(i,k)$ belonging to different clusters. Thus the covariance $S_{ik}$ of the survey variable between these two rows is not estimable. Indeed only 9 of the $36\left(=8\times 9/2\right)$ covariances can be estimated in the simulation design. The variances (2.3) and (3.2) are not estimable. The simulation study investigates the performance of the equal correlation estimator $v_{\text{ec}}$ where the common correlation is estimated, through (4.3), using the 9 covariances that are estimable.

**Table 5.1**
**The variance of $\hat{\bar{y}}$ evaluated using (2.3) and the expectations, and standard deviations between parenthesis, of two variance estimators, the stratified estimator $(\text{str})$ and the equal correlation estimator $(\text{ec})$.**

| $M$ | $\rho$ | $\text{var}\left(\hat{\bar{y}}\right)$ | $v_{\text{str}}$ | $v_{\text{ec}}$ |
|---|---|---|---|---|
| 36 | 0 | 0.888 | 0.922 (0.694) | 0.968 (0.759) |
| 36 | 0.18 | 0.582 | 0.796 (0.244) | 0.684 (0.360) |
| 36 | 0.40 | 0.587 | 0.738 (0.309) | 0.521 (0.338) |
| 72 | 0 | 0.521 | 0.504 (0.230) | 0.449 (0.251) |
| 72 | 0.18 | 0.318 | 0.421 (0.136) | 0.362 (0.151) |
| 72 | 0.40 | 0.334 | 0.561 (0.183) | 0.321 (0.195) |

The population sampled are the same as those investigated in Section 4.1. Even if the sampling designs differ, the expectations and variances of the stratified variance estimator $v_{\text{str}}$ are identical in the two experiments. Indeed the $v_{\text{str}}$ entries for Tables 4.1 and 5.1 are the same up to Monte Carlo errors. The bias of the equal correlation estimator $v_{\text{ec}}$ ranges between 3% and 20%. It is smaller than that of $v_{\text{str}}$ when $\rho>0.$ Thus variance estimation is a problem when the first level column sample size is 1. Estimator $v_{\text{str}}$ provides an upper bound for the variance while the validity of $v_{\text{ec}}$ rests on an homogeneity assumption that can be verified, at least in part, by comparing the correlation coefficients that are estimable.

## 5.2 A complex example

This section discusses a complex sampling design presented in Ida et al. (2018) that involves a $54\times 33$ population matrix. The goal is to estimate the fishing effort $\bar{y}$ over $M=33$ days at 9 sites, grouped in three clusters. In Table 5.2 the sites are numbered from 1 to 9. Sites 1, 3, 7, and 9 have more fishermen; their planned number of visits is about twice that for the other sites. Each day a site can be visited at 6 time points, 2 in the morning (AM), 2 in the afternoon (PM) and 2 in the evening (EV). So for each day there are $N=9\times 6=54$ site-time-points. The sampling design selects 4 site-time-points on each day, under two constraints: the sites visited must be in the same cluster and the four visits must be selected as two blocks of two visits in either the morning, the afternoon or the evening. Three levels of sampling are needed to

address these constraints. Because of these constraints it is not feasible to have the number of visits to the more important sites, 1, 3, 7, and 9, exactly equal to twice that at the other sites. The row totals given in Table 5.2 correspond to an approximate solution to the determination of the site-time-point sample sizes. Other approximate solutions are possible; those considered in Ida et al. (2018) are obtained by running the cube algorithm, modified for highly stratified populations by Hasler and Tillé (2014), for the three levels of sampling.

Appendix shows how to obtain a matrix of sample indicators for the $54 \times 33$ population matrix with column total $n = 4$ and row totals given in Table 5.2. There are three level of sampling. Level 1 selects one cluster for each day (column), level 2 proceeds cluster by cluster and selects two time periods for each day it is visited. Level 3 sampling is applied within each cluster period; it selects the two sites that will be visited at the two time point in the period. Level three sampling is stratified: one site is selected for each time point. This involves 13 $\mathbf{Z}$ matrices of sample indicators.

**Table 5.2**
**Vector of 54 row totals for the $54 \times 33$ matrix $\mathbf{Z}$ for the design of Ida et al. (2018).**

| Cluster | Period | Time-point-site | | | | | | Tot |
|---|---|---|---|---|---|---|---|---|
| | | 1 – 1 | 1 – 2 | 1 – 3 | 2 – 1 | 2 – 2 | 2 – 3 | |
| 1 | AM | 4 | 1 | 3 | 3 | 2 | 3 | 16 |
| 1 | PM | 3 | 2 | 3 | 3 | 2 | 3 | 16 |
| 1 | EV | 3 | 2 | 3 | 3 | 1 | 4 | 16 |
| 1 | Tot | 10 | 5 | 9 | 9 | 5 | 10 | 48 |
| | | 1 – 4 | 1 – 5 | 1 – 6 | 2 – 4 | 2 – 5 | 2 – 6 | |
| 2 | AM | 1 | 2 | 2 | 2 | 1 | 2 | 10 |
| 2 | PM | 2 | 2 | 1 | 2 | 2 | 1 | 10 |
| 2 | EV | 2 | 2 | 2 | 2 | 2 | 2 | 12 |
| 2 | Tot | 5 | 6 | 5 | 6 | 5 | 5 | 32 |
| | | 1 – 7 | 1 – 8 | 1 – 9 | 2 – 7 | 2 – 8 | 2 – 9 | |
| 3 | AM | 3 | 2 | 4 | 4 | 1 | 4 | 18 |
| 3 | PM | 3 | 2 | 3 | 3 | 1 | 3 | 16 |
| 3 | EV | 3 | 2 | 4 | 4 | 1 | 3 | 18 |
| 3 | Tot | 9 | 6 | 11 | 11 | 5 | 10 | 52 |

In Table 5.2, equation (5.1) means that the total sample size for each cluster is a multiple of 4 while that for each of the 9 cluster-periods is even. Once a matrix of sample indicators has been selected, either with the cube algorithm or by selecting the 13 random $\mathbf{Z}$ matrices implementing the hierarchical design presented in the appendix, the conditional approach of Section 3 is a relatively straightforward method to estimate the variance. This involves two $54 \times 54$ matrices, an estimated covariance matrix $\hat{\mathbf{S}}$ and $\boldsymbol{\Delta}^c$, evaluated using (3.1). To account for the non estimability of some covariances found in the example considered in Section 5.1, the two variance estimators investigated in the simulation study reported in Table 5.1 could be used. Taking $\hat{\mathbf{S}}$ as a diagonal matrix of row variances gives the stratified variance estimator. The equal correlation estimator $\hat{\mathbf{S}}$ can also be evaluated. Thus two methods of variance estimation are available for this complex problem.

# 6. Discussion

Many samplings problems face operational constraints that need to be addressed when designing a survey, see for instance Vallée, Ferland-Raymond, Rivest and Tillé (2015) for a forestry example. One strategy to address these constraints is to use the cube method, possibly within a multi-level design, after a careful specification of the selection probabilities. This paper proposes an alternative strategy for cross-classified populations where the constraints can be expressed in terms of fixed row and column sample sizes. As illustrated in Section 5.2, this strategy involves setting up a population matrix and target row totals for the matrix of sample indicators. Sample selection is done by selecting relatively small sample indicator matrices $\mathbf{Z}$ uniformly over the set of feasible matrices at each level of the design. The hypergeometric sampling algorithm of Rivest and Ebouele (2020) is suited for this problem as it does not need an initial value $\mathbf{Z}_0$, that is required by MCMC algorithms. One advantage of this approach is that the constraints on row and column totals are always verified while the cube method sometimes fails to meet them exactly. An interesting feature of the methodology proposed in this paper is the availability of variance estimators. If, at all the levels, the column sample sizes are larger than 2, then an unbiased variance estimator for the Horvitz-Thompson estimator of the mean of the survey variable can be constructed.


# Acknowledgements

This project benefitted from the financial assistance of the Canada Research Chair in Statistical Sampling and Data Analysis and from a discovery grant from the Natural Sciences and Engineering Research Council of Canada. The constructive comments of the special editor and of the referees are gratefully acknowledged. I am also grateful to Anne-Sophie Julien, who contributed to the construction of the sample design proposed in Section 5.2.


# Appendix:

### Proof of (4.2)

One has

$$\begin{aligned}
&\frac{1}{2M(M-1)}\left\{\sum_{j\neq\ell}^{M}(y_{ij}-y_{k\ell})^2-(M-1)\sum_{j}^{M}(y_{ij}-y_{kj})^2\right\}\\
&\qquad=\frac{1}{2M(M-1)}\left\{\sum_{j,\ell=1}^{M}(y_{ij}-y_{k\ell})^2-M\sum_{j}^{M}(y_{ij}-y_{kj})^2\right\}\\
&\qquad=\frac{1}{M-1}\left\{\sum_{j=1}^{M}y_{ij}y_{kj}-\frac{\sum_{j=1}^{M}y_{ij}\sum_{\ell=1}^{M}y_{k\ell}}{M}\right\}\\
&\qquad=S_{ik}.
\end{aligned}$$

## The matrices of sample indicators needed to construct Figure 5.1

The $3 \times 20$ matrix for selecting clusters $Z3,\ Z2,$ and $Z1$ respectively within each column is given by

$$\mathbf{Z}_1 = \begin{pmatrix} 1&1&1&1&0&1&1&1&1&0&0&0&1&1&1&1&1&1&1&1 \\ 1&1&0&0&1&0&1&1&0&1&1&1&1&0&0&1&1&0&1&0 \\ 0&0&1&1&1&1&0&0&1&1&1&1&0&1&1&0&0&1&0&1 \end{pmatrix}.$$

The three matrices of sample indicators for level 2 sampling in respectively zones 3, 2 and 1 are

$$\mathbf{z}_{Z3} = \begin{pmatrix} 0&0&0&0&0&1&1&0&1&1&1&1&1&1&0&0 \\ 1&1&1&1&0&0&1&0&0&0&1&1&1&0&0&0 \\ 0&0&0&0&1&1&0&1&1&1&0&0&0&1&1&1 \\ 1&1&1&1&1&0&0&1&0&0&0&0&0&0&1&1 \end{pmatrix}$$

$$\mathbf{z}_{Z2} = \begin{pmatrix} 1&1&1&0&0&1&0&0&1&1&1&1 \\ 1&0&1&1&1&1&1&1&0&0&0&1 \\ 0&1&0&1&1&0&1&1&1&1&1&0 \end{pmatrix} \quad \mathbf{z}_{Z1} = \begin{pmatrix} 1&1&1&0&0&1&0&1&1&0&1&1 \\ 1&1&0&1&1&0&1&0&1&1&0&1 \\ 0&0&1&1&1&1&1&1&0&1&1&0 \end{pmatrix}.$$

The (10, 5) and (8, 5) entries in Figure 5.1 are equal to 1. This information can be retrieved from $\mathbf{Z}_1$ and $\mathbf{Z}_{Z1}$: in the fifth column of $\mathbf{Z}_1$ zones 2 and 1 are sampled. It is the third times that zone 1 is sampled so that the third column of $\mathbf{Z}_{Z1}$ informs us that the first and third row of $Z1$ are sampled in that column. This translates into black boxes for entries (10, 5) and (8, 5) of Figure 5.1.

## Section 5.2: Some matrices of sample indicators for a three level sampling problem

To simplify the presentation we use matrices $\mathbf{Z}_0$ where identical columns are pooled together. A matrix of sample indicators is obtained by taking a random permutation of $\mathbf{Z}_0$. For the first level of sampling $\mathbf{Z}_0$ can be expressed using row vectors of ones, $\mathbf{1}_n^\top$, and zeros, $\mathbf{0}_n^\top$, where $n$ is the length of the vector. The level 1 $\mathbf{Z}_0$ matrix is

$$\mathbf{Z}_0^{(1)} = \begin{pmatrix} \mathbf{1}_{12}^\top & \mathbf{0}_8^\top & \mathbf{0}_{13}^\top \\ \mathbf{0}_{12}^\top & \mathbf{1}_8^\top & \mathbf{0}_{13}^\top \\ \mathbf{0}_{12}^\top & \mathbf{0}_8^\top & \mathbf{1}_{13}^\top \end{pmatrix}.$$

The second cluster is selected 8 times. A $\mathbf{Z}_0$ matrix for choosing the periods, AM, PM or EV, on the days that cluster two is selected is

$$\mathbf{Z}_{0,2}^{(2)} = \begin{pmatrix} 1&1&1&1&1&0&0&0 \\ 1&1&0&0&0&1&1&1 \\ 0&0&1&1&1&1&1&1 \end{pmatrix},$$

where the exponent gives the sampling level and the index accompanying 0 is the cluster to which this matrix $\mathbf{Z}_0$ applies. The row totals for $\mathbf{Z}_{0,2}^{(2)}$, 5, 5, and 6 are half the AM, PM, EV totals for cluster 2 in

Table 5.2. We now consider period AM for cluster 2. Considering row 1 of $\mathbf{Z}_{0,2}^{(2)}$, AM is selected 5 times. The third level matrix has five columns and 6 rows, corresponding to the time-point-site (1-4, 1-5, 1-6, 2-4, 2-5, 2-6) with row totals (1, 2, 2, 2, 1, 2) as given in the AM line for cluster 2, see Table 5.2. Here one has to stratify by time point: one site need to be selected at each time point. A candidate $\mathbf{Z}_0$ where the first three rows are for time point 1 and the last 3 for time point 2 is

$$\mathbf{Z}_{0,2,\mathrm{AM}}^{(3)} = \left(\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ \hline 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right).$$

Thus the set of possible level 3 matrices $\mathbf{Z}_{2,\mathrm{AM}}^{(3)}$ is obtained by permuting the columns of the first three and of the last three rows of $\mathbf{Z}_{0,2,\mathrm{AM}}^{(3)}$ independently. The third level of sampling involves 9 matrices similar to $\mathbf{Z}_{0,2,\mathrm{AM}}^{(3)}$.

# References


Barvinok, A. (2010). On the number of matrices and a random matrix with prescribed row and column sums and 0-1 entries. *Advances in Mathematics*, 224(1), 316-339.

Chen, X.-H., and Dempster, A.P. (1994). Weighted finite population sampling to maximize entropy. *Biometrika*, 81(3), 457-469.

Deville, J.-C., and Maumy-Bertrand, M. (2006). Extension of the indirect sampling method and its application to tourism. *Survey Methodology*, 32, 2, 177-185. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2006002/article/9552-eng.pdf.

Deville, J.-C., and Tillé, Y. (2004). Efficient balanced sampling: the cube method. *Biometrika*, 91(4), 893-912.

Grafström, A., and Matei, A. (2015). Coordination of conditional Poisson samples. *Journal of Official Statistics*, 31(4), 649-672.

Hasler, C., and Tillé, Y. (2014). Fast balanced sampling for highly stratified population. *Computational Statistics & Data Analysis*, 74, 81-94.

Ida, I.O., Rivest, L.-P. and Daigle, G. (2018). Using balanced sampling in creel surveys. *Survey Methodology*, 44, 2, 239-252. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2018002/article/54954-eng.pdf.

Juillard, H., Chauvet, G. and Ruiz-Gazen, A. (2017). Estimation under cross-classified sampling with application to a childhood survey. *Journal of the American Statistical Association*, 112, 850-858.

Kozfkay, J.R., and Dillon, J.C. (2010). Creel survey methods to assess catch, loss, and capture frequency of white sturgeon in the Snake River, Idaho. *North American Journal of Fisheries Management*, 30(1), 221-229.

Matei, A., and Tillé, Y. (2005). Maximal and minimal sample co-ordination. *Sankhyā: The Indian Journal of Statistics*, 590-612.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Szoecs, E. and Wagner, H. (2020). *vegan: Community Ecology Package*.

Rivest, L.-P. (2021). Limiting properties of an equiprobable sampling scheme for 0-1 matrices. *Statistics & Probability Letters*, 172, 109047.

Rivest, L.-P., and Ebouele, S.E. (2020). Sampling a two dimensional matrix. *Computational Statistics & Data Analysis*, 149, 106971.

Tillé, Y. (2006). *Sampling Algorithms*. New York: Springer Science & Business Media.

Vallée, A.-A., Ferland-Raymond, B., Rivest, L.-P. and Tillé, Y. (2015). Incorporating spatial and operational constraints in the sampling designs for forest inventories. *Environmetrics*, 26(8), 557-570.