Weighting estimation under bipartite incidence graph sampling

Martina Patone[1] 🟢 · Li-Chun Zhang[1,2,3]

Abstract

Bipartite incidence graph sampling provides a unified representation of many sampling situations for the purpose of estimation, including the existing unconventional sampling methods, such as indirect, network or adaptive cluster sampling, which are not originally described as graph problems. We develop a large class of design-based linear estimators, defined for the sample edges and subjected to a general condition of design unbiasedness. The class contains as special cases the classic Horvitz-Thompson estimator, as well as the other unbiased estimators in the literature of unconventional sampling, which can be traced back to Birnbaum et al. (1965). Our generalisation allows one to devise other unbiased estimators in future, thereby providing a potential of efficiency gains. Illustrations are given for adaptive cluster sampling, line-intercept sampling and simulated graphs.

Keywords Multiplicity estimator · Priority rule · Graph sampling · Ancestral observation procedure · Rao-Blackwell method

1 Introduction

Birnbaum et al. (1965) study the situation where patients are sampled indirectly via the hospitals from which they receive treatment. Insofar as a patient may be treated at more than one hospital, the patients are not nested in the hospitals like elements in clustered sampling.

Birnbaum and Sirken consider three estimators for such *indirect sampling*. The first one is the classic Horvitz-Thompson (HT) estimator (Horvitz and Thompson 1952)

✉ Li-Chun Zhang
  L.Zhang@soton.ac.uk

[1] University of Southampton, Southampton, UK

[2] Statistics Norway, Oslo, Norway

[3] University of Oslo, Oslo, Norway

based on all the sample patients, each of which is weighted by the inverse of the probability of being included in the sample. The second estimator is based on all the sample hospitals and a constructed value for each of them, and the third one is only based on a sub-sample of hospitals determined by a *priority rule*. In particular, the estimator using all the sample hospitals is often referred to as a Hansen-Hurwitz (HH) type estimator. The HH-type estimator and its variations are used for *network sampling* (Sirken 1970, 2005); it is recast as a "generalised weight share method" (Lavalleè 2007); and a modified HH-type estimator is considered for *adaptive cluster sampling* (Thompson 1990, 1991).

All the sampling techniques mentioned above are considered somewhat unconventional, compared to the standard sampling methods using stratification or multistage selection. Unconventional sampling techniques are often characterised by the presence of some rules of observation, *in addition* to the probability design of an *initial sample*. For example, under network sampling (Sirken 1970), "siblings report each other" are needed to reach a "network" of siblings following an initial sample of households. Under adaptive cluster sampling (Thompson 1990), sample propagation depends on the "network" relationship among the units and the values of the surveyed units. Moreover, unconventional sampling require that information of "multiplicity" of sources is collected *in addition to* the sample. For instance, in the example of indirect sampling of patients via hospitals, one needs to identify all the relevant hospitals outside the initial sample, in order to compute the inclusion probability of a sample patient. The same requirement exists as well for any other unconventional sampling, such as "counting rules" of links between population elements and selection units under network sampling (Sirken 2005), or the relationship between edge units and their neighbouring networks under adaptive cluster sampling (Thompson 1990).

Zhang and Patone (2017) formally define sampling from finite graphs, in analogy to sampling from finite populations (Neyman 1934), extending the previous works by Frank (1971, 1980a, 1980b, 2011), which deal with different graph motifs separately. In particular, they show that each of the aforementioned unconventional sampling techniques can be given different graph sampling representations. Zhang and Oguz-Alper (2020) identify sufficient and necessary conditions for feasible representation of sampling from arbitrary graphs as *bipartite incidence graph sampling (BIGS)*, including indirect, network and adaptive cluster sampling. For instance, the nodes can be the hospitals and the patients and an edge exists between a hospital and any patient that receive treatment at the hospital. This is a bipartite graph since the nodes of the graph are bi-partitioned, where an edge can exist only between two nodes in different parts, but not between any two nodes in the same part.

Under graph sampling (Zhang and Patone 2017), one needs to specify an *observation procedure*, by which the edges of the sample graph are observed following an initial sample of nodes. As demonstrated by Zhang and Oguz-Alper (2020), BIGS can provide a unified representation of various situations of sampling, which are originally described in other terms, where one part of the nodes refer to the initial sampling units and the other part the measurement units of interest, to be referred to as *motifs*, such that the edges represent the observational links between sampling units and motifs. More examples will be given later in this paper. Also, the

observation procedure needs to be *ancestral* (Zhang and Patone 2017), so that one knows which other out-of-sample nodes could have led to the motifs in the sample graph, had they been selected in the initial sample of nodes. The information of multiplicity or ancestry is apparent under BIGS, which is simply the knowledge of the nodes (representing sampling units) that are adjacent to the node representing a sample motif in the BIG.

BIGS can provide a unified representation of many so-called unconventional sampling techniques in the literature, and the three estimators considered by (Birnbaum and Sirken 1965) are applicable under any BIGS.

Our aim in this paper is to formulate a large class of unbiased incidence weighting estimators, which includes the three estimators of (Birnbaum and Sirken 1965) as special cases but are not limited to them. This allows one to study design-based estimation under the general setting of *ancestral BIGS* (satisfying the requirement of ancestral observation), where the results are immediately applicable to all the relevant situations. Notice that we do not consider model-based estimation in this paper, which requires additional assumptions but would allow one to draw conclusions about the superpopulation from which the given population graph is taken from.

We shall develop the class of unbiased incidence weighting estimators, based on the sample edges that link the sampling units to the observed motifs. As will be explained, all the three estimators used by (Birnbaum and Sirken 1965) are special cases of this class of estimators, which is an insight hitherto unknown in the literature. Many other unbiased estimators can be devised as members of the proposed class, and one can apply the Rao-Blackwell method (Rao 1945, Blackwell 1947) to the non-HT estimators, to generate distinct unbiased estimators that can improve the estimation efficiency. Thus, the discovery of the class of incidence weighting estimators provides a potential for efficiency gains.

Below, in Sect. 2, we formally introduce ancestral BIGS, and develop the incidence weighting estimators. The general condition of unbiased estimation is established. New understandings of the three aforementioned estimators are discussed. We consider also the application of Rao-Blackwell method, which motivates a new subclass of the HH-type estimators. Illustrations are given in Sect. 3 of adaptive cluster sampling (Thompson 1990), line-intercept sampling (Becker 1991) and simulated graphs, which demonstrate the scope and flexibility of the proposed approach across a variety of situations. Some concluding remarks are given in Sect. 4.

2 Incidence weighting estimator under BIGS

Denote by $\mathcal{B} = (F, \Omega; H)$ a bipartite simple directed graph, where (F, Ω) form a bipartition of the node set $F \cup \Omega$, and each edge in H points from one node in F to another in Ω. No edge exists between any two nodes in F or any two in Ω. For BIGS from $\mathcal{B}$, let F be the set of initial sampling units, and Ω the population of *motifs* that are of interest, where a motif is a subgraph exhibiting a particular pattern, for example a pair of nodes with directed edges to each other, or three nodes forming a triangle in an undirected simple graph. An edge $(i\kappa)$ that is incident to $i \in F$ and

$\kappa \in \Omega$ exists, if and only if the selection of i in a sample s from F leads to the observation of motif κ in Ω, hence the edges (and the graph $\mathcal{B}$) are defined to be directed. The edge set H is unknown to start with. Let the size of F be $M = |F|$, and that of Ω be $N = |\Omega|$, where N is generally unknown. The incidence relationships corresponding to the edges in H represent thus the observational links between the sampling units and the motifs of interest.

Zhang and Oguz-Alper (2020, Theorem 1) establish the sufficient and necessary conditions, by which an arbitrary instance of graph sampling can be given a feasible BIGS representation. They examine and discuss the BIGS representation of indirect, network and adaptive cluster sampling. For instance, for indirect sampling of patients via hospitals, let F consist of the hospitals and Ω the patients, where $(i\kappa) \in H$ iff patient κ receives treatment at hospital i. For network sampling of siblings via households, one can let F consist of the households and Ω the networks of siblings, i. e. each κ represents a group of people who are siblings of each other, where $(i\kappa) \in H$ iff at least one of the siblings in κ belongs to household i. Adaptive cluster sampling will be discussed in Sect. 3.

Let $\alpha_i = \{\kappa : \kappa \in \Omega, (i\kappa) \in H\}$ be the successors of i in $\mathcal{B}$. Given the initial sample s from F, the observation procedure of BIGS is *incident* (Zhang and Patone 2017), such that all the nodes in α_i are included in the sample graph provided $i \in s$; hence, the term BIGS. Let $\Omega_s = \bigcup_{i \in s} \alpha_i$, which consists of all the sample motifs. Following the general definition of sample graph (Zhang and Patone 2017), the *sample BIG* is given by

$$\mathcal{B}_s = (s, \Omega_s; H_s) \qquad \text{where} \quad H_s = (s \times \Omega_s) \cap H$$

is the sample of edges. To be able to calculate the inclusion probabilities of each κ in Ω_s, the observation procedure needs to be *ancestral* as well. Let $\beta_\kappa = \{i : i \in F, (i\kappa) \in H\}$ be the *ancestors* (or predecessors) of κ in $\mathcal{B}$. Let $\beta(\Omega_s) = \bigcup_{\kappa \in \Omega_s} \beta_\kappa$. The knowledge of ancestry (or multiplicity) amounts to the observation of $\beta(\Omega_s) \setminus s$, although these nodes are not part of the sample graph $\mathcal{B}_s$, such as the out-of-sample hospitals of the sample patients.

Example 1 Consider ancestral BIGS from the population BIG below.
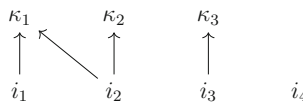
We have $F = \{i_1, i_2, i_3, i_4\}$ and $\Omega = \{\kappa_1, \kappa_2, \kappa_3\}$ and $H = \{(i_1\kappa_1), (i_2\kappa_1), (i_2\kappa_2), (i_3\kappa_3)\}$. Suppose $s = \{i_1, i_3\} \subset F$. By incident observation procedure, we have $\Omega_s = \{\kappa_1, \kappa_3\}$ and $H_s = \{(i_1\kappa_1), (i_3\kappa_3)\}$, and the sample graph $\mathcal{B}_s = (s, \Omega_s; H_s)$ as defined above. In addition, we observe $\beta(\Omega_s) \setminus s = \{i_2\}$, where i_2 is not part of the sample BIG. Notice that the ancestry knowledge requires one to obtain *additionally* the information identifying all the ancestors of all the observed sample motifs in Ω_s. For instance, for each patient κ sampled from the hospital i, all the hospitals (other than i) in which k receives treatments must be identified, whether

or not they are among the actual sample of hospitals. This can e.g. be achieved by adding a survey question to each sample patient in Ω_s, which enumerates all the relevant hospitals. Sometimes, it may be more natural to survey the units in s instead. For instance, when sampling children via their parents, where the mother and father are used as separate sampling units in F, one can ask the in-sample parent about the out-of-sample parent(s). Finally, it may be possible or preferable to retrieve the ancestry knowledge from external sources, such as the Birth Register when sampling children via parents.

Notice also that, in computer science (e.g. Leskovec and Faloutsos 2006; Hu and Lau 2013), one may be concerned with situations where the graph is in principle known but is too large or dynamic to be fully processed or stored for practical purposes. Taking a sub-graph according to some chosen probability scheme is then a possible approach. For an example the whole Twitter graph consisting of users and their following/follower relationships can be constructed by the company at any given time point. However storing every instance of the graph might be unfeasible due to the enormous amount of memory required and the fact that the graph is changing all the time. Taking a sample may suffice for the purpose of estimating e.g. the follower to following ratio. As another example, let F be the products available in an online market place and Ω the buying customers, and let $(i\kappa) \in H$ iff the customer κ has bought the product i. Again, the whole graph is available to the owner of the market, but sampling may be preferred for various market analytics. Of course, in these situations the ancestry knowledge of the sample Ω_s is guaranteed.

Sometimes, either the design or circumstances may prevent one from obtaining the complete ancestry knowledge, such that not all the ancestors β_κ of an observed motif κ are known. Without losing generality, suppose one only manages to obtain information about a subset of β_κ, denoted by β_κ^*, where β_κ^* is non-empty now that κ is already observed. It is then both necessary and possible to modify the sampling strategy (including the estimators described below), an example of which will be discussed in Sect. 3.1 later. Moreover, we refer to Zhang and Oguz-Alper (2020) for a treatment of incomplete ancestry knowledge, which can arise in a number of situations of graph sampling.

2.1 The incidence weighting estimator

Let y_κ be an unknown constant associated with motif κ, for $\kappa \in \Omega$, given the population graph $\mathcal{B}$. The aim is to estimate the total $\theta = \sum_{\kappa \in \Omega} y_\kappa$, including e.g. $y_\kappa \equiv 1$. Let $\delta_i = 1$ or 0 indicates if $i \in s$ or not, so that $\pi_i = \Pr(\delta_i = 1) = E(\delta_i)$, for $i \in F$. Given the sample graph $\mathcal{B}_s$, let $\{W_{i\kappa}; (i\kappa) \in H_s\}$ be the *incidence weights* of the sample edges, and $W_{i\kappa} \equiv 0$ if $(i\kappa) \notin H_s$. The *incidence weighting estimator (IWE)* is given by

$$\hat{\theta} = \sum_{(i\kappa)\in H_s} W_{i\kappa}\frac{y_\kappa}{\pi_{(i\kappa)}} = \sum_{(i\kappa)\in H_s} W_{i\kappa}\frac{y_\kappa}{\pi_i} \tag{1}$$

where $\pi_{(i\kappa)} = \Pr\left((i\kappa) \in H_s\right) = \pi_i$ under BIGS. Notice that the definition (1) allows for sample dependent weights $W_{i\kappa}$.

Proposition 1 *The IWE by (1) is unbiased for θ provided, for each $\kappa \in \Omega$,*

$$\sum_{i\in\beta_\kappa} E(W_{i\kappa}|\delta_i = 1) = 1 \ . \tag{2}$$

Proof The expectation of $\hat{\theta}$ with respect to the sampling distribution of s is given by

$$E(\hat{\theta}) = \sum_{i\in F}\frac{E(\delta_i)}{\pi_i}E\left(\sum_{\kappa\in\alpha_i} W_{i\kappa}y_\kappa|\delta_i = 1\right) = \sum_{\kappa\in\Omega}y_\kappa\sum_{i\in\beta_\kappa}E(W_{i\kappa}|\delta_i = 1) = \theta$$

since $\pi_{(i\kappa)} = \pi_i$ under BIGS, and $\sum_{i\in\beta_\kappa} E(W_{i\kappa}|\delta_i = 1) = 1$ by stipulation. $\square$

The condition (2) ensures that the IWE is unbiased under repeated sampling. When the weights are constant of sampling, denoted by $\omega_{i\kappa}$ for distinction, it reduces to $\sum_{i\in\beta_\kappa} \omega_{i\kappa} = 1$ for any $\kappa \in \Omega$. Let π_{ij} be the second-order sample inclusion probability of $i,j \in F$.

Proposition 2 *The BIG sampling variance of an unbiased IWE is given by*

$$V(\hat{\theta}) = \sum_{\kappa\in\Omega}\sum_{\ell\in\Omega}(\Delta_{\kappa\ell} - 1)y_\kappa y_\ell \tag{3}$$

where

$$\Delta_{\kappa\ell} = \sum_{i\in\beta_\kappa}\sum_{j\in\beta_\ell}\frac{\pi_{ij}}{\pi_i\pi_j}E\left(W_{i\kappa}W_{j\ell}|\delta_i\delta_j = 1\right) \ .$$

Proof Given unbiased $\hat{\theta}$, we have $V(\hat{\theta}) = E(\hat{\theta}^2) - \theta^2$, where

$$E(\hat{\theta}^2) = \sum_{\kappa\in\Omega}\sum_{\ell\in\Omega}y_\kappa y_\ell\sum_{i\in\beta_\kappa}\sum_{j\in\beta_\ell}E\left(\frac{\delta_i\delta_j}{\pi_i\pi_j}W_{i\kappa}W_{j\ell}\right)$$

$$= \sum_{\kappa\in\Omega}\sum_{\ell\in\Omega}y_\kappa y_\ell\sum_{i\in\beta_\kappa}\sum_{j\in\beta_\ell}\frac{\pi_{ij}}{\pi_i\pi_j}E(W_{i\kappa}W_{j\ell}|\delta_i\delta_j = 1)$$

since $W_{i\kappa}W_{j\ell} = 0$ if $\delta_i\delta_j = 0$ under BIGS, for any $(i\kappa), (j\ell) \in H$. The result follows now from taking the difference of $E(\hat{\theta}^2)$ and $\theta^2 = \left(\sum_{\kappa\in\Omega}y_\kappa\right)^2$. $\square$

2.2 HT-type estimator

Let $\pi_{(\kappa)} = \Pr(\kappa \in \Omega_s)$ and $\pi_{(\kappa\ell)} = \Pr(\kappa \in \Omega_s, \ell \in \Omega_s)$ for $\kappa, \ell \in \Omega$, where parentheses are used in the subscript to distinguish these inclusion probabilities of the motifs from those of the sampling units. The HT-estimator is given by

$$\hat{\theta}_y = \sum_{\kappa \in \Omega_s} y_\kappa / \pi_{(\kappa)}$$

where $V(\hat{\theta}_y) = \sum_{\kappa \in \Omega} \sum_{\ell \in \Omega} (\pi_{(\kappa\ell)}/\pi_{(\kappa)}\pi_{(\ell)} - 1)y_\kappa y_\ell$. Under BIGS, we have

$$\pi_{(\kappa)} = 1 - \bar{\pi}_{\beta_\kappa} = 1 - \Pr(\beta_\kappa \cap s = \emptyset)$$
$$\pi_{(\kappa\ell)} = 1 - (\bar{\pi}_{\beta_\kappa} + \bar{\pi}_{\beta_\ell} - \bar{\pi}_{\beta_\kappa \cup \beta_\ell}) \ .$$

where $\bar{\pi}_{\beta_k}$ is the exclusion probability of β_k in s, which is the probability that none of the ancestors of κ in $\mathcal{B}$ is included in the initial sample s, and the knowledge of the out-of-sample ancestors $\beta_\kappa \setminus s$ is required to compute $\bar{\pi}_{\beta_\kappa}$. Similarly for $\bar{\pi}_{\beta_\kappa \cup \beta_\ell}$.

The HT-estimator is a special case of the IWE, where the weights $W_{i\kappa}$ for each k and s satisfy

$$\sum_{i \in s \cap \beta_\kappa} W_{i\kappa}/\pi_i = 1/\pi_{(\kappa)} \ . \tag{4}$$

Notice that these weights $W_{i\kappa}$ are not constant of sampling if $|\beta_\kappa| > 1$, since they depend on how s intersects β_κ. For Example 1 earlier, we have $W_{i_2\kappa_2} = \pi_{i_2}/\pi_{(\kappa_2)}$ and $W_{i_3\kappa_3} = \pi_{i_3}/\pi_{(\kappa_3)}$ by (4), since both κ_2 and κ_3 have only one ancestor in the BIG. Moreover,

$$\begin{cases} W_{i_1\kappa_1} = \pi_{(\kappa_1)}^{-1}\pi_{i_1} & \text{if } s \cap \{i_1, i_2\} = \{i_1\} \\ W_{i_2\kappa_1} = \pi_{(\kappa_1)}^{-1}\pi_{i_2} & \text{if } s \cap \{i_1, i_2\} = \{i_2\} \\ (W_{i_1\kappa_1}, W_{i_2\kappa_1}) = \pi_{(\kappa_1)}^{-1}(a\pi_{i_1}, (1-a)\pi_{i_2}) & \text{if } s \cap \{i_1, i_2\} = \{i_1, i_2\} \end{cases}$$

The value a does not matter, since the coefficient of y_{κ_1} in the IWE (1) is $\sum_{i \in s \cap \beta_\kappa} W_{i\kappa}/\pi_i$.

To see that the weights given by (4) satisfy the condition (2) generally, let ϕ_{s_κ} be the probability that the *sample intersection* is $s_\kappa = s \cap \beta_\kappa$ for $\kappa \in \Omega$, where $\pi_{(\kappa)} = \sum_{s_\kappa} \phi_{s_\kappa}$ over all possible s_κ. Given (4), for any $\kappa \in \Omega$, we have then

$$\sum_{i \in \beta_\kappa} E(W_{i\kappa}|\delta_i = 1) = \sum_{i \in \beta_\kappa} \sum_{s_\kappa \ni i} \frac{\phi_{s_\kappa}}{\pi_i} W_{i\kappa} = \sum_{s_\kappa} \phi_{s_\kappa} \sum_{i \in s_\kappa} \frac{W_{i\kappa}}{\pi_i} = \sum_{s_\kappa} \frac{\phi_{s_\kappa}}{\pi_{(\kappa)}} = 1 \ .$$

Arguing similarly in terms of the joint probability that the sample intersections for κ and ℓ are s_κ and s_ℓ, it can be shown that $\Delta_{\kappa\ell}$ in (3) reduces to $\pi_{(\kappa\ell)}/\pi_{(\kappa)}\pi_{(\ell)}$ given (4) and (2).

More generally, let $\eta_{s_\kappa} = \pi_{(\kappa)} \sum_{i \in s_\kappa} W_{i\kappa}/\pi_i$ for any weights $W_{i\kappa}$ that are not constants of sampling. To satisfy the condition (2), for any $\kappa \in \Omega$, the weights must be such that

$$\sum_{s_\kappa} \phi_{s_\kappa} \eta_{s_\kappa} = \pi_{(\kappa)} \ . \tag{5}$$

The HT-estimator is the special case where $\eta_{s_\kappa} \equiv 1$. It is possible to assign η_{s_κ} that differs from 1 for different sample intersects s_κ, subjected to the restriction (5). Any estimator satisfying (5) but not (4) may be referred to as a *HT-type estimator*.

2.3 HH-type estimator

While a HT-type estimator uses sample dependent weights $W_{i\kappa}$, a HH-type estimator uses weights $\omega_{i\kappa}$ that are constant of sampling. The condition (2) is reduced to $\sum_{i \in \beta_\kappa} \omega_{i\kappa} = 1$, for any $\kappa \in \Omega$. Thus, for Example 1 earlier, we have now $\omega_{i_1\kappa_1} + \omega_{i_2\kappa_1} = 1$ and $\omega_{i_2\kappa_2} = \omega_{i_3\kappa_3} = 1$. Birnbaum and Sirken (1965) observe that

$$\sum_{\kappa \in \Omega} y_\kappa = \sum_{\kappa \in \Omega} y_\kappa \sum_{i \in \beta_\kappa} \omega_{i\kappa} = \sum_{i \in F} \sum_{\kappa \in \alpha_i} \omega_{i\kappa} y_\kappa \ .$$

It follows that the HH-type estimator given by

$$\hat{\theta}_z = \sum_{i \in s} z_i/\pi_i \qquad \text{and} \qquad z_i = \sum_{\kappa \in \alpha_i} \omega_{i\kappa} y_\kappa \tag{6}$$

is unbiased for θ under repeated sampling, where z_i is a constructed constant for each initial sample unit i. The BIG sampling variance of $\hat{\theta}_z$ is given by

$$V(\hat{\theta}) = \sum_{i \in F} \sum_{j \in F} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) z_i z_j \ .$$

Notice that one only needs z_i for the initial sample units in order to apply $\hat{\theta}_z$, which is possible provided ancestral BIGS. Moreover, the HH-type estimator (6) defines actually a family of estimators, depending on the choice of $\omega_{i\kappa}$, although (Birnbaum and Sirken 1965) use only the equal weights $\omega_{i\kappa} = 1/|\beta_\kappa|$. The corresponding $\hat{\theta}_z$ is referred to as the *multiplicity estimator*, denoted by $\hat{\theta}_{z\beta}$. Variations of the multiplicity estimator under other settings of indirect, network sampling are considered by Sirken (1970), Sirken and Levy (1974), Sirken (2004) and Lavallèe (2007). Unlike the HT-estimator, it is in principle possible to apply the Rao-Blackwell method to improve the HH-type estimator, to which we return in Sect. 2.5. Some other HH-type estimators will be discussed then.

2.4 Priority-rule estimator

Birnbaum and Sirken (1965) invent a third estimator based on a *prioritised subset* of H_s, where they let $I_{i\kappa} = 1$ if $i = \min\left(s \cap \beta_\kappa\right)$ and 0 otherwise, i.e. if unit i happens to

be enumerated first in the frame F among all the in-sample ancestors of κ, for each $\kappa \in \Omega_s$. For Example 1 earlier, we have $I_{i_2\kappa_2} = 1$ whenever $\kappa_2 \in \Omega_s$ and $I_{i_3\kappa_3} = 1$ whenever $\kappa_3 \in \Omega_s$, since both κ_2 and κ_3 have only one ancestor in the BIG. The priority-rule only matters for κ_1 here. If $\{i_1, i_2, i_3, i_4\}$ is the frame arranged in the order of enumeration, then we would have $I_{i_1\kappa_1} = 1$ if $i_1 \in s$ whether or not $i_2 \in s$, and $I_{i_2\kappa_1} = 1$ only if $i_2 \in s$ and $i_1 \notin s$. Whereas if $\{i_4, i_3, i_2, i_1\}$ is the frame arranged in the order of enumeration, then we would have $I_{i_2\kappa_1} = 1$ if $i_2 \in s$ whether or not $i_1 \in s$, and $I_{i_1\kappa_2} = 1$ only if $i_1 \in s$ and $i_2 \notin s$.

The *priority-rule estimator* based on $\{(i\kappa) : I_{i\kappa} = 1, (i\kappa) \in H_s\}$ is given by

$$\hat{\theta}_p = \sum_{(i\kappa) \in H_s} \frac{I_{i\kappa}\omega_{i\kappa}y_\kappa}{p_{i\kappa}\pi_i} \tag{7}$$

where $p_{i\kappa} = \text{Pr}\left(I_{i\kappa} = 1 | (i\kappa) \in H_s\right) = \text{Pr}\left(I_{i\kappa} = 1 | \delta_i = 1\right)$ is the conditional probability that $(i\kappa)$ is prioritised given $(i\kappa) \in H_s$, and $\omega_{i\kappa} = 1/|\beta_\kappa|$ are the equal weights for any $\kappa \in \Omega$. Clearly, other priority rules or choices of $\omega_{i\kappa}$ are possible.

One can easily recognise $\hat{\theta}_p$ as a special case of IWE with $W_{i\kappa} = I_{i\kappa}\omega_{i\kappa}/p_{i\kappa}$. It can satisfy the unbiasedness condition (2), *provided $p_{i\kappa} > 0$ for all $(i\kappa) \in H_s$*, in which case $E(W_{i\kappa}|\delta_i = 1) = \omega_{i\kappa}$. Birnbaum and Sirken (1965) did not provide an expression of $V(\hat{\theta}_p)$, but indicated that it is unwieldy. Now that $\hat{\theta}_p$ is a special case of IWE, its variance follows readily from Proposition 2. Let $p_{i\kappa,j\ell} = \text{Pr}\left(I_{i\kappa}I_{j\ell} = 1 | \delta_i\delta_j = 1\right)$, we have

$$\Delta_{\kappa\ell} = \sum_{i \in \beta_\kappa} \sum_{j \in \beta_\ell} \frac{\pi_{ij}p_{i\kappa,j\ell}}{\pi_i\pi_j p_{i\kappa}p_{j\ell}} \omega_{i\kappa}\omega_{j\ell}$$

in (3), such that

$$V(\hat{\theta}_p) = \sum_{(i\kappa) \in H} \sum_{(j\ell) \in H} \left(\frac{\pi_{ij}p_{i\kappa,j\ell}}{\pi_i\pi_j p_{i\kappa}p_{j\ell}} - 1\right)\omega_{i\kappa}\omega_{j\ell}y_\kappa y_\ell$$

because $\sum_{i \in \beta_\kappa} \omega_{i\kappa} = 1$ for any $\kappa \in \Omega$. An unbiased variance estimator can be given by

$$\widehat{V}(\hat{\theta}_p) = \sum_{(i\kappa) \in H_s} \sum_{(j\ell) \in H_s} \left(\frac{\pi_{ij}p_{i\kappa,j\ell}}{\pi_i\pi_j p_{i\kappa}p_{j\ell}} - 1\right)\frac{\omega_{i\kappa}\omega_{j\ell}}{\pi_{ij}}y_\kappa y_\ell \ .$$

The priority probabilities $p_{i\kappa}$ and $p_{i\kappa,j\ell}$ depend on the priority rule, as well as the sampling design. The details for the estimtator of Birnbaum and Sirken (1965) under initial simple random sampling (SRS) without replacement of s are given in Appendix A.

It should be noticed that the priority rule is not part of sampling; the sample graph $\mathcal{B}_s$ includes all the edges incident to every sample unit in s. Had one applied subsampling by randomly selecting one of the edges incident to each i in s with some designed probabilities, the sample graph would have contained one and only one edge from each sample unit. Instead, the priority rule selects only one sample edge incident to each motif in Ω_s for the purpose of estimation.

There is a possibility that a unit i can be sampled but never prioritised, in which case $\hat{\theta}_p$ would be biased. For an extreme example, suppose a motif κ is incident to all the sampling units in F, then the last unit in F can never be prioritised (for κ) according to the priority rule of Birnbaum and Sirken (1965), as long as $|s| > 1$. Generally, $\hat{\theta}_p$ is *biased* under this priority rule, provided there exists at least one motif κ in Ω, where

$$|\beta_\kappa| > 1 \quad \text{and} \quad \Pr\left(|s_\kappa| > 1 \mid \kappa \in \Omega_s\right) = 1$$

such that the ancestor $i = \max(\beta_\kappa)$ has no chance of being prioritised when it is in s. The probability above depends on the ordering of sampling units in F, as well as the initial sample size. Given any ordering of the units in F, as the initial sample increases, it is possible for $\hat{\theta}_p$ to behave more erratically and become biased eventually.

2.5 Using Rao-Blackwell method

The minimal sufficient statistic under BIGS is $\{(\kappa, y_\kappa) : \kappa \in \Omega_s\}$, or simply Ω_s as long as one keeps in mind that the y-values are constants associated with the motifs. Let $\hat{\theta}$ be an unbiased IWE. Applying the Rao-Blackwell method to $\hat{\theta}$ yields $\hat{\theta}_{RB} = E(\hat{\theta}|\Omega_s)$ as an improved estimator, if the conditional variance $V(\hat{\theta}|\Omega_s)$ is positive. Since the HT-estimator $\hat{\theta}_y$ is fixed conditional on Ω_s, we have $\hat{\theta}_{yRB} \equiv \hat{\theta}_y$. For a non-HT estimator, it is in principle possible that the RB method can improve its efficiency, as illustrated below.

Example 2 Consider the BIG in Example 1. Given $|s| = 1$, there are 4 distinct initial samples, leading to 4 distinct Ω_s, such that $V(\hat{\theta}|\Omega_s) = 0$ and $\hat{\theta}_{RB} = \hat{\theta}$ for any unbiased IWE. Given $|s| = 2$, there are 6 different initial samples, leading to 5 distinct Ω_s, where both $s = \{i_1, i_2\}$ and $s' = \{i_2, i_4\}$ lead to the same motifs $\{\kappa_1, \kappa_2\}$, so that $\hat{\theta}_{RB} \neq \hat{\theta}$ given motif sample $\{\kappa_1, \kappa_2\}$, if $\hat{\theta}(s) \neq \hat{\theta}(s')$. Take e.g. the HH-type estimator $\hat{\theta}_z$ by (6), we have

$$\hat{\theta}_z(s) = \frac{\omega_{i_1\kappa_1}}{\pi_{i_1}} y_{\kappa_1} + \frac{\omega_{i_2\kappa_1}}{\pi_{i_2}} y_{\kappa_1} + \frac{\omega_{i_2\kappa_2}}{\pi_{i_2}} y_{\kappa_2} \neq \hat{\theta}_z(s')$$

$$= \frac{\omega_{i_2\kappa_1}}{\pi_{i_2}} y_{\kappa_1} + \frac{\omega_{i_2\kappa_2}}{\pi_{i_2}}$$

$$y_{\kappa_2}\hat{\theta}_{zRB} = \frac{p(s)}{p(s) + p(s')} \cdot \frac{\omega_{i_1\kappa_1}}{\pi_{i_1}} y_{\kappa_1} + \frac{\omega_{i_2\kappa_1}}{\pi_{i_2}} y_{\kappa_1} + \frac{\omega_{i_2\kappa_2}}{\pi_{i_2}} y_{\kappa_2} \ .$$

The calculation required for the RB method may be intractable, if the conditional sample space of s given Ω_s is large and the initial sampling design $p(s)$ is not fully specified, which is common in practice for designs with unequal inclusion probabilities over F. Moreover, the result of RB method is generally not a unique minimum variance unbiased estimator under BIGS, because the minimal sufficient

statistic is not complete. It is thus worth exploring other useful choices of the IWE. Due to the inherent shortcoming of the priority-rule estimator pointed out earlier, we concentrate on the HH-type estimator $\hat{\theta}_z$ below.

Consider the special case where $|\alpha_i| \equiv 1$ in the population BIG, such as when sampling households via persons. Suppose first with-replacement sampling of s, where the different draws generate an IID sample, and compare $\hat{\theta}_y$ and $\hat{\theta}_z$ based on a single draw. Let p_i and $p_{(\kappa)} = \sum_{i \in \beta_\kappa} \pi_i$ be the respective *selection* probabilities. We have $p_{ij} = p_i$ if $i = j$ and 0 if $i \neq j$, and $p_{(\kappa\ell)} = p_{(\kappa)}$ if $\kappa = \ell$ and 0 if otherwise, now that $|\alpha_i| \equiv 1$. We have

$$V\left(\sum_{i \in s} z_i/p_i\right) - V\left(\sum_{\kappa \in \Omega_s} y_\kappa/p_{(\kappa)}\right) = \sum_{\kappa \in \Omega}\left(\sum_{i \in \beta_\kappa} \omega_{i\kappa}^2/p_i - 1/p_{(\kappa)}\right)y_\kappa^2 = 0$$

if $\omega_{i\kappa} \equiv p_i/p_{(\kappa)}$ for $i \in \beta_\kappa$, given which we have $\hat{\theta}_z = \hat{\theta}_{zRB}$. The variance of any other $\hat{\theta}_z$ would be larger, as long as $\omega_{i\kappa}/p_i$ is not a constant over β_κ, because

$$E\left(\frac{z_i}{p_i}|\kappa \in \Omega_s\right) = \sum_{i \in \beta_\kappa}\frac{p_i w_{i\kappa} y_\kappa}{p_{(\kappa)}p_i} = \frac{y_\kappa}{p_{(\kappa)}} \quad \text{and} \quad V\left(\frac{z_i}{p_i}|\kappa \in \Omega_s\right) = y_\kappa^2 V\left(\frac{\omega_{i\kappa}}{p_i}|\kappa \in \Omega_s\right) > 0.$$

A similar argument holds approximately for the choice $\omega_{i\kappa} \propto \pi_i$ under sampling without replacement of s, provided $\pi_{ij} \approx \pi_i \pi_j$ and $\pi_{(\kappa\ell)} \approx \pi_{(\kappa)}\pi_{(\ell)}$, as in the case of sampling households via persons with a small sampling fraction $|s|/|F|$. This can make z_i/π_i more similar to each other over F, which is advantageous with respect to the anticipated mean squared error of $\hat{\theta}_z$ under the sampling design and a population model of z_i, according to Theorem 6.2 of Godambe and Joshi ([1965]).

To make z_i/π_i more similar to each other over F without the restriction $|\alpha_i| = 1$, one may consider setting $\omega_{i\kappa} < \omega_{j\kappa}$ if $|\alpha_i| > |\alpha_j|$, despite $\pi_i = \pi_j$, because there are more motifs contributing to z_i than z_j. Thus, under general unequal-probability sampling of s, it may be reasonable to consider the *probability and inverse-degree adjusted (PIDA)* weights

$$\omega_{i\kappa} \propto \pi_i/|\alpha_i|^\gamma \tag{8}$$

subjected to the condition ([2]), where $\gamma > 0$ is a tuning constant of choice. Denote by $\hat{\theta}_{z\alpha\gamma}$ the corresponding PIDA-IWE. The multiplicity estimator $\hat{\theta}_{z\beta}$ becomes a special case of $\hat{\theta}_{z\alpha\gamma}$ given $\gamma = 0$ and constant π_i over F.

Notice that to apply the weights ([8]) with $\gamma \neq 0$, one needs to know $|\alpha_i|$ for all $i \in \beta_\kappa$ and $\kappa \in \Omega_s$, in addition to the ancestral observation of β_κ. For instance, under indirect sampling of children via parents, one would need to collect the number of children for the out-of-sample parents in $\beta(\Omega_s) \setminus s$ as well. For network sampling of siblings via households, one would need to collect the number of other sibling networks in each household i with at least one member from a sample sibling network κ.

3 Illustrations

3.1 Adaptive cluster sampling

Consider the example of adaptive cluster sampling (ACS) discussed by Thompson (1990). The population F consists of 5 grids, with y-values $\{1, 0, 2, 10, 1000\}$. Each grid has either one or two neighbours which are adjacent in the given sequence, as in the graph G below, where as Thompson (1990) we simply denote each grid by its y-value.

$$G: \qquad 1 \longleftrightarrow 0 \longleftrightarrow 2 \longleftrightarrow 10 \longleftrightarrow 1000$$

Given an initial sample of size 2 by SRS from F, one would survey all the neighbour grids (in both directions if possible) of a sample grid i if y_i exceeds the threshold value 5 but not otherwise. The observation procedure is repeated for all the neighbour grids, which may or may not generate further grids to be surveyed. The process is terminated, when the last observed grids are all below the threshold. The interest is to estimate the total amount of species (or mean per grid) over the given area.

In particular, the grid 2 is a so-called *edge unit*, which can be observed from 10 or 1000, but would not lead to 10 or 1000 if only 2 is selected in s. The inclusion probability of grid 2 under ACS cannot be calculated correctly when it is selected in s but not 10 or 1000, in which case the knowledge of multiplicity (or ancestry) is lacking. Thompson (1990) proposes a modified HT-estimator which uses the grid 2 in estimation, only if it is selected on its own, the probability of which is known from the design of the initial sample.

Zhang and Oguz-Alper (2020) develop feasible BIGS representations of ACS from G above. Here we use one of them to illustrate how the IWE can be applied to ACS. The population BIG is given by $\mathcal{B} = (F, F; H)$, with $\Omega = F$ and H as below.

Zhang and Oguz-Alper (2020) point out that it is possible to consider BIGS from $\mathcal{B}$, where the observational links between (10, 2) and (1000, 2) under ACS are removed to ensure ancestral observation, and apply the classic HT-estimator under this BIGS representation of ACS from G. They show that the two strategies (ACS, modified HT) and (BIGS, HT) actually lead to the same estimator. The difference is that one cannot apply the RB method to the HT-estimator under BIGS, as one can with the modified HT-estimator under ACS. We refer to Zhang and Oguz-Alper (2020) for more details.

Thompson (1990) proposes also a modified HH-type estimator, where an edge unit is used in estimation only if it is selected in s directly. This modified HH-type

estimator is simply the multiplicity estimator $\hat{\theta}_{z\beta}$ under BIGS from $\mathcal{B}$, with equal weights $\omega_{i\kappa} = 1/|\beta_\kappa|$ in (6). The two strategies (ACS, modified HH-type) and (BIGS, $\hat{\theta}_{z\beta}$) lead to the same estimator. Moreover, application of the RB method to $\hat{\theta}_{z\beta}$ is the same as that for the modified HH-type estimator; we refer to Thompson (1990) for the details.

Finally, since the contiguous grids that form a network are all observed together under ACS if any of them is observed, ancestral BIGS from $\mathcal{B}$ entails the observation of $|\alpha_i|$ needed for the PIDA weights given by (8). However, since $|\alpha_i|$ is the same for all the grids in the same network and the initial sampling is SRS, the weights by (8) are all equal in this case, so that the estimator $\hat{\theta}_{z\alpha\gamma}$ coincides with the multiplicity estimator $\hat{\theta}_{z\beta}$.

3.2 Line-intercept sampling

Line-intercept sampling (LIS) is a method of sampling habitats in a region, where a habitat is sampled if a chosen line segment transects it. The habitat may e.g. be animal tracks, roads, forestry, which are of irregular shapes. Kaiser (1983) considers the general situation, where a point is randomly selected on the map and an angle is randomly chosen, yielding a line segment of fixed length or transecting the whole area in the chosen direction. Repetition generates an IID sample of lines. In the simplest setting, each transect line is selected at random by selecting randomly a position along a fixed *baseline* that traverses the whole study area, in the direction perpendicular to the baseline. We apply IWE under BIGS to the following example of LIS (Becker 1991) under this simple setting.

The aim is to estimate the total number of wolverines in the mapped area, as sketched in Fig. 1. Four systematic samples A, B, C and D, each containing 3 positions, are drawn on the baseline that is equally divided into 3 segments of length 12 miles each. Following the 12 selected lines *and* any wolverine track that intercepts them yields 4 observed tracks, denoted by $\kappa = 1, ..., 4$ and heuristically indicated by the dashed lines in Fig. 1. Let y_κ be the associated number of wolverines, and L_κ the length of the projection of κ on the baseline. From top to bottom and left to right, we observe $(y_1, L_1) = (1, 5.25)$, $(y_2, L_2) = (2, 7.5)$, $(y_3, L_3) = (2, 2.4)$ and $(y_4, L_4) = (1, 7.05)$.

3.2.1 Feasible BIGS representation of LIS

First we construct a feasible BIGS representation of LIS in this case. Given the observed tracks, partition the baseline into 7 *projection segments*, each with associated length x_i, for $i = 1, ..., 7$ from left to right, where x_1 refers to the overlapping projection of $\kappa = 1$ and 2, x_2 the projection of $\kappa = 2$ that does not overlap with $\kappa = 1$, x_3 the distance between projections of $\kappa = 2$ and 3, x_4 the projection of $\kappa = 3$, x_5 the distance between projections of $\kappa = 3$ and 4, x_6 the projection of $\kappa = 4$, and x_7 the distance between $\kappa = 4$ and right-hand border. The
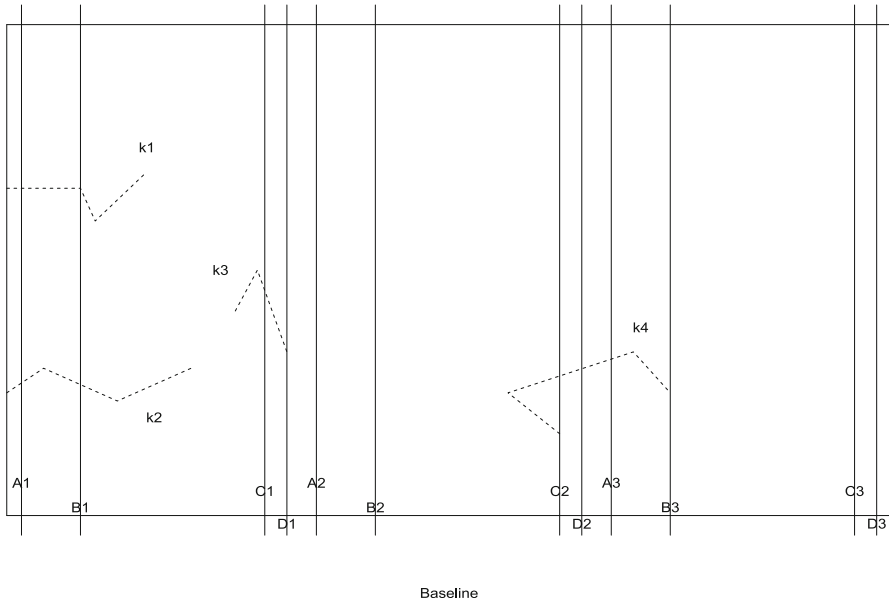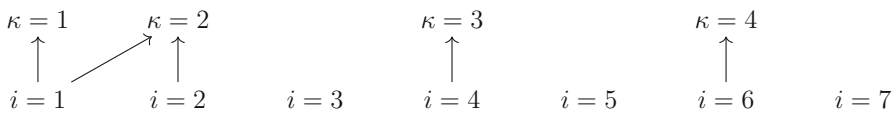
Fig. 1 LIS for wolverinne with 4 systematic samples (A, B, C, D) of 3 positions each

probability that the i-th projection segment is selected by a systematic sample is $p_i = x_i/12$. The 4 systematic samples are IID.

The sample BIG on the r-th draw is given by $\mathcal{B}_r = (s_r, \Omega_r; H_r)$, where s_r contains the selected projection segments, and α_i the wolverine tracks that intercept the sampled line originating from $i \in s_r$, such that $\Omega_r = \bigcup_{i \in s_r} \alpha_i$ and $H_r = \bigcup_{i \in s_r} i \times \alpha_i$. In this example, we have $s_1 = s_2 = \{1, 5, 6\}$, yielding $\Omega_1 = \Omega_2 = \{1, 2, 4\}$ on the first two draws A and B, and $s_3 = s_4 = \{4, 6, 7\}$, yielding $\Omega_3 = \Omega_4 = \{3, 4\}$ on the last two draws C and D. The distinct projection segments selected over all the draws are $s = \bigcup_{r=1}^{4} s_r = \{1, 4, 5, 6, 7\}$, and the distinct tracks are $\Omega_s = \bigcup_{r=1}^{4} \Omega_r = \{1, 2, 3, 4\}$. Let $F^* = \{1, 2, ..., 7\}$ contain the 7 projection segments constructed from (s, Ω_s), and $H_s = \bigcup_{r=1}^{4} H_r$. Let $\mathcal{B}^* = (F^*, \Omega_s; H_s)$ be given as below:

$$
\begin{array}{cccc}
\kappa = 1 & \kappa = 2 & \kappa = 3 & \kappa = 4 \\
\uparrow & \nearrow\ \uparrow & \uparrow & \uparrow \\
i = 1 \quad i = 2 & i = 3 \quad i = 4 & i = 5 \quad i = 6 & i = 7
\end{array}
$$

Let β_κ^* be the ancestors of κ in $\mathcal{B}^*$, where $\beta_1^* = \{1\}$, $\beta_2^* = \{1, 2\}$, $\beta_3^* = \{4\}$ and $\beta_4^* = \{6\}$.

Let $\Omega = \{1, ..., \kappa, ..., N\}$ contain all the wolverine tracks in the area, where $N \geq 4$ given the sample Ω_s. Let $F(\Omega) = \{1, ..., i, ..., M\}$ be the sampling frame, which consists of all the projection segments constructed from Ω. Let $H = \{(i\kappa); i \in F, \kappa \in \Omega\}$, where an edge exists from i to κ provided κ intercepts

any line that originates from the i-th projection segment. The population BIG is given by $\mathcal{B} = (F, \Omega; H)$.

In practice, only $\mathcal{B}^*$ can be constructed but not $\mathcal{B}$. The two are not the same generally, in that one needs to further partition the projection segments of F^* in F based on Ω, in order to accommodate the unobserved tracks in $\Omega \setminus \Omega_s$. For instance, suppose there is a track that can only be intercepted from the 7-th projection segment in F^* and the track does not reach the right-hand border, then this projection segment would be partitioned into 3 segments in F, and (F, H) would differ from (F^*, H^*) accordingly.

Under LIS, field observation along a line has an actual width of detectability. Dividing the baseline accordingly yields thus a known sampling frame F' of detectability partitions. Let $\mathcal{B}' = (F', \Omega; H')$ be the corresponding BIG. By Theorem 1 of Zhang and Oguz-Alper (2020), LIS can be represented as BIGS from $\mathcal{B}'$ where, in particular, the observation procedure of LIS ensures that BIGS from $\mathcal{B}'$ is ancestral for Ω_s.

Now, as along as the unit of detectability is negligible in scale compared to the baseline, one can assume the elements of F' to be nested in those of F^* (or F), such that the selection probability of each observed track κ with respect to BIGS from $\mathcal{B}'$ can be correctly calculated using $\mathcal{B}^*$ (or $\mathcal{B}$). Thus, the strategy BIGS-IWE defined for $\mathcal{B}'$ can be applied using the observed $\mathcal{B}^*$, just as when $\mathcal{B}$ were known.

3.2.2 Estimators

The HT-estimator is given by Thompson (2012, Ch. 19.1). In the present set-up, since the projection segments are non-overlapping, the selection probability of track κ on each draw is given by $p_{(\kappa)}$, where $p_{(1)} = 0.4375$, $p_{(2)} = 0.625$, $p_{(3)} = 0.2$ and $p_{(4)} = 0.5875$. The inclusion probability of $\kappa \in \Omega_s$ is one minus the probability that κ is not selected on any of the 4 draws, $\pi_{(\kappa)} = 1 - (1 - p_{(\kappa)})^4$, where $\pi_{(1)} = 0.90$, $\pi_{(2)} = 0.98$, $\pi_{(3)} = 0.59$ and $\pi_{(4)} = 0.97$. Denote by $p_{(\kappa \cup \ell)}$ the probability of selecting either κ or ℓ on a given draw. The second-order inclusion probability of $\kappa \neq \ell \in \Omega_s$ is given by

$$\pi_{(\kappa \ell)} = \pi_{(\kappa)} + \pi_{(\ell)} - 1 + \left(1 - p_{(\kappa \cup \ell)}\right)^4$$

where $\pi_{(12)} = 0.90$, $\pi_{(13)} = 0.51$, $\pi_{(14)} = 0.88$, $\pi_{(23)} = 0.57$, $\pi_{(24)} = 0.95$ and $\pi_{(34)} = 0.59$. The HT-estimator $\hat{\theta}_y$ and its estimated variance are given in Table 1.

An unbiased estimator of θ from the r-th draw is $\tau_r = \sum_{\kappa \in \Omega_r} y_\kappa / p_{(\kappa)}$, where $\tau_1 = \tau_2 = 7.1878$ and $\tau_3 = \tau_4 = 11.7021$. Becker (1991) uses the HH-type estimator over all the draws:

$$\hat{\theta}_{HH} = \sum_{r=1}^{4} \tau_r / 4 \ .$$

Under BIGS from $\mathcal{B}^*$, let $\hat{\theta}_{z,r} = \sum_{i \in s_r} z_i / p_i$ be the IWE on the r-th draw, given by (6). Given $\omega_{i\kappa} = p_i / p_{(\kappa)}$ for $i \in \beta_\kappa$, where $\omega_{11} = \omega_{43} = \omega_{64} = 1$, $\omega_{12} = p_1 / p_{(2)} = 0.7$ and $\omega_{22} = p_2 / p_{(2)} = 0.3$, we obtain $\hat{\theta}_{HH}$ above as an IWE estimator, since

Table 1 IWE under BIGS from $\mathcal{B}^*$

	$\hat{\theta}_y$	$\hat{\theta}_{z\alpha0}$	$\hat{\theta}_{z\beta}$	$\hat{\theta}_{z\alpha.5}$
Estimate	7.57	9.44	8.99	9.27
Variance	5.27	1.70	2.46	1.97

$$\hat{\theta}_{z,r} = \sum_{i \in s_r} \frac{1}{p_i} \sum_{\kappa \in \alpha_i} \frac{p_i}{P_{(\kappa)}} y_\kappa = \sum_{\kappa \in \Omega_r} \frac{y_\kappa}{P_{(\kappa)}} = \tau_r \ .$$

Since these weights are given by (8) with $\gamma = 0$, we denote $\hat{\theta}_{HH}$ by $\hat{\theta}_{z\alpha0}$ in Table 1.

For the multiplicity estimator $\hat{\theta}_{z\beta}$ with equal weights, we have $\omega_{11} = \omega_{43} = \omega_{64} = 1$, and $\omega_{12} = \omega_{22} = 0.5$. The resulting IWE on each draw are $\hat{\theta}_{z\beta,1} = \hat{\theta}_{z\beta,2} = 6.2736$ and $\hat{\theta}_{z\beta,3} = \hat{\theta}_{z\beta,4} = 11.7021$. The IWE over all the draws is given in Table 1. Next, the unequal weights by (8) can be calculated, since α_i is observed under ancestral BIGS from $\mathcal{B}^*$ given the observation procedure of LIS. Let $\gamma = 0.5$. We have $\omega_{11} = \omega_{43} = \omega_{64} = 1$, $\omega_{12} = 0.6226$ and $\omega_{22} = 0.3773$. The corresponding IWE is 6.8341 on the first two draws, and 11.7021 on the last two draws. This estimator is denoted by $\hat{\theta}_{z\alpha.5}$ in Table 1.

Given the systematic sampling design of the transect lines, the tracks $\{1, 2, 4\}$ can only be observed if a position is selected in the left part of 1st projection segment, which would only result in $\{1, 5, 6\}$ as the sampled projection segments. Similarly, the tracks $\{3, 4\}$ can only be observed if a position is selected in 4th projection segment, which would only result in $\{4, 6, 7\}$ as the sampled projection segments. Thus, applying the RB method would not change any unbiased IWE based on the observed sample BIGS in this case.

The estimator $\hat{\theta}_{HH}$ of Becker (1991) is the IWE $\hat{\theta}_{z\alpha0}$. The HT-estimator $\hat{\theta}_y$ noted by Thompson (2012) can be given as the IWE with weights satisfying (4). Other unbiased IWE can be used for LIS under BIGS from $\mathcal{B}^*$, two of which are as given in Table 1. Neither the HT-estimator $\hat{\theta}_y$ nor the multiplicity estimator $\hat{\theta}_{z\beta}$ is efficient here. Efficiency gains can be achieved using the PIDA weights (8). In this case, adjusting the equal weights by the selection probability while disregarding the degrees of the initial sample units performs well, where $\hat{\theta}_{z\alpha0}$ has the lowest estimated variance. Of course, the true variance of $\hat{\theta}_{z\alpha0}$ may or may not be smaller than that of, say, $\hat{\theta}_{z\alpha.5}$. Meanwhile, setting $\gamma = 1.227$ would numerically reproduce the equal weights $\omega_{12} = \omega_{22} = 0.5$ based on the observed sample. It seems that the IWE by (8) has the potential to approximate the relatively more efficient estimators in different situations, if one is able to choose the coefficient γ in (8) appropriately.

3.3 A simulation study

Two graphs $\mathcal{B} = (F, \Omega; H)$ and $\mathcal{B}' = (F, \Omega; H')$ are constructed for this simulation study. Both $\mathcal{B}$ and $\mathcal{B}'$ have the same node sets F and Ω, and $|F| = 54$ and $|\Omega| = 310$.

The edge sets $|H|$ and $|H'|$ have the size $|H| = |H'| = 1200$, but different distributions of the degree on the motifs in F, as shown in Fig. 2. The distribution of the degree of the motifs in F is relatively uniform over a small range of values in $\mathcal{B}$, but much more skewed and asymmetric in $\mathcal{B}'$.

Let $\theta = |\Omega|$, and $y_\kappa \equiv 1$ for $\kappa \in \Omega$. We consider the following 7 estimators of θ under BIGS from $\mathcal{B}$ or $\mathcal{B}'$ with SRS of s, where $m = |s|$ varies from 2 to 53:

- the IWE $\hat{\theta}_y$ with weights satisfying (4) (the HT estimator);
- the IWE $\hat{\theta}_{z\alpha\gamma}$ with weights satisfying (8) for $\gamma = 0, 1, 2$, (the multiplicity estimator);
- the IWE $\hat{\theta}_p$ by (7) (the priority-rule estimator of Birnbaum and Sirken (1965)). We explore different ordering of the motifs in F: random, ascending or descending yielding three estimators, denoted by $\hat{\theta}_{pR}$, $\hat{\theta}_{pA}$ and $\hat{\theta}_{pD}$, respectively.

Table 2 gives the relative efficiency of the 6 other estimators against the HT-estimator, for a selected set of initial sample sizes, each based on 10000 simulations of BIGS from either $\mathcal{B}$ or $\mathcal{B}'$. All the results are significant with respect to the simulation error.

We notice that all the three priority-rule estimators $\hat{\theta}_{pR}$, $\hat{\theta}_{pA}$ and $\hat{\theta}_{pD}$ are biased when the sample size is large enough. This happens at $m = 45$ for $\mathcal{B}$ and $m = 46$ for $\mathcal{B}'$. Note that the maximum degree of the motifs is 10 in $\mathcal{B}$ and 9 in $\mathcal{B}'$. Moreover, the variance of any priority-rule estimators decreases as the sample size m increases until a threshold value after which the variance starts to increase. In these simulation the threshold is somewhere between 10 and 30.

The sampling variance of the priority-rule estimator is also affected by the ordering of the sampling units in F. The variance tends to be lowest when F is arranged in descending ordering by $|\alpha_i|$, whereas ascending ordering tends to yield the largest variance. Without prioritisation, the value z_i is a constant of sampling given $\omega_{i\kappa}$. Due the randomness induced by the priority-rule, z_i varies over different samples. A sampling unit with large $|\alpha_i|$ has a large range of possible z_i values and placing such a unit towards the end of the ordering tends to increase the sample
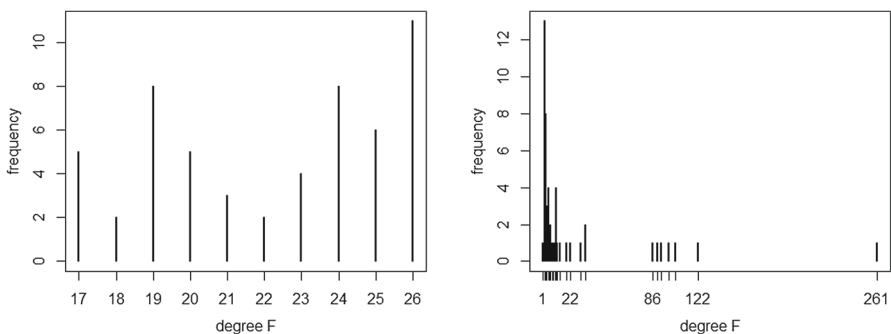

Fig. 2 Distribution of degree $|\alpha_i|$ in $\mathcal{B}$ (left) and $\mathcal{B}'$ (right)

Table 2 Relative efficiency of IWE (against $\hat{\theta}_y$) for $\mathcal{B}$ and $\mathcal{B}'$, 10000 simulations

m	$\mathcal{B}$						$\mathcal{B}'$					
	$\hat{\theta}_{z\alpha0}$	$\hat{\theta}_{z\alpha1}$	$\hat{\theta}_{z\alpha2}$	$\hat{\theta}_{pR}$	$\hat{\theta}_{pA}$	$\hat{\theta}_{pD}$	$\hat{\theta}_{z\alpha0}$	$\hat{\theta}_{z\alpha1}$	$\hat{\theta}_{z\alpha2}$	$\hat{\theta}_{pR}$	$\hat{\theta}_{pA}$	$\hat{\theta}_{pD}$
5	0.96	0.55	0.49	0.80	1.43	0.68	1.22	0.23	0.18	0.97	1.16	0.83
11	0.95	0.55	0.48	0.97	2.57	0.84	1.74	0.33	0.25	0.89	1.54	0.45
17	0.99	0.57	0.51	2.34	4.98	2.57	2.67	0.51	0.39	0.82	2.30	0.24
29	1.31	0.75	0.67	26.7	30.1	33.2	7.96	1.54	1.17	12.0	12.1	29.3

variance of $\{z_i : i \in s\}$ due to prioritisation. It then makes sense that descending ordering by $|\alpha_i|$ may work better than ascending ordering. However, one may not know $\{|\alpha_i| : i \in F\}$ in practice, in which case applying $\hat{\theta}_p$ given whichever ordering of F can be a haphazard business.

Given initial SRS, the different HH-type estimators here differ only with respect to the use of $|\alpha_i|$ in the PIDA weights (8) via the choice of γ. The equal-weights esmator $\hat{\theta}_{z\alpha0}$ is the least efficient of the three HH-type estimators, especially for $\mathcal{B}'$ where the distribution of $|\alpha_i|$ is more skewed. The differences between the other two estimators $\hat{\theta}_{z\alpha1}$ and $\hat{\theta}_{z\alpha2}$ are relatively small, compared to their differences to $\hat{\theta}_{z\alpha0}$, so that a non-optimal choice of $\gamma \neq 0$ is less critical than simply setting $\gamma = 0$. Taken together, these results suggest that the extra effort that may be required to obtain $|\alpha_i|$ is worth considering in practice, and a sensible choice of γ depending on the distribution of $|\alpha_i|$ over F if it is known, or $\mathcal{B}_s$ if it is only observed in the sample BIG, is an interesting question to be studied.

Finally, both $\hat{\theta}_{z\alpha1}$ and $\hat{\theta}_{z\alpha2}$ are more efficient than the HT-estimator when m is small, whereas the HT-estimator improves more quickly as m becomes larger, especially for $\mathcal{B}'$. The matter depends on the sampling fractions $|\Omega_s|/|\Omega|$ and $|s|/|F|$, as well as the respective inclusion probabilities of motifs and sampling units. The interplay between them is complex as it depends on the population BIG. Further research is needed in this respect.

4 Concluding remarks

In this paper we develop a large class of incidence weighting estimators (1) under BIGS. The IWE is applicable to all situations of unconventional sampling techniques that require a specific observation procedure in addition to an initial sample, which can be represented by ancestral BIGS, including indirect, network, adaptive cluster and line-intercept sampling. The condition (2) ensures exactly design-unbiased IWE, which synthesises and generalises the conditions underlying the other unbiased estimators known in the literature.

The classic HT-estimator from finite-population sampling is shown to be a special case of IWE, with any sample dependent weights satisfying the restriction (4), which provides a novel insight. A more general restriction (5) is given for sample dependent weights. It will be intriguing to investigate other HT-type estimators satisfying this restriction.

The priority-rule estimator invented by Birnbaum and Sirken (1965) is another a special case of IWE. However, it may become biased as the initial sample size increases and behave erratically long before that, such that its application may be a haphazard business if one is unable to control the interplay between the ordering of sampling units and the priority-rule of Birnbaum and Sirken (1965). It remains to be seen whether one is able to overcome these shortcomings by future developments.

The HH-type estimators used in the literature are also members of the proposed class. While it is in principle possible to apply the Rao-Blackwell method to an HH-type estimator to improve its efficiency, the computation may be intractable if the conditional sample space of s is large and/or if the initial sampling design $p(s)$ is not fully specified. However, consideration of the Rao-Blackwell method and the degrees (in the BIG) of the sampling units points to the PIDA weights (8) for IWE, as a general alternative to the commonly used equal weights and the corresponding multiplicity estimator. The numerical illustration of line-intercept sampling and the simulation results suggest that the PIDA weights can easily outperform the equal weights. Further study is warranted, in order to identify the sensible choice of the PIDA weights in applications.

Finally, other incidence weights can be explored subjected to the condition (2), beyond those examined in this paper. This is clearly another direction of future research.

Appendix

Priority probabilities of $\hat{\theta}_p$

For each $\kappa \in \Omega_s$ and $i \in \beta_\kappa$, let $d_{i(\kappa)} = \sum_{j \in F : j < i} I_{(j\kappa) \in H}$ be the number of sampling units where higher priority than i under the priority rule $\min(s \cap \beta_\kappa)$. Assume SRS of s, where $m = |s|$. We have

$$p_{i\kappa} = \binom{M - 1 - d_{i(\kappa)}}{m - 1} / \binom{M - 1}{m - 1} .$$

The joint priority probability of $(i\kappa)$ and $(j\ell)$ given $\delta_i \delta_j = 1$ is

$$
p_{i\kappa,j\ell} = \begin{cases} p_{i\kappa} & \text{if } \kappa = \ell, i = j \\ 0 & \text{if } \kappa = \ell, i \neq j \\ \dbinom{M-1-d_{i(\kappa,\ell)}}{m-1} \Big/ \dbinom{M-1}{m-1} & \text{if } \kappa \neq \ell, i = j \\ \dbinom{M-2-d_{i(\kappa)j(\ell)}}{m-2} \Big/ \dbinom{M-2}{m-2} & \text{if } \kappa \neq \ell, i \neq j \text{ and } |\beta_\kappa^i \cap \{j\}| + |\beta_\ell^j \cap \{i\}| = 0 \\ 0 & \text{if } \kappa \neq \ell, i \neq j \text{ and } |\beta_\kappa^i \cap \{j\}| + |\beta_\ell^j \cap \{i\}| > 0 \end{cases}
$$

where β_κ^i is the subset ancestors of κ with higher priority than i, and $d_{i(\kappa,\ell)} = |\beta_\kappa^i \cup \beta_\ell^i|$ is the number of units in $\beta_\kappa \cup \beta_\ell$ with higher priority than i, and $d_{i(\kappa),j(\ell)} = |\beta_\kappa^i \cup \beta_\ell^j|$.

Data Availability The datasets generated during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no conflicts of interest to declare that are relevant to the content of this article.

References

Becker EF (1991) A terrestrial furbearer estimator based on probability sampling. J Wildl Manag 55:730–737

Birnbaum ZW, Sirken MG (1965) Vital and health statistics. Design of sample surveys to estimate the prevalence of rare diseases: three unbiased estimates. Government Printing Office, Washington

Blackwell D (1947) Conditional expectation and unbiased sequential estimation. Ann Math Stat 18:105–110

Frank O (1971) Statistical inference in graphs. Forsvarets forskningsanstalt, Stockholm

Frank O (1980) Estimation of the number of vertices of different degrees in a graph. J Stat Plan Inference 4:45–50

Frank O (1980) Sampling and inference in a population graph. Int Stat Rev 48:33–41

Frank O (2011) The SAGE handbook of social network analysis. Survey sampling in networks. Sage, California, pp 389–403

Godambe VP, Joshi VM (1965) Admissibility and bayes estimation in sampling finite populations. Ann Math Stat 36:1707–1722

Horvitz DG, Thompson DJ (1952) A generalization of sampling without replacement from a finite universe. J Am Stat Assoc 47:663–685

Hu P, Lau WC (2013) A survey and taxonomy of graph sampling. https://arxiv.org/pdf/1308.5865.pdf

Kaiser L (1983) Unbiased estimation in line-intercept sampling. Biometrics 39:965–976

Lavalleè P (2007) Indirect sampling. Springer, Berlin Heidelberg

Leskovec J, Faloutsos C (2006) Sampling from Large Graphs. In proceedings of ACM SIGKDD

Neyman J (1934) On the two different aspects of the representative method: the method of stratified sampling and the method of purposive selection. J R Stat Soc 97:558–625

Rao CR (1945) Information and accuracy attainable in the estimation of statistical parameters. Bull Calcutta Math Soc 37:81–91

Sirken MG (1970) Household surveys with multiplicity. J Am Stat Assoc 65:257–266

Sirken MG, Levy PS (1974) Multiplicity estimation of proportion based on ratios of random variable. J Am Stat Assoc 69:68–73

Sirken MG (2004) Network sample survey of rare and elusive populations: a historical review, innovative methods for surveying difficult-to reach population. In: Preceedings of statistics canada symposium

Sirken MG (2005) In encyclopedia of biostatistics. Network sampling. Wiley, New York

Thompson SK (1990) Adaptive cluster sampling. J Am Stat Assoc 85:1050–1059

Thompson SK (1991) Adaptive cluster sampling: designs with primary and secondary units. Biometics 47:1103–1115

Thompson SK (2012) Sampling. Wiley, New York

Zhang LC, Oguz-Alper M (2020) Bipartite incidence graph sampling. *Submitted*. https://arxiv.org/abs/2003.09467

Zhang LC, Patone M (2017) Graph sampling. Metron 75:277