Instituto Nacional de Estadística

Working Papers
13/2024

On Admissibility in Bipartite Incidence Graph Sampling

Pedro García Segador
Li-Chun Zhang

First draft: October 2024

This draft: October 2024

On Admissibility in Bipartite Incidence Graph Sampling

Abstract

In bipartite incidence graph sampling, the target study units may be formed as connected population elements, which are distinct to the units of sampling and there may exist generally more than one way by which a given study unit can be observed via sampling units. This generalizes finite-population element or multistage sampling, where each element can only be sampled directly or via a single primary sampling unit. We study the admissibility of estimators in bipartite incidence graph sampling and identify other admissible estimators than the classic Horvitz-Thompson estimator. Our admissibility results encompass those for finite-population sampling.

Keywords

graph sampling, admissibility, sufficiency, Rao-Blackwellization

Authors and Affiliations

Pedro García Segador, INE (Spain)

Li-Chun Zhang, University of Southampton (UK) and Statistisk sentralbyra (Norway)

On Admissibility in Bipartite Incidence Graph Sampling

Pedro García-Segador
National Statistics Institute (Spain)

Li-Chun Zhang
University of Southampton, UK
Statistisk sentralbyrå, Norway

October 23, 2024

Abstract

In bipartite incidence graph sampling, the target study units may be formed as connected population elements, which are distinct to the units of sampling and there may exist generally more than one way by which a given study unit can be observed via sampling units. This generalizes finite-population element or multistage sampling, where each element can only be sampled directly or via a single primary sampling unit. We study the admissibility of estimators in bipartite incidence graph sampling and identify other admissible estimators than the classic Horvitz-Thompson estimator. Our admissibility results encompass those for finite-population sampling.

Keywords: graph sampling, admissibility, sufficiency, Rao-Blackwellization

1 Introduction

In the development of finite-population sampling theory since Neyman (1934), the estimator of Horvitz and Thompson (1952) has a central position. Whilst uniformly minimum variance unbiased estimator (UMVUE) does not exist generally (e.g., Godambe, 1955; Hanurav, 1966; Basu, 1971), the Horvitz-Thompson estimator (HTE) is the only UMVUE in the class of linear unbiased estimators for a special type of sampling designs, called unicluster designs. However, unicluster designs (such as systematic sampling) do not satisfy all the needs of efficiency or practicability in applications, and it is not possible to estimate the sampling variance unbiasedly for these designs (Hanurav, 1966).

Godambe (1960) shifts the attention to admissibility as an alternative criterion and proves that the HTE is admissible in the class of linear unbiased estimators. The linearity restriction is removed subsequently (Godambe and Joshi, 1965; Joshi, 1965, 1966), and the admissibility of HTE is extended to the entire class of unbiased estimators. Ramakrishnan (1973) gives an alternative, simpler proof of this result.

In *bipartite incidence graph sampling* (Zhang, 2021a, 2021b), shorthanded as BIGS, the sampling units are formally distinguished from the target study units and there can exist *more than one way* by which a given study unit may be observed via sampling units. In other words, let a bipartite digraph have two sets of nodes representing sampling and study units, respectively, such that directed edges only exist from the first set of nodes to the other, where each edge represents the incidence observation relationship between a sample

unit and a study unit (regardless the actual operations involved or if the relationships may be unknown in advance). For example, sampling 'influencers' on a social media platform via the edges from an initial sample of users is a case of BIGS, where an edge exists from any user to each one she 'follows' and an influencer can be sampled via more than one user. Similarly, if one samples webpages by following the links from other webpages, where the study units are the webpages with links from others and the sampling units are all the webpages.

Such *multiplicity of access* to a given study unit distinguishes BIGS from 'standard' finite-population element or multistage sampling. For instance, in two-stage sampling (e.g., Cochran, 1977), each element can be sampled via one and only one primary sampling unit; similarly for each element called ultimate sampling unit in multistage sampling.

On the one hand, BIGS can extend the scope of estimation by allowing the study units to be formed as groups of population elements, such as social networks of individuals (e.g., Goodman, 1961; Frank, 1971, 1980) or a cluster of neighbouring habitats (Thompson, 1990). On the other hand, it enables a unified treatment (Zhang, 2021b) of so-called 'non-standard' population sampling techniques as special cases of BIGS, such as multiplicity or indirect sampling (Birnbaum and Sirken, 1965; Lavallée, 2007), network sampling (Sirken, 2005) and adaptive cluster sampling (Thompson, 1990), as well as breadth-first or depth-first techniques that are explicitly devised for graph data, such as snowball sampling (Goodman, 1961; Frank and Snijders, 1994; Zhang and Oguz-Alper, 2020; Zhang and Patone, 2017) and random-walk type sampling (e.g., Thompson, 2006; Avrachenkov et al., 2010; Zhang, 2021, 2024).

A large family of unbiased *incidence weighting estimators (IWEs)* has been proposed for BIGS (Zhang, 2021b; Patone and Zhang, 2022), which includes the HTE as a special case and many others. There is thus a need to study admissibility in BIGS, since there may be other admissible estimators than the HTE which may or may not be IWEs.

The paper continues as follows: in Section 2, we introduce formally the notation, BIGS, IWE and other basic concepts. In Section 3, we prove the two main results of this work regarding the admissibility of unbiased estimators in BIGS, which encompass the existing results in finite-population sampling mentioned earlier. A summary is given in Section 4, together with some open problems for future research.

2 Basic concepts

Denote by $\mathcal{B} = (F, \Omega; H)$ a bipartite simple digraph, where F is the non-empty node set of *sampling units*, Ω the non-empty set of *study units* and H the set of *edges* each of which points from a node in F to a node in Ω. Whenever possible, we will denote the elements of F by $\{i_1, i_2, ...\}$ and the elements of Ω by $\{\kappa_1, \kappa_2, ...\}$. Let an initial sample s_0 be taken from F, $s_0 \subseteq F$. The nodes in Ω connected to those in s_0 are called *successors* of s_0, denoted by $\alpha(s_0)$. For any $i \in F$, the set of successors of i in Ω is denoted by α_i. For a subset Λ of Ω, $\Lambda \subseteq \Omega$, the nodes in F connected to Λ are called the *ancestors* of Λ, denoted by $\beta(\Lambda)$, and the set of ancestors of any $\kappa \in \Omega$ is denoted by β_κ.

For BIGS given any initial sample s_0, the study units are given by the *incident observation procedure*, denoted by $\Omega_s = \alpha(s_0)$, together with the corresponding edges emanating from s_0, $H_s = H \cap (s_0 \times \Omega)$. The *sample graph* (from $\mathcal{B}$) is given as $\mathcal{B}_s = (s_0, \Omega_s; H_s)$. Sometimes, we will also use the notation $\Omega_s(s_0)$ to make explicit the choice of s_0. Note that any standard setting of finite-population element or multistage sampling can be represented as BIGS from a graph $\mathcal{B}$ where $|\beta_\kappa| \equiv 1$ for any element κ in the population Ω.

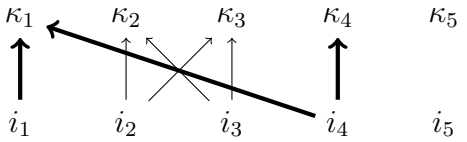

Figure 1: Example of a bipartite graph

Figure 1 provides an illustration. For instance, the successors of i_2 are $\alpha_{i_2} = \{\kappa_2, \kappa_3\}$, and the ancestors of κ_1 are $\beta_{\kappa_1} = \{i_1, i_4\}$. Note that there may exist unconnected elements in F (such as i_5) or Ω (such as κ_5). If the initial sample is $s_0 = \{i_1, i_4\}$ then $\Omega_s = \{\kappa_1, \kappa_4\}$ and the edges H_s of the sample graph $\mathcal{B}_s$ are marked in bold.

Let y_κ be an unknown constant associated with each study unit $\kappa \in \Omega$. Let $y(\Omega) = \{y_\kappa : \kappa \in \Omega\}$, or as a vector $y \in \mathbb{R}^{|\Omega|}$. BIGS from $\mathcal{B}$ allows us to observe $y(\Omega_s) = \{y_\kappa : \kappa \in \Omega_s\}$ associated with the realised sample graph $\mathcal{B}_s$. Let the target of estimation be the total

$$\theta = \sum_{\kappa \in \Omega} y_\kappa.$$

For a given initial sampling design $p(s_0)$ on F, let $\mathcal{S}_0$ be the set of possible initial samples with $p(s_0) > 0$ which is called the *support* of p, such that $\sum_{s_0 \in \mathcal{S}_0} p(s_0) = 1$. Let the initial-sample inclusion probabilities be $\pi_i = \Pr(i \in s_0)$, $\pi_{ij} = \Pr(i \in s_0, j \in s_0)$ and so on, which we distinguish from the study-sample inclusion probabilities written as $\pi_{(\kappa)} = \Pr(\kappa \in \Omega_s)$ or $\pi_{(\kappa\ell)} = \Pr(\kappa \in \Omega_s, \ell \in \Omega_s)$ which are induced from $p(s_0)$ and the given graph $\mathcal{B}$. Due to the multiplicity of access to study units in BIGS, additional knowledge of $\mathcal{B}$ beyond $\mathcal{B}_s$ is needed to calculate $\pi_{(\kappa)}$ or $\pi_{(\kappa\ell)}$. We shall distinguish two levels of knowledge in this paper.

- *Graph knowledge:* when the entire graph $\mathcal{B} = (F, \Omega; H)$ is known.
- *Ancestry knowledge:* when only the ancestor sets $\mathfrak{AK}_s = \{\beta_\kappa : \kappa \in \Omega_s\}$ are known for Ω_s.

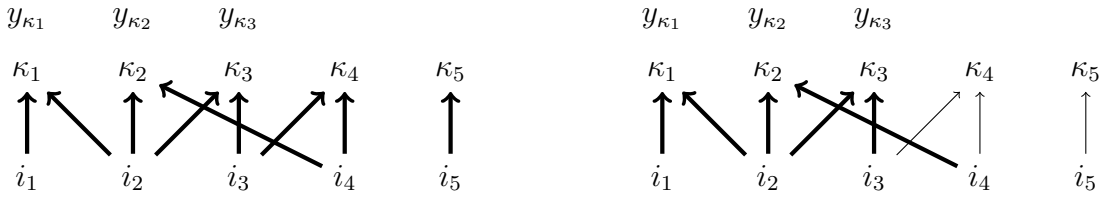

Figure 2: Graph knowledge (left), ancestry knowledge (right) given $s_0 = \{i_1, i_2\}$.

Figure 2 illustrates the differences between the two levels of knowledge, given the initial sample $s_0 = \{i_1, i_2\}$, where the edges highlighted in bold represent the known part of the graph. Note that at least the ancestry knowledge is needed for the calculation of $\pi_{(\kappa)}$ and, consequently, the HTE. Meanwhile, even in situations where the graph knowledge may be available in principle, such as the linked webpages earlier, the graph may be too large or non-static over time such that sampling can be more resource-effective to obtain an estimate of θ than to count θ directly by attempting to process the whole graph.

Now we can give a formal definition of estimator in BIGS.

Definition 1. *An* estimator $e(s_0; y, \mathcal{B})$ *is a real-valued function on* $\mathcal{S}_0 \times \mathbb{R}^{|\Omega|} \times \mathfrak{B}$*, where* $\mathfrak{B}$ *denotes all possible* $\mathcal{B}$ *given* (F, Ω)*, which depends on* y *only through* $\{y_\kappa : \kappa \in \alpha(s_0)\}$ *and*

depends on $\mathcal{B}$ only through $\mathcal{B}_s$ and the given level of knowledge. We may use the simplified notation e, $e(s_0)$ or $e(s_0;\mathcal{B})$ to refer to an estimator when the context is clear.

Notice that, given graph knowledge, the condition that an estimator e only depends on the values of y through the study sample translates to

$$e(s_0; y, \mathcal{B}) = e(s_0; y', \mathcal{B}), \quad \forall y(\Omega_s) = y'(\Omega_s)$$

i.e. for every y' that coincides with y in the realised Ω_s. Whereas, given ancestry knowledge, the conditions that an estimator e must meet would translate to

$$e(s_0; y, \mathcal{B}) = e(s_0; y', \mathcal{B}'), \quad \forall y(\Omega_s) = y'(\Omega_s),\ \mathcal{B}_s = \mathcal{B}'_s,\ \mathfrak{AK}_s = \mathfrak{AK}'_s$$

i.e. for every y' that coincides with y in Ω_s *and* every $\mathcal{B}'$ that coincides with $\mathcal{B}$ in terms of the realised same sample graph and the corresponding ancestor sets of Ω_s.

Next, throughout this work, we shall be working under the condition that every y_κ in Ω has a positive probability to be sampled, which is formally phrased as a property of $\mathcal{B}$.

Definition 2. *A graph $\mathcal{B} = (F, \Omega; H)$ is* covered *by a design $p(s_0)$ on F if*

$$\pi_{(\kappa)} = \Pr(\kappa \in \Omega_s) > 0,\ \forall \kappa \in \Omega.$$

Let us denote by $\mathfrak{B}(p)$ the collection of graphs $\mathcal{B}$ covered by a given $p(s_0)$.

Given $p(s_0)$, we denote by D the class of all unbiased estimators of θ based on BIGS, where an estimator e is unbiased if, over hypothetically repeated sampling,

$$\mathrm{E}\{e(s_0; y, \mathcal{B})\} = \theta$$

for any given $\mathcal{B} \in \mathfrak{B}(p)$ and associated $y \in \mathbb{R}^{|\Omega|}$. A general approach to unbiased estimators is provided by the family of IWEs (Zhang, 2021b), given as

$$\hat{\theta}_{IWE} = \sum_{(i\kappa)\in H_s} W_{i\kappa}\frac{y_\kappa}{\pi_i} = \sum_{\kappa\in\Omega_s}\Big(\sum_{i\in\beta_\kappa\cap s_0}\frac{W_{i\kappa}}{\pi_i}\Big)y_\kappa = \sum_{i\in s_0}\frac{1}{\pi_i}\Big(\sum_{\kappa\in\alpha_i} W_{i\kappa}y_\kappa\Big),$$

where the weights $W_{i\kappa}$ can vary with $\mathcal{B}_s$ generally. Given any $\mathcal{B} \in \mathfrak{B}(p)$, the IWE $\hat{\theta}_{IWE}$ is unbiased iff for each $\kappa \in \Omega$, we have

$$\sum_{i\in\beta_\kappa} \mathrm{E}(W_{i\kappa}|i \in s_0) = 1. \tag{1}$$

(Zhang, 2021b, Theorem 2.1). In particular, if the weight assigned to every $(i\kappa) \in H$ is a constant, regardless of $\mathcal{B}_s$ and denote by $w_{i\kappa}$ for distinction, the IWE is unbiased iff for each $\kappa \in \Omega$, we have

$$\sum_{i\in\beta_\kappa} w_{i\kappa} = 1.$$

Various unbiased estimators in non-standard finite-population sampling applications can be studied as special cases of the IWE; see Zhang (2021b) and Patone and Zhang (2022). The following examples of unbiased IWE will be used in the discussions in this paper. Firstly,

the HTE defined on Ω_s can be written as an IWE, i.e.

$$\hat{\theta}_{HTE} = \sum_{\kappa \in \Omega_s} \frac{y_\kappa}{\pi_{(\kappa)}} = \hat{\theta}_{IWE} \qquad \text{given} \qquad \sum_{i \in s_0 \cap \beta_\kappa} \frac{W_{i\kappa}}{\pi_i} = \frac{1}{\pi_{(\kappa)}} \tag{2}$$

as a condition on the weights $W_{i\kappa}$. Moreover, let $s_\kappa = s_0 \cap \beta_\kappa$ and $\phi_{s_\kappa} = \Pr(s_0 \cap \beta_\kappa = s_\kappa)$, a *HT-type estimator* is given by the weights $W_{i\kappa}$ satisfying

$$\eta_{s_\kappa} = \pi_{(\kappa)} \sum_{i \in s_\kappa} \frac{W_{i\kappa}}{\pi_i} \qquad \text{and} \qquad \sum_{s_\kappa} \phi_{s_\kappa} \eta_{s_\kappa} = \pi_{(\kappa)}$$

which includes the HTE (2) as the special case of $\eta_{s_\kappa} \equiv 1$. Secondly, the so-called *Hansen-Hurwitz type (HH-type) estimator* is an IWE that uses constant weights $w_{i\kappa}$. The earliest example is the *multiplicity estimator* (Birnbaum and Sirken, 1965) using equal weights

$$w_{i\kappa} = |\beta_\kappa|^{-1}. \tag{3}$$

Note that all these IWEs are feasible given the ancestry knowledge of Ω_s.

Let us now give a definition of admissibility in classes of unbiased estimators in BIGS.

Definition 3. *Let $p(s_0)$ be a design on F and $\mathcal{D}$ a class of unbiased estimators, $\mathcal{D} \subseteq D$. An estimator $e_0 \in \mathcal{D}$ is said to be* admissible *in $\mathcal{D}$, if there is no other $e \in \mathcal{D}$ satisfying*

$$\begin{cases} V(e) \leq V(e_0), & \textit{for all } y(\Omega),\ \mathcal{B} \in \mathfrak{B}(p) \\ V(e) < V(e_0), & \textit{for some } y(\Omega),\ \mathcal{B} \in \mathfrak{B}(p) \end{cases}$$

In case an estimator is not admissible, it is said to be inadmissible.

We note immediately that if an estimator e_0 is admissible in $\mathcal{D}$, then there cannot exist another estimator in $\mathcal{D}$ which has the same variance given any $(\mathcal{B}, y)$. The result is given below and its proof in Appendix A (as are all the other proofs in this paper).

Proposition 1. *Let $\mathcal{D} \subseteq D$ be a class of unbiased estimators such that if $e_0, e_0 + d \in \mathcal{D}$, then $e_0 + ad \in \mathcal{D},\ \forall a \in \mathbb{R}$. If e_0 is admissible in the class $\mathcal{D}$, then it is the only estimator $e \in \mathcal{D}$ such that*

$$V(e) \leq V(e_0)\,,\ \forall \mathcal{B} \in \mathfrak{B}(\mathfrak{p}),\ y(\Omega).$$

Note that the hypothesis of $\mathcal{D}$ holds when $\mathcal{D} = D$, since $e_0, e_0 + d \in D \Leftrightarrow \mathrm{E}(d) = 0$, thus $\mathrm{E}(ad) = a\mathrm{E}(d) = 0$ and $e_0 + ad \in D$. All the classes of estimators we will work with in this paper will meet this hypothesis of $\mathcal{D}$.

However, the whole class D is still too broad to establish admissibility *generally*, although we already know that the HTE is admissible in D. Let us therefore introduce an additional property of the estimators that is necessary to our development later.

Let $\mathcal{B} = (F, \Omega; H)$ be a graph and $\Lambda \subseteq \Omega$, we denote by $\mathcal{B}^{(\Lambda)}$ the graph resulting from removing from $\mathcal{B}$ the nodes Λ and any edge in H that ends in Λ. When the set Λ consists of a single element κ, we will use the notation $\mathcal{B}^{(\kappa)}$. Now, in the case $y_\kappa \equiv 0$ for any $\kappa \in \Lambda$, the target total θ is the same in $\mathcal{B}$ and $\mathcal{B}^{(\Lambda)}$, as illustrated in Figure 3, and it seems intuitive that the differences between $\mathcal{B}$ and $\mathcal{B}^{(\Lambda)}$ should not affect the expectation and variance of any potentially admissible estimator of θ. This is formalized in the definition below, and the potential loss of efficiency of non-zero-invariant estimators will be demonstrated later.

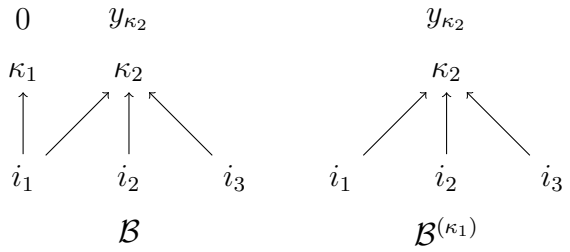

Figure 3: A zero-invariant estimator is the same given $\mathcal{B}$ or $\mathcal{B}^{(\kappa_1)}$.

Definition 4. *An estimator e is* zero-invariant *if for any $s_0 \in \mathcal{S}_0$, we have*

$$e(s_0; y, \mathcal{B}) = e(s_0; y', \mathcal{B}^{(\Lambda)})$$

for any $y(\Omega)$ and $y'(\Omega \setminus \Lambda)$ satisfying $y_\kappa = 0$ if $\kappa \in \Lambda$ and $y_\kappa = y'_\kappa$ if $\kappa \in \Omega \setminus \Lambda$. In the special case of empty Ω_s, $\Omega_s = \emptyset$, we will let the estimator take the value zero. Denote by D^* the class of unbiased zero-invariant estimators.

Note that the HTE is zero-invariant, since $y_\kappa/\pi_{(\kappa)} = 0$ if $y_\kappa = 0$ regardless $\pi_{(\kappa)}$. In this work we study admissibility for classes of unbiased zero-invariant estimators.

3 Admissibility in classes of unbiased estimators

Below we consider admissibility given either graph knowledge or ancestry knowledge.

3.1 Given graph knowledge

Given the graph knowledge, one can determine the study-sample $\Omega_s(s_0)$ following any initial sample s_0, so as to apply Rao-Blackwellization (Rao, 1945; Blackwell, 1947) and the concept of sufficiency (Godambe and Joshi, 1965). Indeed, we shall prove below that any sufficient estimator is admissible in the class D^* of unbiased zero-invariant estimators, and any such estimator can be obtained by a modified Rao-Blackwellization procedure.

Recall that in finite population sampling by $p(s)$, yielding sample s from population U, the minimal sufficient statistic is the unordered sample $\{(i, y_i) : i \in s\}$, with all $\{y_i : i \in U\}$ as the unknown parameters. In BIGS that yields the study sample $\Omega_s(s_0)$, denote the minimal sufficient statistic by

$$D_{s_0} := \{(\kappa, y_\kappa) : \kappa \in \Omega_s(s_0)\}.$$

We shall characterise an estimator as *sufficient* by the following result without proof.

Proposition 2. *In BIGS, an estimator e is* sufficient *for y given $p(s_0)$ and any $\mathcal{B}$, if*

$$e(s_0) = e(s'_0)$$

for $s_0 \neq s'_0$ as long as $\Omega_s(s_0) = \Omega_s(s'_0)$.

Notice that any sufficient estimator can as well be written as $e(\Omega_s)$ with Ω_s as the *distinct* images of all the possible initial samples, including $\Omega_s = \emptyset$, where the sampling probability

of Ω_s follows from $p(s_0)$, denoted by

$$P(\Omega_s) = \Pr\{\alpha(s_0) = \Omega_s\} = \sum_{s_0:\alpha(s_0)=\Omega_s} p(s_0)$$

Clearly, given any estimator e, one can apply Rao-Blackwellization to obtain a sufficient estimator, which is given as

$$e_{RB} = \mathrm{E}(e \mid D_{s_0}) = \sum_{s_0':\alpha(s_0')=\Omega_s(s_0)} \frac{p(s_0')e(s_0')}{P(\Omega_s(s_0))}$$

and we would obtain $e_{RB} = e$ if the estimator e is sufficient, such as the HTE.

Although Rao-Blackwellization does not yield UMVUE because D_{s_0} is not complete in BIGS (nor in sampling from Ω directly), we shall prove below that it is closely related to admissibility. However, it is necessary to modify the procedure since e_{RB} is not zero-invariant generally, which can have undesirable consequences, as the example below illustrates.

Example 1. *Consider again the graphs in Figure 3. Let there be two possible initial samples* $s_0^* = \{i_1, i_2\}$ *and* $s_0^{**} = \{i_2, i_3\}$*, where* $p(s_0^*) = 1/3$ *and* $p(s_0^{**}) = 2/3$*. We have* $\pi_1 = 1/3$*,* $\pi_2 = 1$ *and* $\pi_3 = 2/3$*. Let* e *be the multiplicity estimator* (3)*, where*

$$\begin{cases} e(s_0^*; y', \mathcal{B}^{(\kappa_1)}) = (\pi_1^{-1} + \pi_2^{-1})\frac{y_{\kappa_2}}{3} = \frac{4}{3}y_{\kappa_2} \\ e(s_0^{**}; y', \mathcal{B}^{(\kappa_1)}) = (\pi_2^{-1} + \pi_3^{-1})\frac{y_{\kappa_2}}{3} = \frac{5}{6}y_{\kappa_2} \end{cases} \quad \textit{or} \quad \begin{cases} e(s_0^*; y, \mathcal{B}) = 3y_{\kappa_1} + \frac{4}{3}y_{\kappa_2} \\ e(s_0^{**}; y, \mathcal{B}) = \frac{5}{6}y_{\kappa_2} \end{cases}$$

given $\mathcal{B}^{(\kappa_1)}$ *or* $\mathcal{B}$*. The corresponding* e_{RB} *is*

$$e_{RB}(s_0^*; y', \mathcal{B}^{(\kappa_1)}) = e_{RB}(s_0^{**}; y', \mathcal{B}^{(\kappa_1)}) = y_{\kappa_2}$$

given $\mathcal{B}^{(\kappa_1)}$*; whereas*

$$e_{RB}(s_0^*; y, \mathcal{B}) = e(s_0^*; y, \mathcal{B}) \not\equiv e(s_0^{**}; y, \mathcal{B}) = e_{RB}(s_0^{**}; y, \mathcal{B})$$

because $\alpha(s_0^*) \neq \alpha(s_0^{**})$ *given* $\mathcal{B}$*. Thus, given* $y_{\kappa_1} = 0$ *and* $\mathcal{B}$ *instead of* $\mathcal{B}^{(\kappa_1)}$*, the variation of* e_{RB} *over* s_0 *causes extra variance unnecessarily. Note that being zero-invariant, the HTE is always equal to* y_{κ_2} *here given* $\mathcal{B}^{(\kappa_1)}$ *or* $\mathcal{B}$*.*

Let use therefore modify Rao-Blackwellization so that it becomes zero-invariant.

Definition 5. *The* zero-invariant Rao-Blackwell (ZRB) *estimator derived from any given unbiased estimator* e *is*

$$e_{ZRB}(s_0) \coloneqq ZRB(e) = \sum_{s_0' \in [s_0]} \frac{p(s_0')e(s_0')}{P\left(\Omega_s(s_0) \cap \Omega_{(0)}\right)}$$

where $\Omega_{(0)} = \{\kappa \in \Omega : y_\kappa \neq 0\}$*, and* $\Omega_s(s_0) \cap \Omega_{(0)}$ *contains only the non-zero study units in the realised sample* $\Omega_s(s_0)$*, and* $[s_0] = \{s_0' : \alpha(s_0') \cap \Omega_{(0)} = \Omega_s(s_0) \cap \Omega_{(0)}\}$*.*

By definition, e_{ZRB} is zero-invariant if e is zero-invariant, and it is unbiased if e unbiased. Moreover, e_{ZRB} has a smaller mean squared error if e is not constant over $[s_0]$, as illustrated with $\mathcal{B}$ in Example 1 above. The operation, denoted by ZRB(e), achieves thus the intended

effect of Rao-Blackwellization in BIGS. The results are formalized below, where the theorem is the main result of admissibility given graph knowledge.

Lemma 1. *Let p be a design and any $\mathcal{S}_0^* \subseteq \mathcal{S}_0$. The minimum of $\sum_{s_0 \in \mathcal{S}_0^*} p(s_0)e^2(s_0)$ subjected to the constraint $\sum_{s_0 \in \mathcal{S}_0^*} p(s_0)e(s_0) = Y^*$ is only attained at*

$$e(s_0) = \frac{Y^*}{\sum_{s_0 \in \mathcal{S}_0^*} p(s_0)}.$$

Proposition 3. *Let e be an unbiased zero-invariant estimator in BIGS. The corresponding e_{RB} is inadmissible in D, since*

$$V(e_{ZRB}) \leq V(e_{RB}) \leq V(e),$$

where the first inequality is strict for some graph $\mathcal{B}$ and vector y.

Theorem 1. *Assume graph knowledge. Let $p(s_0)$ be a design, where $\pi_i > 0$ for any $i \in F$ and e_0 an estimator in the class D^*. Then the following statements are equivalent:*

i) e_0 is sufficient;

ii) e_0 is admissible in the class D^*;

iii) e_0 is given by ZRB(e) for some estimator $e \in D^*$.

We note that the graph knowledge is needed for ZRB(e) but not for the proof of (i) and (ii) in Theorem 1. For instance, the sufficient, zero-invariant and admissible HTE requires only the ancestry knowledge. Moreover, now that BIGS encompasses finite-population sampling, Theorem 1 applies directly to all the non-standard population sampling problems, which clarifies the connection of admissibility to sufficiency and Rao-Blackwellization.

For example, Thompson (1990) proposes the *modified HTE* for adaptive cluster sampling and applies Rao-Blackwellization to it, where the units with y-values below the fixed adaptive threshold are used in the estimator *only when* they are *directly selected* in the initial sample, not otherwise when they are observed via the units with y-values above the threshold. The modified HTE is thus zero-invariant but not sufficient, which is why it can be modified by Rao-Blackwellization. Theorem 1 shows that this indeed yields an admissible estimator in the class D^* for adaptive cluster sampling; indeed, applying ZRB to any unbiased zero-invariant estimator would yield an admissible estimator for adaptive cluster sampling.

Returning to any unbiased IWE estimator that is zero-invariant due to its linearity in y, applying ZRB to it would yield an admissible estimator in D^*. The admissible HTE is a special case that is zero-invariant and sufficient, such that it is unchanged by ZRB.

Corollary 1. *Assume graph knowledge. Let $p(s_0)$ be a design, where $\pi_i > 0$ for any $i \in F$. The estimator e_{ZRB} derived from some unbiased IWE e is admissible in the class D^*.*

A final remark on the assumed graph knowledge is worthwhile. While the condition is easy to state and sufficient for Theorem 1, what is actually needed is the part of the graph that enables one to apply ZRB to the *realised* study sample Ω_s. It is then not necessary to know the whole graph, but it suffices to know all the successors of any node in the ancestor set $\mathfrak{AR}_s$. One may refer to this level of knowledge as the *successor-ancestry knowledge*, which can replace the assumption of graph knowledge in order to apply Theorem 1. Figure 4 illustrates the different levels of knowledge when the initial sample is $s_0 = \{i_1, i_2\}$.

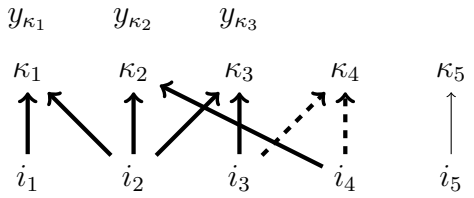

Figure 4: Ancestry knowledge (bold) and successor-ancestry knowledge (bold and dashed), given $s_0 = \{i_1, i_2\}$. Graph knowledge comprises all edges (see also Figure 2) regardless s_0.

3.2 Given ancestry knowledge

When we only have the ancestry knowledge, the HTE estimator is still admissible because it is sufficient and zero-invariant. However, for other estimators that are not sufficient, we can no longer invoke ZRB as by Theorem 1. We therefore need to identify some additional properties of the estimators which are shared by the HTE but not limited to it.

The first additional property we require for the main result to be developed concerns analytic functions. A mathematical function is analytic at a point if its local behavior can be expressed as an infinite sum of powers (i.e. its Taylor series) converging in a region around that point. An important property of nonzero analytic functions $f(x)$ is that their zeros are isolated, which means there is a neighborhood around each zero x_0, $f(x_0) = 0$, where the function does not vanish except at x_0.

We shall call an estimator *analytic* if it is an analytic function of $y(\Omega_s)$. For example, every IWE estimator is linear and thus analytic, whereas ZRB(e) is not generally continuous at zero and therefore is not analytic. Below we will denote by D^{**} *the class of unbiased, zero-invariant and analytic estimators* and study admissibility in D^{**}.

Next, we will call an estimator *elemental* if it takes the same expression as the HTE for a special type of graphs, called elemental graphs, as formalised below.

Definition 6. *In BIGS, an estimator e is* elemental *if $e(s_0, \mathcal{B}) = y_\kappa/\pi_i$ whenever $i \in s_0$, given any* elemental *graph $\mathcal{B} = (F, \Omega; H)$ with $\Omega = \{\kappa\}$ and $H = \{(i, \kappa)\}$.*

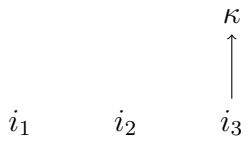

Figure 5: Example of elemental graph.

Figure 5 shows an example of elemental graph. We would have $e(s_0; y, \mathcal{B}) = y_\kappa/\pi_3$ given any sample s_0 containing i_3. Note that if we assume an estimator e is linear, elemental and zero-invariant, then for any graph where $\beta_\kappa = 1$, $\forall \kappa \in \Omega$, this estimator will match the HTE in BIGS, which will also coincide with the HTE by sampling from Ω directly. Being elemental is thus a property of the HTE which may be shared by other estimators such as the multiplicity estimator.

Next, we will invoke some linear algebra concepts that are relevant to elemental estimators and the proof of the main result later.

Definition 7. *Let the* sample-space matrix $\mathbb{S}^p$ *have rows corresponding to the sampling units F and columns to all the possible samples in the support of $p(s_0)$, $s_0 \in \mathcal{S}_0$. The element for*

row i and column s_0 is $\mathbb{S}^p_{i,s_0} = 1$ if $i \in s_0$ and 0 otherwise. Likewise, we define the matrix $\tilde{\mathbb{S}}^p$ whose element is $\tilde{\mathbb{S}}^p_{i,s_0} = p(s_0)$ if $i \in s_0$ and 0 otherwise.

Definition 8. *An estimator e is* sample-space spanned *if there exist functions a_i that do not depend on (i.e. vary with) the sample s_0 such that:*

$$e(s_0; y, \mathcal{B}) = \sum_{i \in s_0} a_i(y, \mathcal{B}), \quad \forall s_0 \in \mathcal{S}_0.$$

Proposition 4. *Let $e \in D^*$. If e is sample-space spanned then e is elemental.*

Now, given a linear map defined by matrix A, the *kernel* of A, $\mathrm{Ker}(A)$, is the vector space formed by all the vectors v, such that $A \cdot v = 0$. The *row space*, $\mathrm{Row}(A)$, is the linear space spanned by the row vectors of A. By definition, an estimator e is sample-space spanned if and only if $e \in \mathrm{Row}(\mathbb{S}^p)$ given any y and $\mathcal{B}$. Note that $\mathrm{Row}(A)$ is orthogonal to $\mathrm{Ker}(A)$, denoted by $\mathrm{Ker}(A)^\perp = \mathrm{Row}(A)$. Moreover, let $v \cdot_p w = \sum_i p_i v_i w_i$ be the p-weighted scalar product of vectors v and w, and $W^{\perp_p}$ the orthogonal complement of a vector space W, such that $v \cdot_p w = 0$ for any $v \in W^{\perp_p}$, $w \in W$. We can now state the following result.

Lemma 2. *Let $p(s_0)$ be a design with associated matrices $\mathbb{S}^p$ and $\tilde{\mathbb{S}}^p$. We have*

$$Ker(\tilde{\mathbb{S}}^p)^{\perp_p} = Row(\mathbb{S}^p).$$

It follows from Lemma 2 that

$$\mathrm{Row}(\mathbb{S}^p) \cap \mathrm{Ker}(\tilde{\mathbb{S}}^p) = \mathbf{0} \quad \text{and} \quad \mathbb{R}^{|\mathcal{S}_0|} = \mathrm{Row}(\mathbb{S}^p) \oplus \mathrm{Ker}(\tilde{\mathbb{S}}^p) \tag{4}$$

where $\mathbf{0}$ is the null-vector and $\oplus$ denotes direct sum. However, the decomposition (4) of vector space $\mathbb{R}^{|\mathcal{S}_0|}$ does not imply that every estimator in BIGS can be given as the sum of an estimator that is sample-space spanned and another that lies in the kernel. To make this distinction we introduce the next definition.

Definition 9. *Let $p(s_0)$ be a design and $V \subseteq \mathbb{R}^{|\mathcal{S}_0|}$ a vector space. We define the associated* estimator space*, $\mathfrak{E}(V)$, as the set of all estimators e such that $e(s_0; y, \mathcal{B}) \in V$, $\forall (y, \mathcal{B})$. Denote by $\mathfrak{E}_c(V)$ the subset of estimators in $\mathfrak{E}(V)$ whose expectation is c over sampling.*

If $V \cap W = \mathbf{0}$, then $\mathfrak{E}(V) \oplus \mathfrak{E}(W)$ contains the estimators that can be uniquely expressed as the sum of an estimator in V and another in W. In particular, $\mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big)$ is just the set of unbiased sample-space spanned estimators of θ, and $\mathfrak{E}_0\big(\mathrm{Ker}(\tilde{\mathbb{S}}^p)\big)$ is the set of estimators d in the specified kernel space satisfying $\mathrm{E}(d) = 0$. Their relevance to delineating admissibility is clarified in the following proposition and used in the proof of the theorem next, which states that every estimator in $\mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big)$ is admissible in D^{**}.

Proposition 5. *Let $p(s_0)$ be a design and $e = e_0 + d$ an estimator, where $e_0 \in \mathfrak{E}_\theta\big(Row(\mathbb{S}^p)\big)$ and $d \in \mathfrak{E}_0\big(Ker(\tilde{\mathbb{S}}^p)\big)$. Then $V(e_0) \leq V(e)$ for every $(y, \mathcal{B})$, with strict inequality whenever $d \not\equiv 0$ for given $(y, \mathcal{B})$, i.e., if d is not constant 0. In particular, if $d \not\equiv 0$, e is inadmissible.*

Theorem 2. *Assume ancestry knowledge in BIGS and let p be a design for which $\pi_i > 0$, $\forall i \in F$. If e_0 in D^{**} is a sample-space spanned estimator, i.e. $e_0 \in \mathfrak{E}_\theta\big(Row(\mathbb{S}^p)\big)$, then e_0 is admissible in the class D^{**}.*

Note that sample-space spanned estimators can be seen as a special case of estimators in $\mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big) \oplus \mathfrak{E}_0\big(\mathrm{Ker}(\tilde{\mathbb{S}}^p)\big)$ with the trivial kernel element $d \equiv 0$. Hence, we can summarize the results obtained so far for the estimators in D^{**} as follows:

- any estimator e_0 in $\mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big)$ is admissible in D^{**}, whereas any estimator $e = e_0 + d$ with $e_0 \in \mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big)$ and non-trivial $d \in \mathfrak{E}_0\big(\mathrm{Ker}(\tilde{\mathbb{S}}^p)\big)$ is inadmissible in D^{**};
- the admissibility of estimator e is unknown generally if $e \notin \mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big) \oplus \mathfrak{E}_0\big(\mathrm{Ker}(\tilde{\mathbb{S}}^p)\big)$.

In particular, any IWE with constant weights $w_{i\kappa}$ is a linear (hence, analytic) sample-space spanned estimator, which can be given as $\hat{\theta}_{IWE} = \sum_{i\in s_0} a_i(y, \mathcal{B})$ where

$$a_i(y, \mathcal{B}) = \sum_{\kappa\in\alpha_i} w_{i\kappa} y_\kappa / \pi_i.$$

Any such unbiased IWE is admissible in D^{**} as summarised below without proof.

Corollary 2. *Assume ancestry knowledge and let $p(s_0)$ be a design for which $\pi_i > 0, \forall i \in F$. An unbiased IWE with constant weights $w_{i\kappa}$ is admissible in the class D^{**}.*

Let us close this section with two further remarks. First, although the admissibility of estimators that lie outside of $\mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big) \oplus \mathfrak{E}_0\big(\mathrm{Ker}(\tilde{\mathbb{S}}^p)\big)$ is yet unknown generally, it is still possible to obtain specific admissibility results by making restrictions on the designs $p(s_0)$, similarly to the aforementioned result that the HTE is the only UMVUE given unicluster designs in finite-population sampling. Below we give such a result.

Definition 10. *A design $p(s_0)$ over F is said to be* full rank *if $\mathbb{S}^p$ is a full rank matrix, that is, if rank$(\mathbb{S}^p) = |\mathcal{S}_0|$.*

It is not difficult to verify that, for designs with fixed sample size $|s_0|$, all minimum support designs (e.g. Tillé, 2006) are full rank. Moreover, systematic sampling designs are full rank, but the simple random sampling design is not full rank. It now follows from Theorem 2 that every estimator in D^{**} is admissible given full-rank designs.

Corollary 3. *Assume ancestry knowledge and let $p(s_0)$ be a full-rank design for which $\pi_i > 0$, $\forall i \in F$. Every estimator in D^{**} is admissible in the class D^{**}.*

Second, the construction of Proposition 5 to inadmissible estimators can be generalized to orthogonal pairs of estimator spaces other than $\mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big) \oplus \mathfrak{E}_0\big(\mathrm{Ker}(\tilde{\mathbb{S}}^p)\big)$, denoted by $\mathfrak{E}_\theta(V) \oplus \mathfrak{E}_0(W)$ given $W = V^{\perp_p}$, where $\mathfrak{E}_\theta(V)$ is a space of unbiased estimators of θ and $\mathfrak{E}_0(W)$ its orthogonal space of mean-zero estimators. Below we give such an example.

Example 2. *Consider simple random sampling design $p(s_0) \equiv 1/6$ with $|s_0| = 2$ given the graph in Figure 1, where we remove the disconnected elements i_5 and κ_5. Let an IWE e_0 with non-constant weights $W_{i\kappa}$ be given as follows. Denote by $12 < 13 < 14 < 23 < 24 < 34$ the lexicographical order among the 6 samples in $\mathcal{S}_0$, based on which we define the weights $W_{i\kappa}$ and e_0 to be*

$$W_{i\kappa}(s_0) = \frac{w_{i\kappa}\pi_i I_i(s_0)}{p(s_0)} \quad \textit{and} \quad e_0 = \sum_{\kappa\in\Omega_s} \left(\sum_{i\in\beta_\kappa\cap s_0} \frac{w_{i\kappa} I_i(s_0)}{p(s_0)} \right) y_\kappa$$

where $w_{i\kappa}$ can be any constant weights satisfying $\sum_{i\in\beta_\kappa} w_{i\kappa} = 1$, $\forall\kappa$, and $I_i(s_0)$ is an indicator that s_0 is the first sample in the lexicographical order where $s_0 \ni i$. Since $I_i(s_0) = 1$ only for one particular s_0 and $I_i(s_0) = 0$ otherwise, each y_κ is used for estimation given only one particular s_0 as well; hence the expression of e_0. The estimator e_0 is unbiased because

$$\sum_{i\in\beta_\kappa} E(W_{i\kappa}|i \in s_0) = \sum_{i\in\beta_\kappa}\sum_{s_0\ni i} \frac{p(s_0)W_{i\kappa}}{\pi_i} = \sum_{i\in\beta_\kappa} w_{i\kappa} \sum_{s_0\ni i} I_i(s_0) = \sum_{i\in\beta_\kappa} w_{i\kappa} = 1.$$

Note that $e_0(s_0) = 0$ when $s_0 = \{i_2, i_3\}$, $\{i_2, i_4\}$ or $\{i_3, i_4\}$, where $I_i(s_0) = 0$ for any $i \in s_0$. In other words, the estimator e_0 resides in the vector space

$$V = \{(a, b, c, 0, 0, 0) : a, b, c \in \mathbb{R}\}.$$

Next, to construct a mean-zero estimator d that belongs to the orthogonal vector space, let us repeat the construction of e_0 with different ordered samples. Let e_1 be given as e_0 except that $I_i(s_0)$ is based on the reverse lexicographical order $34 < 24 < 23 < 14 < 13 < 12$; and let e_2 be based on the order $24 < 34 < 23 < 14 < 13 < 12$. As both e_1 and e_2 are unbiased, their difference $d = e_2 - e_1$ has zero mean, which resides in the vector space

$$W = \{(0, 0, 0, 0, f, g) : f, g \in \mathbb{R}\}.$$

Table 1 lists the estimates e_0, e_1, e_2 and d given each of the 6 samples s_0, where we use the multiplicity weights $w_{i\kappa} = 1/|\beta_\kappa|$ for this illustration.

Table 1: Estimators e_0, e_1, e_2 and d given $w_{i\kappa} = 1/|\beta_\kappa|$.

s_0	$e_0(s_0)$	$e_1(s_0)$	$e_2(s_0)$	$d(s_0)$
$\{i_1, i_2\}$	$3(y_{\kappa_1} + y_{\kappa_2} + y_{\kappa_3})$	0	0	0
$\{i_1, i_3\}$	$3(y_{\kappa_2} + y_{\kappa_3})$	0	0	0
$\{i_1, i_4\}$	$3(y_{\kappa_1} + 2y_{\kappa_4})$	$3y_{\kappa_1}$	$3y_{\kappa_1}$	0
$\{i_2, i_3\}$	0	0	0	0
$\{i_2, i_4\}$	0	$3(y_{\kappa_2} + y_{\kappa_3})$	$3(y_{\kappa_1} + y_{\kappa_2} + y_{\kappa_3} + 2y_{\kappa_4})$	$3(y_{\kappa_1} + 2y_{\kappa_4})$
$\{i_3, i_4\}$	0	$3(y_{\kappa_1} + y_{\kappa_2} + y_{\kappa_3} + 2y_{\kappa_4})$	$3(y_{\kappa_2} + y_{\kappa_3})$	$-3(y_{\kappa_1} + 2y_{\kappa_4})$

Clearly, V and W are orthogonal, but $\mathfrak{E}_\theta(V) \oplus \mathfrak{E}_0(W) \neq \mathfrak{E}_\theta\big(Row(\mathbb{S}^p)\big) \oplus \mathfrak{E}_0\big(Ker(\tilde{\mathbb{S}}^p)\big)$, since e_0 is not a sample-space spanned estimator. One can now follow the proof of Proposition 5 in Appendix A to show that $V(e_0) < V(e_0 + d)$; hence, $e = e_0 + d$ is inadmissible.

4 Concluding remarks

In this work, we have analyzed the admissibility of many unbiased estimators in BIGS, depending on the available knowledge of the graph, and identified many other admissible estimators in BIGS than the HTE.

In the case of graph knowledge, the investigation has naturally led us to the concepts of sufficiency, Rao-Blackwellization and its zero-invariant modification ZRB. In Theorem 1, we have proved that for unbiased zero-invariant estimators, admissibility is equivalent to sufficiency and the application of ZRB. Thus, given graph knowledge (or successor-ancestry knowledge which enables the application of ZRB for the realised sample graph), it will always be better to perform ZRB of any unbiased estimator to ensure its admissibility.

In the case of ancestry knowledge, we have established in Theorem 2 that all sample-space spanned estimators are admissible in the class of unbiased, zero-invariant and analytic estimators, as well as devised a constructive approach to identify inadmissible estimators therein by means of orthogonal vector spaces (Proposition 5 and Example 2).

Table 2: Admissibility of some IWE estimators in BIGS

Knowledge	Estimator class	HTE	HT-type	HH-type
Graph	D^*	Admissible	Inadmissible	Inadmissible
Ancestry	D^{**}	Admissible	Unknown generally	Admissible

The results regarding the examples of IWE introduced at the beginning are summarized in Table 2. As one can see, with graph knowledge, the HTE is admissible because it is sufficient, but the others are inadmissible because they are not sufficient. However, provided ancestry knowledge, the HH-type estimators are admissible in the class D^{**} and cannot be dominated by the HTE; whereas the HT-type estimators with non-constant weights $W_{i\kappa}$ are not covered by Theorem 2, so their admissibility is still unknown generally, although there are special cases such as Corollary 3 if one only considers restricted designs.

Finally, since all the sampling-unbiased estimators known in the literature are analytic, restricting to the class D^{**} is not a limitation that matters much for the practice of sampling theory, compared to studying the classes D^* or D more broadly. An obvious topic for future research would be to further delineate admissibility for the estimators in D^{**}, possibly by other means than pairwise orthogonal vector spaces devised in this paper.

A Proofs

Proof of Proposition 1.

Proof. Let us assume that e meets the condition of the statement. Since no strict inequality can occur because e_0 is admissible, the variance must always be equal $\mathrm{V}(e) = \mathrm{V}(e_0)$, that is, if $e = e_0 + d$, then

$$\mathrm{V}(e) = \mathrm{V}(e_0) + \mathrm{V}(d) + 2\mathrm{Cov}(e_0, d) \Leftrightarrow \mathrm{V}(d) = -2\mathrm{Cov}(e_0, d).$$

Observe that $\mathrm{Cov}(e_0, d) \leq 0$. If for some graph $\mathcal{B}$ and vector y, $e \neq e_0 \Rightarrow d \neq 0$ and $\mathrm{Cov}(e_0, d) < 0$, then, since $e_\alpha = e_0 + \alpha d \in \mathcal{D}$ for any $\alpha \in (0, 1)$, we get

$$\mathrm{V}(e_\alpha) = \mathrm{V}(e_0) + \alpha^2\mathrm{V}(d) + 2\alpha\mathrm{Cov}(e_0, d) = \mathrm{V}(e_0) + 2\alpha(1-\alpha)\mathrm{Cov}(e_0, d) \leq \mathrm{V}(e_0),$$

and $\mathrm{V}(e_\alpha) < \mathrm{V}(e_0)$ for the cases where $\mathrm{Cov}(e_0, d) < 0$. This is a contradiction with the fact that e_0 is admissible, such that $e = e_0$ is the only possibility. □

Proof of Lemma 1.

Proof. By the Lagrangian of the optimization problem, we have

$$L(e,\lambda) = \sum_{s_0\in\mathcal{S}_0^*} p(s_0)e^2(s_0) - \lambda\left(\sum_{s_0\in\mathcal{S}_0^*} p(s_0)e(s_0) - Y^*\right),$$
$$\frac{\partial L}{\partial e(s_0)} = 2p(s_0)e(s_0) - \lambda p(s_0) = 0 \quad \Rightarrow \quad e(s_0) = \frac{\lambda}{2}, \ \ \forall s_0 \in \mathcal{S}_0^*.$$

Using the constraint we obtain

$$\lambda = \frac{2Y^*}{\sum_{s_0\in\mathcal{S}_0^*} p(s_0)} \quad \Rightarrow \quad e(s_0) = \frac{\lambda}{2} = \frac{Y^*}{\sum_{s_0\in\mathcal{S}_0^*} p(s_0)}.$$

Since the Hessian is positive definite, the solution corresponds to a local minimum. As we are minimizing a convex function over a convex set, the minimum is also global. Moreover, since the function is strictly convex, the minimum is unique and the lemma holds. □

Proof of Proposition 3.

Proof. The second inequality is simply the Rao-Blackwell theorem. Let us focus on the first inequality. If $y(\Omega)$ are all non-zero, then $e_{ZRB} = e_{RB}$. Otherwise, let $y(\Lambda)$ consist of all the zero values, where $\Lambda \subset \Omega$. Between $\mathcal{B}^{(\Lambda)}$ and $\mathcal{B}$ let us define

$$\mathfrak{O}(\mathcal{B},\Lambda) = \{\Omega_s(s_0,\mathcal{B}^{(\Lambda)}) : \exists s_0, s_0' \text{ s.t. } \alpha(s_0;\mathcal{B}) \neq \alpha(s_0';\mathcal{B}) \text{ and } \alpha(s_0;\mathcal{B}^{(\Lambda)}) = \alpha(s_0';\mathcal{B}^{(\Lambda)})\}$$

and work by induction on the cardinality $K = |\mathfrak{O}(\mathcal{B},\Lambda)|$.

First, if $K = 1$, let the only element of $\mathfrak{O}(\mathcal{B},\Lambda)$ be $\Omega_s(s_0';\mathcal{B}^{(\Lambda)}) = \Omega_s(s_0';\mathcal{B}) \cap \Omega_{(0)}$, and $[s_0'] = \{s_0 : \alpha(s_0;\mathcal{B}^{(\Lambda)}) = \Omega_s(s_0';\mathcal{B}^{(\Lambda)})\}$. Since both e_{RB} and e_{ZRB} are unbiased, and equal to each other whenever $s_0 \notin [s_0']$, the problem reduces to comparing the second moments for the elements in $[s_0']$. Additionally, note that the estimators e, e_{RB} and e_{ZRB} satisfy:

$$\sum_{s_0\in[s_0']} p(s_0)e(s_0) = Y^* := \theta - \sum_{s_0\notin[s_0']} p(s_0)e(s_0).$$

Applying Lemma 1 with $\mathcal{S}_0^* = [s_0']$ yields $V(e_{ZRB}) \leq V(e_{RB})$ where the inequality is strict if $|[s_0']| > 1$.

Next, suppose $K > 1$. Define estimators $e_{ZRB,1}, e_{ZRB,1}, ..., e_{ZRB,K} = e_{ZRB}$, where $e_{ZRB,1}$ is given by applying ZRB only to the first element of $\mathfrak{O}(\mathcal{B},\Lambda)$, and $e_{ZRB,2}$ by applying ZRB only to the first two elements of $\mathfrak{O}(\mathcal{B},\Lambda)$, and so on. Applying Lemma 1 to compare $e_{ZRB,k}$ and $e_{ZRB,k-1}$, for $k = K, ..., 2$, as well as $e_{ZBR,1}$ and e_{RB}, we obtain

$$\mathrm{V}(e_{ZRB}) = \mathrm{V}(e_{ZRB,K}) \leq \mathrm{V}(e_{ZBR,K-1}) \leq \cdots \leq \mathrm{V}(e_{ZBR,1}) \leq \mathrm{V}(e_{RB}).$$

This completes the proof. □

Proof of Theorem 1.

Proof. i) ⇒ ii) We shall use a double induction on $N = |F|$ and $K = |\Omega|$. For $N = 1$ the only possible design is $s_0 = \{1\}$ and $p(s_0) = 1$. Since e_0 is unbiased, i.e. $p(s_0)e_0(s_0) = \theta$, the only possible estimator is $e_0(s_0) = \theta$ which is therefore trivially admissible.

We now assume that the result is true for N. To show that the theorem holds for $N+1$, we will show that for any given design $p(s_0)$ given $|F| = N+1$, where $\pi_i > 0$ for any $i \in F$, for any estimator $e(s_0; y, \mathcal{B}) \in D^*$ such that

$$\mathrm{V}\left(e(s_0; y, \mathcal{B})\right) \leq \mathrm{V}\left(e_0(s_0; y, \mathcal{B})\right), \quad \forall \mathcal{B} \in \mathfrak{B}(p),\ y(\Omega), \tag{5}$$

it holds that $e(s_0; y, \mathcal{B}) = e_0(s_0; y, \mathcal{B})$ for every $s_0 \in \mathcal{S}_0$.

We will prove the above property by induction on $K = |\Omega|$. If $K = 1$, the graphs associated with this case are in the form shown in Figure 6. There could be elements in F that do not connect to κ, and if there is a sample formed exclusively by these elements, the value of the estimator, due to the zero-invariance property, would be zero.

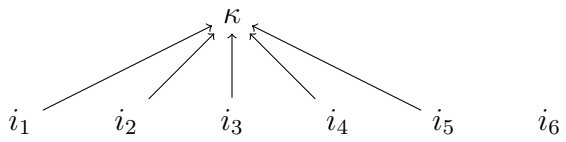

Figure 6: Illustration of a graph in the case $K = 1$.

Note that either $\alpha(s_0) = \emptyset$ and the estimator takes the value zero, or $\alpha(s_0) = \{\kappa\}$. Since e_0 is sufficient, for any pair of samples s_0, s_0' such that $\alpha(s_0) = \alpha(s_0') = \{\kappa\}$ the estimator takes the same value, i.e. $e_0(s_0; y, \mathcal{B}) = e_0(s_0', y, \mathcal{B})$. As the estimator is unbiased, we have:

$$\sum_{\alpha(s_0)=\{\kappa\}} p(s_0) e_0(s_0; y, \mathcal{B}) = \theta \quad \Leftrightarrow \quad e_0(s_0; y, \mathcal{B}) \sum_{\alpha(s_0)=\{\kappa\}} p(s_0) = \theta$$

$$\Leftrightarrow \quad e_0(s_0; y, \mathcal{B}) = \frac{\theta}{\sum_{\alpha(s_0)=\{\kappa\}} p(s_0)},\ \forall s_0 \text{ s.t. } \alpha(s_0) = \{\kappa\}.$$

If we apply Lemma 1, considering that $Y^* = \theta$ and $\mathcal{S}_0^*$ is the set of samples that satisfy $\alpha(s_0) = \{\kappa\}$, we obtain that the estimator above corresponds to the one that minimizes the variance. As e meets inequality (5), then $e = e_0$ and the result holds for $K = 1$.

Let us assume that the result is true for K and prove it for $K+1$. Let $\mathcal{B} \in \mathfrak{B}(p)$ such that $|F| = N+1$ and $|\Omega| = K+1$ and $\kappa' \in \Omega$. Let $y_{\kappa'} = 0$. Now consider the graph $\mathcal{B}^{(\kappa')}$ where we remove κ' and any edge that ends at κ'. Since e and e_0 are zero-invariant we have

$$e(s_0, \mathcal{B}) = e(s_0, \mathcal{B}^{(\kappa')}) = e_0(s_0, \mathcal{B}^{(\kappa')}) = e_0(s_0, \mathcal{B}),$$

where the first and third equalities are due to zero-invariance and the second one due to the induction hypothesis since $\mathcal{B}^{(\kappa')}$ has K study units.

In this way, we have shown that for any value of y that vanishes at some element $\kappa' \in \Omega$ and for any graph $\mathcal{B}$, $e(s_0; y, \mathcal{B}) = e_0(s_0; y, \mathcal{B})$ for any $s_0 \in \mathcal{S}_0$. Therefore, considering any sample s_0 such that $\alpha(s_0) \neq \Omega$ and a general vector y, we can choose some $\kappa' \in \Omega \setminus \alpha(s_0)$ and, by the same argument, show that $e(s_0; y, \mathcal{B}) = e_0(s_0; y, \mathcal{B})$ for any graph.

Let us see what happens if the sample s_0 satisfies $\alpha(s_0) = \Omega$. In this case, as e meets e_0

for the rest of samples s_0' with $\alpha(s_0') \neq \Omega$, from unbiasedness and inequality (5) we get

$$\sum_{\alpha(s_0)=\Omega} p(s_0)e(s_0; y, \mathcal{B}) = Y^* = \theta - \sum_{\alpha(s_0)\neq\Omega} p(s_0)e_0(s_0; y, \mathcal{B}), \tag{6}$$

$$\sum_{\alpha(s_0)=\Omega} p(s_0)e^2(s_0; y, \mathcal{B}) \leq \sum_{\alpha(s_0)=\Omega} p(s_0)e_0^2(s_0; y, \mathcal{B}). \tag{7}$$

Since e_0 is sufficient and $\alpha(s_0) = \Omega$ we get from (6):

$$\sum_{\alpha(s_0)=\Omega} p(s_0)e_0(s_0; y, \mathcal{B}) = Y^* \quad \Rightarrow \quad e_0(s_0; y, \mathcal{B}) = \frac{Y^*}{\sum_{\alpha(s_0)=\Omega} p(s_0)}.$$

If we apply Lemma 1 for minimizing the left-hand side of (7) subject to (6), considering $\mathcal{S}_0^*$ as the set of samples that satisfy $\alpha(s_0) = \Omega$, we obtain the above estimator as the one that minimizes the variance and then by (7), $e(s_0; y, \mathcal{B}) = e_0(s_0; y, \mathcal{B})$ and the result holds.

ii) $\Rightarrow$ *iii)* Note that if e_0 is not ZRB(e) of any estimator $e \in D^*$, then $e_0 \neq \text{ZRB}(e_0)$ and by Proposition 3, ZRB(e_0) improves upon e_0 for some graph. Thus, e_0 is not admissible.

Another way to prove this is by applying the Rao-Blackwell Theorem, which shows that $\text{V}(e_{ZRB}) \leq \text{V}(e_0)$. If e_0 were admissible, we would have $e_0 = ZRB(e_0)$ by Proposition 1.

iii) $\Rightarrow$ *i)* If $e_0 = \text{ZRB}(e)$ for some $e \in D^*$, then e_0 is a sufficient estimator in D^*. □

Proof of Proposition 4.

Proof. Let B be an elemental graph with the only edge $H = \{(j, \kappa)\}$. As e is zero-invariant, we have $a_i = 0$ if $i \neq j$, and $e(s_0; y, \mathcal{B}) = a_j(y_\kappa, \mathcal{B})$ for every s_0 containing j. It follows from unbiasedness that $a_j(y_\kappa, \mathcal{B}) = y_\kappa/\pi_i$, i.e., e is elemental. □

Proof of Lemma 2.

Proof. Between weighted and unweighted scalar products, we have

$$v \in \text{Ker}(\tilde{\mathbb{S}}^p)^{\perp_p} \iff \sum_{s_0\in\mathcal{S}_0} p(s_0)v(s_0)d(s_0) = 0, \ \forall d \in \text{Ker}(\tilde{\mathbb{S}}^p) \iff pv \in \text{Ker}(\tilde{\mathbb{S}}^p)^{\perp}.$$

Since $\text{Ker}(\tilde{\mathbb{S}}^p)^{\perp} = \text{Row}(\tilde{\mathbb{S}}^p)$, we have $pv \in \text{Row}(\tilde{\mathbb{S}}^p) \iff v \in \text{Row}(\mathbb{S}^p)$. □

Proof of Proposition 5.

Proof. We have $\text{V}(e) = \text{V}(e_0) + \text{V}(d) + 2\text{Cov}(e_0, d)$. Since $\text{E}(e) = \theta$ and $\text{E}(d) = 0$, we obtain

$$\text{Cov}(e_0, d) = \sum_{s_0} p(s_0)\{e_0(s_0) - \theta\}d(s_0) = \sum_{s_0} p(s_0)e_0(s_0)d(s_0) = e_0 \cdot_p d = 0$$

due to the orthogonality established by Lemma 2, such that $\text{V}(e_0) \leq \text{V}(e)$ for every $(y, \mathcal{B})$, and $\text{V}(e_0) < \text{V}(e)$ whenever $\text{V}(d) > 0$ because $d \not\equiv 0$ for given $(y, \mathcal{B})$. □

Proof of Theorem 2.

Proof. Suppose there exists some estimator $e \in D^{**}$ such that

$$\mathrm{V}\left(e(s_0; y, \mathcal{B})\right) \leq \mathrm{V}\left(e_0(s_0; y, \mathcal{B})\right), \quad \forall \mathcal{B} \in \mathfrak{B}_N,\ y \in \mathbb{R}^{|\Omega(\mathcal{B})|}.$$

We shall prove that $e(s_0; y, \mathcal{B}) = e_0(s_0; y, \mathcal{B})$, $\forall s_0 \in \mathcal{S}_0$, $y \in \mathbb{R}^{|\Omega(\mathcal{B})|}$ and $\mathcal{B} \in \mathfrak{B}(p)$.

To do so, we adopt the following approach. First, suppose that e differs to e_0 given some graph $\mathcal{B}$. Note that, since they are both unbiased, we can express e_0 and e given $\mathcal{B}$ as

$$e_0(s_0; y, \mathcal{B}) = \theta + h_0(s_0; y, \mathcal{B}) \quad \text{and} \quad e(s_0; y, \mathcal{B}) = \theta + h(s_0; y, \mathcal{B})$$

where $\sum_{s_0 \in \mathcal{S}_0} p(s_0) h_0(s_0) = \sum_{s_0 \in \mathcal{S}_0} p(s_0) h(s_0) = 0$.

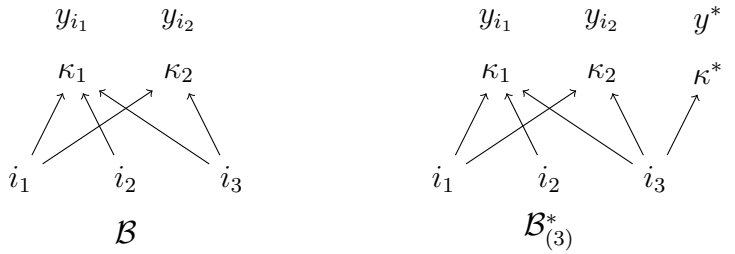

Figure 7: Illustration of $\mathcal{B}^*_{(i_3)}$ constructed from $\mathcal{B}$.

Next, let us construct a new graph $\mathcal{B}^*_{(j)}$ from $\mathcal{B}$ by adding a new element κ^* to Ω and the edge (j, κ^*) to H, with the value y^* associated with κ^*, as illustrated in Figure 7. We shall prove that there exists some vector $\overline{y}$ associated with some $\mathcal{B}^*_{(j)}$, consisting of $y(\Omega)$ and y^*, where the variance of e_0 is less than that of e, contradicting the hypothesis that $\mathrm{V}(e_0) \geq \mathrm{V}(e)$ and resulting in the only possibility that e and e_0 could not differ given $\mathcal{B}$.

To argue for the contradiction, we shall distinguish whether e is elemental or not, noting that e_0 is elemental by Proposition 4.

Case 1: Suppose e is not elemental. Let us take $y = \mathbf{0}$ and compare the variance of e_0 and e. As e is not elemental there exists some $j^* \in F$ such that for the elemental graph $\mathcal{B}^*_{(j^*)}$ there is some sample s_0^* such that $e(s_0^*; y^*, \mathcal{B}^*_{(j^*)}) \neq y^*/\pi_{j^*}$. Now that both e_0 and e are zero-invariant, we have $e_0(s_0) = e(s_0) = 0$ for all $s_0 \not\ni j^*$. Moreover, since e_0 and e are unbiased estimators, both satisfy

$$\sum_{s_0 \in \mathcal{S}_0} p(s_0) e(s_0; \bar{y}, \mathcal{B}^*_{(j^*)}) = \sum_{s_0 \ni j^*} p(s_0) e(s_0; \bar{y}, \mathcal{B}^*_{(j^*)}) = y^*.$$

It follows that had $e(s_0)$ been the same for every sample $s_0 \ni j^*$, we could only have $e(s_0; y^*, \mathcal{B}^*_{(j^*)}) = y^*/\pi_{j^*}$ whenever $s_0 \ni j^*$. Therefore, $e(s_0)$ cannot be constant over s_0 if it is not elemental. Thus, by Lemma 1, the variance of e_0 achieves the minimum since e_0 is elemental, such that $\mathrm{V}(e) > \mathrm{V}(e_0)$, which yields the contradiction.

Case 2: Suppose e is elemental, i.e. $e(s_0; \bar{y}, \mathcal{B}^*_{(j)}) = y^*/\pi_j$ whenever $s_0 \ni j$ given any choice of $j \in F$ in the case $y = \mathbf{0}$. As the estimators e and e_0 are elemental and analytic, they are, in particular, continuous, and the following holds:

$$\lim_{y \to \mathbf{0}} e_0(s_0; \bar{y}, \mathcal{B}^*_{(j)}) = \lim_{y \to \mathbf{0}} e(s_0; \bar{y}, \mathcal{B}^*_{(j)}) = y^*/\pi_j,$$

for every $s_0 \ni j$. Whereas, by the zero-invariance property, we have that, for any $s_0 \not\ni j$,

$$\lim_{y\to\mathbf{0}} h_0(s_0) = \lim_{y\to\mathbf{0}} h(s_0) = 0.$$

Now, the variance of e_0 given the graph $\mathcal{B}^*_{(j)}$ can be decomposed as

$$\begin{aligned}
\mathrm{V}\left(e_0(s_0;\mathcal{B}^*_{(j)})\right) &= \sum_{s_0\not\ni j} p(s_0)\left(\theta + h_0(s_0) - \theta - y^*\right)^2 + \sum_{s_0\ni j} p(s_0)\left(e_0(s_0) - \theta - y^*\right)^2 \\
&= \sum_{s_0\not\ni j} p(s_0)h_0^2(s_0) + (1-\pi_j)(y^*)^2 - 2y^*\sum_{s_0\not\ni j} p(s_0)h_0(s_0) + \\
&\quad + \sum_{s_0\ni j} p(s_0)\left(e_0(s_0) - \theta - y^*\right)^2.
\end{aligned}$$

Similarly for e, such that $\Delta = \mathrm{V}(e_0(s_0;\mathcal{B}^*_{(j)})) - \mathrm{V}(e(s_0;\mathcal{B}^*_{(j)}))$ is given as

$$\begin{aligned}
\Delta &= \sum_{s_0\ni j} p(s_0)\left(e_0(s_0) - \theta - y^*\right)^2 - \sum_{s_0\ni j} p(s_0)\left(e(s_0) - \theta - y^*\right)^2 \\
&\quad + \sum_{s_0\not\ni j} p(s_0)\left(h_0^2(s_0) - h^2(s_0)\right) - 2y^*\sum_{s_0\not\ni j} p(s_0)\left(h_0(s_0) - h(s_0)\right)
\end{aligned}$$

where, as $y \to \mathbf{0}$, the sum of the first two terms above approaches zero because the estimators are elemental, and the third term approaches zero as obtained above. Thus, for every $j \in F$ and $\epsilon > 0$ there exists some $\delta > 0$ such that, for every $\|y\| < \delta$, the sum of the first three terms are bounded by ϵ absolutely. We can then write, for $\epsilon^*_{(j)} \in (-\epsilon, \epsilon)$ and $\forall j \in F$,

$$\mathrm{V}\left(e_0(s_0;\mathcal{B}^*_{(j)})\right) - \mathrm{V}\left(e(s_0;\mathcal{B}^*_{(j)})\right) = \epsilon^*_{(j)} - 2y^*\sum_{s_0\not\ni j} p(s_0)\left(h_0(s_0) - h(s_0)\right).$$

Note that since e differs to e_0 given $\mathcal{B}$, there exists some sample s_0^* such that $h_0(s_0^*) \neq h(s_0^*)$ for some value of y. Now that $d(s_0) = h(s_0) - h_0(s_0)$ is a non-zero analytic function, all its zeros are isolated, and there exist values of y arbitrarily close to zero where $d(s_0^*) \neq 0$. Choose a small enough value for y such that $d(s_0^*) \neq 0$ and consider the following.

Subcase 2.1: Suppose that

$$\sum_{s_0\not\ni j} p(s_0)d(s_0) = 0,$$

for all $j \in F$. Then, as $\sum_{s_0\in\mathcal{S}_0} p(s_0)d(s_0) = 0$ due to the unbiasedness, it holds that

$$\sum_{s_0\ni j} p(s_0)d(s_0) = 0,\ \forall j \in F.$$

In other words, $d \in \mathfrak{E}_0\big(\mathrm{Ker}(\tilde{\mathbb{S}}^p)\big)$. By Proposition 5, e_0 improves e with strict inequality when $d \neq 0$, in which case $\mathrm{V}(e) \leq \mathrm{V}(e_0)$, which yields the contradiction we seek.

Subcase 2.2: Suppose there exists some $j^* \in F$ such that

$$\sum_{s_0\not\ni j^*} p(s_0)d(s_0) \neq 0.$$

Now, if we take j^* and a small enough value of y, we have:

$$\mathrm{V}\left(e_0(s_0;\mathcal{B}^*_{(j^*)})\right) - \mathrm{V}\left(e(s_0;\mathcal{B}^*_{(j^*)})\right) < 0 \Leftrightarrow \epsilon^*_{(j^*)} < 2y^* \sum_{s_0 \not\ni j^*} p(s_0)\left(h_0(s_0) - h(s_0)\right).$$

To find y^* that satisfies the previous inequality, it is enough to let y^* be a sufficiently large positive value if $\sum_{s_0 \not\ni j^*} p(s_0)\left(h_0(s_0) - h(s_0)\right) > 0$ or a sufficiently large negative value if $\sum_{s_0 \not\ni j^*} p(s_0)\left(h_0(s_0) - h(s_0)\right) < 0$. Thus, there exist some $\overline{y}$ and $\mathcal{B}^*_{(j^*)}$ where $\mathrm{V}(e) \leq \mathrm{V}(e_0)$ does not hold, yielding the contraction we seek. □

Proof of Corollary 3.

Proof. Note that if $p(s_0)$ is full-rank then by the rank-nullity theorem $\mathrm{Ker}(\tilde{\mathbb{S}}^p) = \mathrm{Ker}(\mathbb{S}^p) = \mathbf{0}$. By equation (4), $\mathrm{Row}(\mathbb{S}^p)$ is the complete space $\mathbb{R}^{|\mathcal{S}_0|}$ and every estimator in D^{**} belongs to $\mathfrak{E}_\theta\big(\mathrm{Row}(\mathbb{S}^p)\big)$. It follows then from Theorem 2 that every such estimator is admissible. □

References

[1] Avrachenkov, K., Ribeiro, B. and Towsley, D. (2010). Improving Random Walk Estimation Accuracy with Uniform Restarts. *Research report*, RR-7394, INRIA.

[2] Basu, D. (1971). An essay on the logical foundations of survey sampling, part I (with discussion). In *Foundations of Statistical Inference*, eds V.P. Godambe and D.A. Sprott, Toronto: Holt, Rinehart and Winston, 203–24.

[3] Birnbaum, Z.W. and Sirken, M.G. (1965). *Design of Sample Surveys to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimates*. Vital and Health Statistics, Ser. 2, No.11. Washington:Government Printing Office.

[4] Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *Annals of Mathematical Statistics*, 18: 105-110.

[5] Cochran, W.G. (1977). *Sampling Techniques (3rd ed.)*. New York: Wiley.

[6] Frank, O. (1980). Sampling and inference in a population graph. *International Statistical Review*, 48:33-41.

[7] Frank, O. (1971). *Statistical inference in graphs*. Stockholm: Försvarets forskningsanstalt.

[8] Frank O. and Snijders T. (1994). Estimating the size of hidden populations using snowball sampling. *Journal of Official Statistics*, 10:53-53.

[9] Godambe, V.P. (1955). A unified Theory of Sampling from Finite Populations. *Journal of the Royal Statistical, Society B*, 17:269-278.

[10] Godambe, V.P. (1960). An optimum property of Regular Maximum Likelihood Estimation. *The Annals of Mathematical Statistics*, 31:1208-1211.

[11] Godambe, V.P. and Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite populations I. *The Annals of Mathematical Statistics*, 36:1707-1722.

[12] Goodman, L.A. (1961). Snowball sampling. *Annals of Mathematical Statistics*, 32:148-170.

[13] Hanurav, T.V. (1966) Some aspects of Unified Sampling Theory, *Sankhya A*, 28:175-204.

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

[15] Joshi, V.M. (1965). Admissibility and Bayes estimation in sampling finite populations II. *The Annals of Mathematical Statistics*, 36:1723-1729.

[16] Joshi, V.M. (1966). Admissibility and Bayes estimation in sampling finite populations IV. *The Annals of Mathematical Statistics*, 37:1658-1670.

[17] Lavallée, P. (2007). *Indirect Sampling*. New York: Springer.

[18] Patone, M. and Zhang, L.-C. (2022). Weighting estimation under bipartite incidence graph sampling. *Statistical Methods & Applications*, 32:447-467.

[19] Ramakrishnan, M.K. (1973). An alternative proof of the admissibility of the Horvitz-Thompson estimator. *The Annals of Statistics*, 1:577-579.

[20] Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of Calcutta Mathematical Society*, 37:81-91.

[21] Sirken, M.G. (2005). *Network Sampling*. In *Encyclopedia of Biostatistics*, John Wiley & Sons, Ltd. DOI:10.1002/0470011815.b2a16043

[22] Tillé, Y. (2006). *Sampling Algorithms*. Springer.

[23] Thompson, S.K. (2006). Targeted random walk designs. *Survey Methodology*, 32:11-24.

[24] Thompson, S.K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association*, 85:1050-1059.

[25] Zhang, L.-C. (2024). Graph spatial sampling. *Stat*, 13:e708. https://doi.org/10.1002/sta4.708

[26] Zhang, L.-C. (2021). Graph sampling by lagged random walk. *Stat*, 11:e444. https://doi.org/10.1002/sta4.444

[27] Zhang, L.-C. (2021a). Graph sampling – An introduction. *The Survey Statistician*, 83:27-37.

[28] Zhang, L.-C. (2021b). *Graph Sampling*. Chapman and Hall, CRC Press.

[29] Zhang, L.-C. and Oguz-Alper, M. (2020). Bipartite incidence graph sampling. https://arxiv.org/abs/2003.09467

[30] Zhang, L.-C. and Patone, M. (2017). Graph sampling. *Metron*, 75:277-299. DOI:10.1007/s40300-017-0126-y