On the two-phase framework for joint model and design-based inference

Susana Rubin-Bleuer[1] and Ioana Schiopu Kratina

Statistics Canada

[1]This research work was partly supported by the Sabbatical Research Program of Statistics Canada.

ABSTRACT. We establish a mathematical framework that formally validates the two-phase "super-population viewpoint" proposed by Hartley and Sielken (1975), by defining a product probability space which includes both the design space and the model space. We develop a general methodology that combines finite population sampling theory and classical theory of infinite population sampling to account for the underlying processes that produce the data. Key results in this article are: the sample estimator and the model statistic are asymptotically independent; if a sequence converges in design law, it also converges in the law of the product space; and the distribution theory of the sample estimating equation estimator around a super-population parameter. We also study the interplay between dependence and independence of random variables when viewed in the design space, the product space and the model space and apply it to show formally that under a "simple random sample without replacement" design, we can "ignore" the design and work on the realm of the model space, but that under "simple random sample with replacement" we cannot ignore the design.
Key words: joint design and model-based inference; product space.

1. Introduction

Classical sampling theory is concerned with inference for finite population parameters. This enables us to work exclusively within a sample probability space, which we design and control, and therefore it is completely known to us. However, there are many situations when we have to resort to postulating a model, e.g., when we wish to draw conclusions on a more general population than the finite population from which we obtained the sample or to perform a test of hypothesis. Even for descriptive analysis in a finite population, we need a model when we have to deal with non-response, small area estimation or measurement errors. Once we incorporate a general population model in our framework, our inference procedures would ideally have to account for the design (unequal selection probabilities, dependent selection indicators, etc.), other survey processes (non-response adjustment, calibration, etc.) and the model defining the relationships amongst the variables being studied.

To this purpose, Hartley and Sielken (1975) introduced the "super-population" approach to describe the relationship between the infinite population (also called super-population) and the finite population from which we select the sample. It regards the sample selected by the surveyor according to a specified design, as the result of a two-phase procedure, where the super-population generates the finite population that could have been observed, had we taken a census. Many authors worked within the two-phase framework and accounted for the variability due to the design and the model by means of the "anticipated variance". The contributions of Fuller (1975), Isaki and Fuller(1982), Godambe and Thompson (1986), Korn and Graubard (1998), Pfeffermann and Sverchkov (1999), Binder and Roberts (1999), Rodríguez(2001), Molina, Smith and Sudgen (2001), are just a few among the vast literature on the subject.

Fuller (1975) established large sample properties of the sample regression estimator around the model parameter with data obtained from stratified cluster samples. His approach could only be applied to designs with "simple random sample without replacement"(SRSWOR) in the first stage sampling within strata. A general approach to estimation of model parameters based on samples drawn from complex surveys has not yet been formally established, even in the case where the (first stage) sampling rate is negligible. As an illustration, let us consider the case of the sample mean. We have:

$$\sqrt{n}(\bar{y} - \mu) = \sqrt{n}(\bar{y} - \bar{Y}) + \sqrt{(n/N)}\sqrt{N}(\bar{Y} - \mu), \qquad (1.1)$$

where μ is the super-population mean, N, $\bar{Y}$ are the finite population size and mean respectively, and n, $\bar{y}$ are the sample size and sample mean, respectively.

The large sample properties of the first term on the right hand side of (1.1) have been studied for many designs. Conditions were given for the distribution of the sample mean around the finite population mean to be approximately normal with mean zero and design variance Γ_d (design-based CLT): for SRSWOR and rejective sampling with varying probabilities by Hájek (1960, 1964), for πps designs by Rosén (1972, 1997), and for stratified multistage probability proportional to size designs, by Krewski and Rao (1981). To infer from this result a design-based CLT for the left hand side of (1.1), we would have to assume not only that the sampling rate n/N converges to zero, but also that the sequence of numbers $\sqrt{N}(\bar{Y} - \mu)$ is bounded as n and $N \to \infty$. As a sequence of numbers, this last condition is very restrictive. However, as a sequence of sums of independent, identically distributed (i.i.d.) random variables in the super-population, $\sqrt{N}(\bar{Y} - \mu)$ is bounded in probability and the second term of the right hand side above would converge to zero in the probability of the model as the sampling rate n/N converges to zero. Thus, when we study the asymptotic properties of the sample means around the super-population mean, it makes more sense to integrate the model and the design under the same umbrella.

In this article, we establish a mathematical framework which formally validates the two-phase "super-population viewpoint" advocated by Hartley and Sielken (1975) by defining a product probability space which includes both the design space and the model space (Definitions 4.1 and 4.2). The product space makes it possible to consider joint convergence of design-based estimators and model based estimators, which are originally defined on different probability spaces (Theorem 5.1). In this set-up, Remarks 4.3 and 5.1 show that the design probability and the distribution of a design-based estimator are second phase concepts, i.e., conditional probabilities given minimal information contained in the model. We describe a general methodology that combines finite population sampling theory and classical theory of infinite population sampling to account for the underlying processes that produce the data. We show that this approach enables us to prove the CLT for estimating equation estimators derived from a complex sample and make inference on a super-population parameter, for sampling designs other than those presented in Fuller(1975) (Section 6). We also show formally that when dealing with survey data we cannot ignore the effect of "with replacement" (WR) designs, even if they do not induce selection bias (Proposition 4.1, Example 4.5). Other applications enable us to adapt survival analysis methods to be used with complex survey data (see Rubin-Bleuer (2001)). We can also accommodate in the product space super-population inference techniques used by other authors. In general, we could apply this approach to most situations where we have a two phase randomization process, e.g. a two-phase sampling selection in a finite population.

We remark that in order to obtain the total (anticipated) variance in (1.1) we must impose (model-based) conditions on the super-population model, which survey statisticians would rather avoid. At the very least, some form of model-based independence is needed. Hence many authors assume that the sampling rate is small enough so they can ignore the variation due to the model component. However, the examples given by Korn and Graubard (1998) show that we should not dismiss the second term in the total variance without checking first that it is indeed sufficiently small relative to the first term. Even an approximate knowledge of the model component of the variance may be used to our advantage in designing a survey.

The article is organized as follows: Sections 2 to 5 below develop the tools necessary to do inference while integrating the design into the model, and Section 6 is an application of the product space methodology. In Section 2 we give a slightly more restrictive definition of the sample design and estimator (Definitions 2.3, 2.4) in order to view them subsequently as random variables in the super-population(Definition 4.2, Remark 4.1). In Section 3 we adopt the super-population definition in Särndal et al (1992) to define what it means for a finite population to be generated by a super-population (Definition 3.1). In Section 4 we define the general product space (Definitions 4.1, 4.3) and show how stochastic dependence is introduced in the product space (Example 4.1); we show different forms the product space can take according to the model and wether the design is single stage or multiple stage (Examples 4.2, 4.3). Example 4.4 shows how the work of Pfeffermann and Sverchkov(1999) on estimation of regression models used with survey data, fits into the product space methodology. We exploit the additional information on the design and the model by deriving conditional probabilities which are used in later applications, and we study the interplay between dependence and independence of random variables when viewed in the design space, the product space and the model space. We show that we can “ignore” the design and work in the realm of the super-population space, when the design is SRSWOR (Examples 4.5, 4.6). This was the approach taken by Fuller(1975) to obtain asymptotic normality of the sample regression estimator around the super-population parameter. We show that, counter-intuitively, we cannot ignore the design if it is “simple random sample with replacement” (SRSWR) (Example 4.5).

In Section 5 we show that if the sample and super-population statistics converge in law in their respective spaces, they also converge in law in the product space. The two terms in the right hand side of (1.1) are not, in general, stochastically independent. We establish here their “asymptotic independence” under mild conditions (Theorem 5.1).

Finally, in Section 6 we establish the asymptotic normality of a sample estimator derived from a general estimating equation, under general conditions. We apply this to obtain the asymptotic normality of the ratio estimator of the average stratum mean under a stratified one-stage p.p.s.-design. We then give the type of conditions in the super-population that imply (design) asymptotic normality of a general sample estimating equation estimator for a two-stage sampling design. In this last example the product space is built with a model probability conditioned to “prior” information, i.e., information known at the time of the design and used for designing the two-stage sample.

2. Finite populations and sampling designs

Definition 2.1 A finite population $U = \{1, ..., N\}$ of size N consists of N units, or labels, with the associated data, i.e. each unit i is associated to a unique real valued vector (y_i, x_i, z_i) $i = 1, ..., N$. The components $y_i \in \mathbb{R}^p$ represent the characteristics of interest, $x_i \in \mathbb{R}^k$ represent the auxiliary information, and $z_i \in \mathbb{R}^q$ contains prior information available at the time of the design of the survey on all units $i = 1, ..., N$. Here p, k and q are positive integers. We write $y^N = (y_i)_{i=1,...,N}$, $x^N = (x_i)_{i=1,...,N}$ and $z^N = (z_i)_{i=1,...,N}$. In what follows "prior" information refers to information about the population available at the time of the design of a survey.

Remark 2.1 In this paper, N will denote the size of the finite population (i.e. the number of ultimate sampling units in the population) for one-stage-sampling schemes, and it will denote the number of clusters or primary sampling units (p.s.u.'s) for multistage schemes, in which case the size of the finite population will be denoted by M.

Definition 2.2 A sample is the realization of a probabilistic (randomized) selection or sampling scheme (Särndal et al, 1992, p. 25). We adopt the comprehensive definition of a sample in Hájek (1981, p.42): it views the sample as "a finite sequence of units or labels of the finite population, which are drawn one by one until the sampling is finished according to some stopping rule. This sequence distinguishes the order of units, may be of variable length and may include one unit of the finite population several times". This definition includes both samples selected "without replacement" (WOR), and "with replacement" (WR). We remark that under a WOR scheme, a sample can be viewed as a subset of labels or units from the finite population U and we may use this conceptual view of the sample when it is more convenient. Our framework accounts for both definitions.

Remark 2.2 For a stratified two-stage design the collection of all possible samples S is completely determined only if we know a priori all the strata and cluster sizes and their respective expected sample sizes. If so, S is well defined, since every label in the finite population must have a positive probability of selection.

In the literature, a design p associated with a sampling scheme is a probability function on the set of all possible samples under this scheme (see for example Särndal et al (1992)). The definition of a sampling design given below is more restrictive than the one above in that it requires measurability of p as a function of the variables containing the prior information.

Definition 2.3 Let U be the finite population of Definition 2.1. Given a sampling scheme, let S be the set of all possible samples under the scheme. Let *C*(S) consist of all subsets of S. *C*(S) is defined to provide a field for the probability space we are about to define. Let $D(z^N) \subseteq \mathbb{R}_+^{q\times N}$ be a subset of values of the prior information. A sampling design associated to a sampling scheme is a function $p\colon C(S)\times\mathbb{R}_+^{q\times N} \to [0,1]$ such that:

(I) $p(s,\cdot)$ is Borel - measurable in $\mathbb{R}_+^{q\times N}$, $\forall s \in S$

(ii) $p(\cdot, z_1, z_2, \ldots z_N)$ is a probability measure on *C*(S), $\forall (z_1,\ldots z_N) \in D(z^N)$

We say that (S, *C*(S), p) is a design probability space.

In all applications we will either take q = 1, or do not consider prior information. Under a two stage design with N primary sampling units (p.s.u.'s) we can carry on the design with prior information on the N p.s.u.'s only. The definition of design can be extended to include prior information on all sampling units.

The definition of a finite population parameter given below is more restrictive than that of Särndal (1992), p. 39. The measurability condition imposed on the parameter and its estimator ensures that, when the finite population is generated by a super-population, the finite population parameter and the estimator can be viewed as real-valued measurable functions (random variables) defined on the probability space associated with the super-population (see Definition 3.1).

Definition 2.4 Consider a finite population as in Definition 2.1. A real-valued, finite population parameter θ_N is a Borel- measurable function defined on a subset $D(y,x,z) \subseteq \mathbb{R}^{(p+k+q)\times N}$. An estimator of this finite population parameter associated with a design, also called sample estimator, is a function $\hat{\theta}_N : S \times D(y, x, z) \to \mathbb{R}$, where the domain $D(y, x, z) \subseteq \mathbb{R}^{(p+k+q)\times N}$, $\hat{\theta}_N(s, \cdot)$ is Borel - measurable. Note that $\hat{\theta}_N(\cdot, y^N, x^N, z^N)$ is C(S) - measurable since S is finite.

Remark 2.3 A sample estimator can be a design-based estimator or a model-assisted estimator depending on how the components of the auxiliary variables are used. For pertinent definitions see for example Särndal et al (1992)**.**

We next describe the design, sample estimator and properties shown by Krewski & Rao (1981) for making inference from stratified samples, which we will use in later applications.

Example 2.1 Stratified two-stage probability proportional to size(PPSWR) (Krewski & Rao, 1981). Let $N = \sum_{h=1}^{L} N_h$ be the number of p.s.u.'s in the finite population. For each stratum h, N_h and M_{hi} are respectively the number of p.s.u.'s in the stratum and the number of ultimate units in p.s.u. hi, $i = 1,\dots, N_h$, and $h = 1,\dots, L$. Let $M_h = \sum_{i=1}^{N_h} M_{hi}$ $M = \sum_{h=1}^{L} M_h$. The prior information are the "sizes" of the p.s.u.'s $z_{hi} = M_{hi}$, $i = 1,\dots N_h$, $h = 1,\dots,L$. Suppose $n_h \geq 2$ p.s.u.'s are selected with replacement in stratum h with probabilities $p_{hi} = M_{hi}/M_h$ $i = 1,\dots N_h$, $h = 1,\dots,L$ at each draw. The selection is done independently in each stratum, and independent sub-samples are taken within those p.s.u.'s selected more than once. The finite population mean is $\theta_N = \sum_{h=1}^{L} W_h \theta_h$, where $W_h = M_h/M$, is the stratum weight, $\theta_h = \bar{Y}_h = \sum_{k=1}^{N_h} y_{hk}/M_h$ is the finite population stratum mean and y_{hk} is the total of p.s.u. hk, $k = 1,\dots, N_h$, $h = 1,\dots, L$. Let $\hat{y}_{hk}$ be an unbiased estimator of the total y_{hk} based on sampling at the second stage. Then a sample estimator of the stratum mean θ_h is given by $\sum_{i=1}^{n_h} \hat{\theta}_h^i/n_h$, where

$\hat{\theta}_h^i = \sum_{k=1}^{N_h} \hat{y}_{hk}\, I_{hk}^i/M_{hk}$ and $I_{hk}^i = 1$ if p.s.u. hk is selected in the sample at the i-th draw in stratum h and 0 otherwise, $k = 1,\dots, N_h$. Finally, a design-unbiased sample estimator of the mean θ_N is

$$\hat{\theta}_N(y^N, M^N) = \sum_{h=1}^{L} W_h \sum_{i=1}^{n_h} \hat{\theta}_h^i/n_h$$

We often refer to conditions C_1 to C_4 (Krewski and Rao (1981), p. 1014) in the Appendix for the asymptotic normality of the sample mean $\hat{\theta}_N$.

3. Super-populations

The following definition is similar to the definition of super-population given in Särndal et al (1992, p. 533).

Definition 3.1 Consider a finite population U of size N as in Definition 2.1. A super-population associated with it consists of a probability space $(\Omega, \mathcal{F}, P)$ and random vectors (Y_i, X_i, Z_i), $Y_i:\Omega \to \mathbb{R}^p$, $X_i:\Omega \to \mathbb{R}^k$, $Z_i:\Omega \to \mathbb{R}^q_+$, such that $Y_i(\omega_0) = y_i$, $X_i(\omega_0) = x_i$, $Z_i(\omega_0) = z_i$, for some $\omega_0 \in \Omega$, i = 1,...N . We write $Y^N = (y_i)_{i=1,...,N}$ and define X^N and Z^N similarly. We say that U is a realization of, or is generated by the super-population. A family of distributions of (Y^N, X^N, Z^N) that is given a priori is called a super-population model. We note that different outcomes ω can generate the finite the same finite population.

Example 3.1 Two-stage super-population model. Let Ω be the conceptual population of people like those living at present in a specific country. Suppose that we can conceive it as composed of L disjoint strata of units hi, i = 1,..., N_h, h=1,...,L where unit hi represents a cluster of individuals. Let $(\Omega, \mathcal{F}, P)$ be the corresponding probability space. Now we assume that Z_{hi} are random variables on the probability space that represent the number of individuals that live in cluster hi. We are interested in characteristics Y_{hij} pertaining to the individuals labelled by hij, living in cluster hi $(i = 1,..., N_h, h = 1,..., L)$. In order to be able to define the super-population according to Definition 3.1, we must know an outcome of the Z_{hi}, say for example, the size of the clusters of the population existing right now. Let $F_M = \{\omega \in \Omega: Z_{hi}(\omega) = M_{hi}, i = 1,..., N_h, h = 1,..., L\}$. In this case, we use this prior information to define the super-population model by conditioning on the σ- field $\mathcal{G}$ generated by the event F_M and we use the resulting conditional probability measure to do inference. The conditional probability measure is defined by $P_M(F, \omega_0) = P(F|F_M) I_{F_M}(\omega_0)$ for $F \in \mathcal{F}$, if $P(F_M) > 0$ (see Chow & Teicher, 1977, Equation 3 p. 222). Now we define the super-population on $(\Omega, \mathcal{F}, P_M)$ by random vectors Y_{hij} of p socio-economic characteristics associated with the individual hij, $Y_{hij} : \Omega \to \mathbb{R}^p$, $j = 1,\ldots, M_{hi}$, $i = 1,\ldots, N_h$, $h = 1,\ldots, L$. The cluster totals $\{Y_{hi} = \sum_{j=1}^{M_{hi}} Y_{hij},\ i = 1,\ldots, N_h,\ h = 1,\ldots,L\}$ are assumed independent random vectors and identically distributed within strata

Example 3.2 Finite population with a response model. In this example, the super-population is determined by the response model. Let Π be a finite population associated with data $\{y_1,\ldots, y_N\}$ (see Definition 2.1). Let the N-product $\Omega = \{0,1\} \times \ldots \times \{0,1\}$ describe the collection of response patterns of the N responding units: with $\omega_i = 1$ if unit i responds and zero otherwise. Thus $\omega = (\omega_1,\ldots, \omega_N) \in \Omega$. Suppose the response model for this population is given by $P(\omega_i = 1) = r_i$, $i = 1,\ldots, N$. We define the super-population by $Y_i: \Omega \to \mathbb{R}$, $Y_i(\omega) = y_i\omega_i$, $i = 1,\ldots, N$. For each ω, the finite population $U = U(\omega)$, generated by the outcome ω and the super-population, consists of all N labels and associated data $\{Y_1(\omega),\ldots, Y_N(\omega)\}$. The finite populations $U(\omega)$ and Π will coincide only for the outcome representing complete response

In what follows, the subscript "d" refers to design randomization and we will use "m" to indicate probabilistic properties related to the space $(\Omega, \mathcal{F}, P)$. As a subscript, "m" will indicate convergence in distribution induced by the model space. E_m and V_m denote, respectively, the model expectation and model variance of a random vector.

We now illustrate how conditions that are sufficient for design - based inference in finite populations can be justified as a consequence of simple moment conditions in the super-population, which, in turn, can be justified by expert knowledge of the model. We give below an application related to the work of Krewski & Rao (1981)**.**

Consider the two-stage super-population model of Example 3.1. Here the cluster sizes are non-stochastic in the super-population space. Define $Y^{N_\nu} = (Y_{hi})_{i = 1,\ldots N_{\nu h},\ h = 1,\ldots,L_\nu}$. We assume that the arrays Y^{N_ν} are nested as N_ν increases. The number of strata $L_\nu \to \infty$, as $\nu \to \infty$. We also assume, defined on the finite population generated by $\omega \in \Omega$, the sampling design of Example 2.1 where $Y_{hi}(\omega) = \sum_{j=1}^{M_{hi}} Y_{hij}(\omega)$, $\omega \in \Omega$. The finite population means are $\theta_{N_\nu} = Y_\nu / M_\nu$ and the corresponding sample estimators of Example 2.1 are $\hat{\theta}_{N_\nu} = \sum_{h=1}^{L_\nu} W_h \hat{\theta}_h$. We show that moment conditions in the super-population yield condition C_1 of Krewski & Rao for asymptotic normality of the sample mean (in the law of the design). We omit indexing the populations.

Proposition 3.1 We assume that no strata is of disproportionate size (Condition C_3). Let $n = n_1 + n_2 + ... + n_L$. Thus $n \to \infty$ as $L \to \infty$. If in addition, we assume the model - based conditions:

$$(M_1) \quad \sum_{h=1}^{L} W_h E_m |Y_{h1}|^{2+\delta} = O(1) \text{ as } n \to \infty, \text{ and } \sum_{h=1}^{\infty} \frac{V_m |Y_{h1}|^{2+\delta}}{h^2} < \infty, \ \delta > 0,$$

then condition C_1 holds

$$(C_1) \quad \sum_{h=1}^{L} W_h E_d |\hat{\theta}_h^i - \bar{Y}_h|^{2+\delta} = O(1) \text{ a.s. } \omega,$$

as $n \to \infty$, where $\hat{\theta}_h^i$ is the estimator of the stratum mean based on the *i-th* draw of stratum h, $1 \le i \le n_h$, defined in Example 2.1. The proof is given in the Appendix

4. The product space

In this section, we first define a measurable space (a product space) that includes the super-population and the design and we define the product probability measure $P_{d,m}$. We then present the conditional probabilities given the design and given the model (Propositions 4.1 and 4.2 respectively), which we require to show that we cannot ignore the design even for self-weighted designs (Example 4.5), and to prove that convergence in design implies convergence in the product space above mentioned (Theorem 5.1). We assume that the size of the finite population is not dependent on the outcome of the super-population. Let p be a design and let $(S, C(S), p)$ be a probability design space defined on the finite population. Recall that once a sampling scheme is determined, for the space S of all possible samples to be well defined, it is necessary to know the number of ultimate units of the population, as well as the number and size of strata, p.s.u.'s, secondary sampling units (s.s.u)'s, etc.

Definition 4.1 Consider a finite population of size N generated by a super-population (Y^N, X^N, Z^N) as in Definition 3.1. We define the product space as a measurable space given by $S\times\Omega$ with the σ-field $C(S)\times\mathcal{F}$.

Next, we show that a design and sample estimator can be viewed as random variables in the product space and we define a probability measure on the product space $(S \times \Omega, C(S) \times \mathcal{F})$.

Definition 4.2 Consider a super-population associated with a finite population as in Definition 4.1. Let $p : C(S) \times \mathbb{R}_+^{q\times N} \to [0,1]$ be a design on the finite population as in Definition 2.3. Let us assume that the range of the Z^N is contained in the domain of the design, i.e., $R(Z^N) \subseteq D(z^N)$. Then the design can be viewed as a random variable $p_{d,m}$ on $(S \times \Omega, C(S)\times\mathcal{F})$ defined by

$$p_{d,m}(s, \omega) = p(s, Z(\omega)), \quad \omega \in \Omega, \; s \in S. \tag{4.2}$$

Definition 4.3 We define $P_{d,m}$ as the σ-additive measure that on elementary rectangles of the product σ - field, has the value

$$P_{d,m}(\{s\}\times F) \equiv \int_F p_{d,m}(s, \omega)\, dP, \qquad s \in S,\; F \in \mathcal{F} \tag{4.3}$$

$P_{d,m}$ is well defined because all sets in $C(S)\times\mathcal{F}$ can be expressed as a finite union of elementary rectangles. In particular, $P_{d,m}(S \times \Omega)=1$. Hence $P_{d,m}$ is a probability measure on the product space.

Remark 4.1 If $\hat{\theta}_N$ is a sample estimator on the design space (S, C(S), p) with associated super-population (Y^N, X^N, Z^N) and the range of the super-population is contained in the domain of $\hat{\theta}_N$, then the sample estimator can be viewed as a random variable on the product space defined by:

$$\hat{\theta}_N(s, \omega) = \hat{\theta}_N(s, Y^N(\omega), X^N(\omega), Z^N(\omega)), \quad \omega \in \Omega,\; s \in S. \tag{4.4}$$

Example 4.1 Stochastic dependence in the product space. Let Y^N and p denote, respectively, the super-population and design of Definition 4.1, with Y^N composed of N independent (not necessarily) identically distributed random variables and p a design associated with a SRSWOR or a SRSWR scheme of sample size n. Let $y^s = \{y_i,\ i \in s\}$ denote the values of y^N associated with the units i in a sample s $\in$ S. We define the k-th draw-selection indicators $\{I_1^k,\ ...,\ I_N^k\}$ by $I_i^k(s) = 1$ if unit i is selected in the sample at the k-th draw and zero otherwise, for k=1,..., n. Then y^s can be written as the sequence of n units

$$y^s = \Big(\sum_{i=1}^{N} y_i I_i^1(s),\ \sum_{i=1}^{N} y_i I_i^2(s),...,\ \sum_{i=1}^{N} y_i I_i^n(s)\Big).$$

Each coordinate of the sequence represents the result of a draw. If the design is WR then $I_i^k(s)$ and $I_j^\ell(s)$ are d-stochastically independent for $k \neq \ell$ and all i, j, whereas if the design is WOR the $I_i^k(s)$ and $I_j^\ell(s)$ are design-stochastically dependent and if $I_i^k(s) = 1$ then $I_i^\ell(s) = 0$ for all $\ell \neq k$. Hence the y^s can be viewed as a vector of random variables W_k in the product space: $W_k(s,\omega) \equiv \sum_{i=1}^{N} Y_i(\omega) I_i^k(s),\ k = 1,...,\ n$. If the random variables in Y^N are identically distributed, under SRSWOR the W_k are stochastically independent in the product space, whereas under SRSWR the W_k are stochastically dependent as variables in the product space. Now, if the components of Y^N are not identically distributed, whether the design p is SRSWOR or SRSWR, the W_k are stochastically dependent random variables in the product space. See the Appendix for the proof

Example 4.2 Two-stage super-population model and two stage design. We assume the two-stage super-population model of Example 3.1 defined on (Ω, $\mathcal{F}$, P_M) and recall that we use as prior information the size of the clusters of a population existing right now to define the super-population model. This minimum necessary information is contained in $F_M = \{\omega \in \Omega:\ Z_{hi}(\omega) = M_{hi},\ i = 1,...,\ N_h,\ h = 1,...,L\}$, where the M_{hi} are as in Example 3.1. We select the sample with probability proportional to those sizes, but we want to make conclusions about a more general population than the finite population living in those clusters now. We set $M^{N_h} = (M_{hi})_{i=1,...,N_h}$. Once the model is defined, we first define a sample space S, as the collection of all possible "stratified clustered" sequences of units (see Section 2) of a finite population associated with the super-population model. We then define a stratified two-stage sampling design $p(s, M^{N_1},..., M^{N_L})$ with L strata, N clusters and M ultimate units. We can then construct the product space S × Ω with probability measure $P_{d,m}$ defined on the elementary rectangles by $P_{d,m}(s \times F) = p(s, M^{N_1},..., M^{N_L})P_M(F)$, for $s \in S$ and $F \in ö$

Example 4.3 Product space for inference on a finite population parameter in the presence of non-response. Let U(ω) be a finite population generated by the super-population of Example 3.2. The mean under complete response $\bar{Y} = \sum_i y_i / N$ is the parameter of interest. Without some model assumptions, there is no unbiased or consistent estimator of $\bar{Y}$. However, if sampling and response mechanisms are considered independent of each other and $\hat{\theta} = (1/N)\sum_{i \in S} y_i \omega_i / \pi_i$ is design-unbiased for $\bar{Y}(\omega) = (1/N)\sum_{i=1}^{N} y_i \omega_i$, then $\hat{\theta}_r = (1/N)\sum_{i \in S} y_i \omega_i / \pi_i r_i$ is unbiased for $\bar{Y}$ in the law of the product space. Here π_i is the probability of selecting unit i, $i = 1,... N$

Example 4.4 Parametric distribution of the sample data (Pfeffermann & Sverchkov, 1999)Let us assume that the selection probabilities depend on the cluster sizes, which can be correlated with Y^N (See Definition 3.1) and the auxiliary information X^N, and that we have the two-stage super-population model of Example 3.1, and the product space defined in Example 4.2. The "parametric distribution of the sample data" proposed by Pfeffermann & Sverchkov (1999), which they use to do inference, can be thought of as the conditional probability measure given a sample and the auxiliary data, i.e., $P_{d,m}(\cdot \mid õ)$ where the field õ is generated by the event $\{s_0\} \times F_x$, with $s_0 \in S$ and $F_x = \{\omega \in \Omega: X_i(\omega) = x_i, i \in s_0\}$. Hence the "parametric distribution of the sample data" would given by

$$P_{d,m}(s \times F \mid õ)(\omega) = P_M(F \mid F_x)(\omega),$$

where $F \in ö$, if $s = s_0$, $\omega \in F_x \cap F_M$ and $P(F_x \cap F_M) > 0$

Proposition 4.1 We denote by P_{m*d} the conditional probability on the product space given the field $C(S) \times \Omega$. On each set $B \in C(S) \times \Omega$, $B = \bigcup_{s \in A} \{s\} \times F_s$, $A \in C(S)$, $F_s \in ö$, we have:

$$P_{m*d}(B\ ,\ s_0)\ =\ P_{d,m}(s_0\ \times\ F_{s_0}) = P_{d,m}(s_0\ \times\ \Omega) \tag{4.5}$$

if s_0 is contained in A, and 0 otherwise. If, in particular, p(s,ω) does not depend on ω, then $P_{m*d}(B\ ,\ s_0)\ =\ P(F_{s_0})I_A(s_0)$, where $I_A(s_0)$ is the value of the indicator function of the set A at s_0. The proof is immediate from the definitions since the conditioning field is the field generated by the partition of the product space into the finite number of sets $\{s{\times}\Omega\ :\ s\ \in\ S\}$ (see Chow and Teicher (1997) Example 1, Section 7.2) $\square$

Remark 4.2 Note that Proposition 4.1 does not imply that if p(s,ω) does not depend on ω we can ignore the design, since information about the selection indicators is contained in the factor $I_A(s_0)$. Example 4.5 below illustrates this point. We also note that the probability measure $P_{m*d}(\ \cdot\ ,\ s_0)$ coincides with the conditional product probability measure given only one sample $s_0 \times \Omega$, when evaluated at s_0: $P_{m*d}(\ \cdot\ ,\ s_0)\ = P_{d,m}(\ \cdot\ |s_0\times\ \Omega)(s_0)$ (see Chow & Teicher, 1997, equation 12, p 215 and Definition, p.223) . Hence the Example 4.5 is valid for both the conditional probability given the entire design and the conditional probability given only one sample s_0.

Example 4.5 Stochastic independence of the sample under P_{m*d}. In the context of Example 4.1, with SRSWOR, under P_{m*d}, the W_k- variables inherit the independence of the original $Y_i,\ i = 1,...,\ N$. A WOR design implies that there are no repetitions in the sample, so the sample is a subset of the $Y_i,\ i = 1,...,\ N$. On the other hand, for SRSWR, under P_{m*d}, the W_k variables do not retain the independence of the $Y_i,\ i = 1,...,\ N$, if the selected sample has repeated labels. For illustration of the mechanism, see the Appendix

Example 4.6 Asymptotic normality in the projected space. Consider a sequence of super-populations associated with finite populations as in Definition 3.1. We assume that $Y_{\nu i}$, i = 1, $\ldots N_\nu$, $\nu \geq 1$ are i.i.d. random variables's on (Ω, $\mathcal{F}$, P) with 0 mean and finite second moment $\sigma^2 \neq 0$. The design is SRSWOR of size n_ν , ν = 1,.... . Given a selected sample s_ν, the space $(S_\nu{\times}\Omega,\ C(S_\nu){\times}\mathcal{F},\ P_\nu)$, $P_\nu = P_{m*d}(\cdot\ ,\ s_\nu)$, is the "projected" probability space onto the model . Let us denote by $W_{\nu k}(s_\nu,\omega)$ the array of random variables obtained from the original $Y_{\nu i}$ and the sample defined as in Example 4.1. Then $(\sigma^2 n_\nu)^{-1/2}[\sum_{k=1}^{n_\nu} W_{\nu k}]$ converges in the law of P_ν to a standard normal random variable. Since each $W_{\nu k}(s_\nu,\omega)$ is equal to one $Y_{\nu i}$, $i = i(k)$, we could write that $(\sigma^2 n_\nu)^{-1/2}[\sum_{i \in s_\nu} Y_{\nu i}]$ is asymptotically normal N(0,1). See the Appendix for the proof

We define now the conditional probability given the σ-field $S \times \mathcal{F}$. It represents the change in $P_{d,m}$ when we have the additional information given by the complete model space $(\Omega, \mathcal{F}, P)$. Note that the information contained in $S \times \mathcal{F}$ is richer than the information contained in the (super-population) model given by $S \times \mathcal{F}_N$, where $\mathcal{F}_N = \sigma(Y^N, X^N, Z^N)$ is the σ-field generated by the super-population.

Proposition 4.2 Let $B = \bigcup_{s \in A} \{s\} \times F_s \in C(S) \times \mathcal{F}$ with all s distinct. We define the set function :

$$P_{d|m}(B, \omega) = \sum_{s \in A} p_{d,m}(s,\omega) I_{F_s}(\omega), \quad \omega \in \Omega. \tag{4.6}$$

Then $P_{d|m}(B, \omega)$ is the (regular) conditional probability measure on $(S \times \Omega, C(S) \times \mathcal{F})$ given the σ - field $S \times \mathcal{F}$. The proof is given in the Appendix (

Remark 4.3 Proposition 4.2 is also valid if we replace everywhere $\mathcal{F}$ by $\mathcal{F}_N = \sigma(Y^N, X^N, Z^N)$ or by $\sigma(Z^N)$. We note that given an outcome $\omega_0 \in \Omega$, $z^N = Z^N(\omega_0)$, the design probability can be "recovered" as a version of P_{d*m} in the following sense: for $A \in C(S)$ and $F_z \equiv \{\omega \in \Omega: Z^N(\omega) = z^N\}$, we have $p(A, z^N) = P_{d*m}(A \times F_z, \omega_0)$.

Consider now a one-stage super-population model composed of L disjoint strata of N_h vectors $(\boldsymbol{Y}_{hi}, Z_{hi})$ for each h =1,..., L. Say, for example, the Y_{hi1}, Y_{hi2}, Z_{hi} are respectively the labour cost , workforce size and annual revenue of business i in stratum h, $i = 1,...,N_h$, $h = 1,...,L$. Suppose the revenue values $z = (z_{hi})$ correspond to an outcome $\omega \in \Omega$ that has occurred. Since the elements in $\boldsymbol{Y}_{hi}$ are somewhat correlated with the revenue, we may select the sample with probability proportional to those revenues, and as in Example 4.2 we might want to learn about businesses in a dynamic population rather than the collection of existing businesses having that revenue now. In this case, we can build the product space before we condition on the revenue outcome, then we may use that outcome, the "prior" information,

$$F_z = \{\omega \in \Omega:\ Z_{hi}(\omega) = z_{hi},\ i = 1,...,\ N_h,\ h = 1,...,\ L\}$$

by conditioning on the field $\mathcal{G}$ generated by the event $S \times F_z$ and we may use the conditional probability measure $P_{d,m}(\cdot|\mathcal{G})$ to do inference.

Proposition 4.3. Conditioning on the prior information. $P_{d,m}(\cdot|\mathcal{G})$ is a regular conditional probability measure and it is constant on $S \times F_z$: for B as in Proposition 4.2 and $\omega_0 \in F_z$, $P_{d,m}(\cdot|\mathcal{G})(s_0,\ \omega_0) = P_{d,m}(B)$ if $P(F_z) = 0$ and if $P(F_z) > 0$, $P_{d,m}(\cdot|\mathcal{G})(s_0,\omega_0) = \sum_{s \in A} p_{d,m}(s,\ z) P(F_s|F_z)$.

This follows from the fact that $\mathcal{G}$ is generated by a partition of S × Ω and Example 1, section 7.2, Chow and Teicher (1997).

5. Convergence in the Product Space and Asymptotic Independence

In this Section we establish results that enable us to determine the limiting distribution of a combination of sample estimators and super-population statistics. We show that convergence in the law of the super-population or in the design law implies convergence in the law of the product space and that under certain conditions the two statistics are "asymptotically independent".

Remark 5.1 Let $(S \times \Omega,\ C(S) \times \mathcal{F},\ P_{d,m})$ be the product space of Definitions 4.1 and 4.3 and $\hat{\theta} \in \mathbb{R}$ an estimator defined on the corresponding design-space. The design-distribution of $\hat{\theta}$ is the $\mathcal{F}_N$-measurable random variable :

$$F(t,\omega) \equiv p(\{s \in S:\ \hat{\theta}(s,\omega) \le t\},\ \omega), \quad t \in \mathbb{R}.$$

With $\mathcal{F}_N = \sigma(Y^N, X^N, Z^N)$ in Proposition 4.2, we note that $F(t,\omega)$ coincides with a version of the conditional probability given in formula (4.6): $F(t,\omega) = P_{d|m}(B,\ \omega)$, where $B = \{(s,\omega):\ \hat{\theta}(s,\omega) \le t\}$, $t \in \mathbb{R}$. Hence by definition (see Chow and Teicher (1997), p.225)

$F(t,\omega)$, $t \in \mathbb{R}^R$, is a version of the conditional distribution of $\hat{\theta}$ given $S \times \mathcal{F}_N$.

Theorem 5.1 We consider a sequence of the product spaces and sample estimators as in Remark 5.1 indexed by $\nu \geq 1$. Let λ_ν, $\lambda \in \mathbb{R}^R$ be random vectors defined on $(\Omega, \mathcal{F}, P)$. We have: (i) If $\lambda_\nu \to \lambda$ in the law of the model (P), then $\lambda_\nu \to \lambda$ in the law of the product space.

(ii) If $F_\nu(t, \omega) \to F(t, \omega)$ in probability P, for all points of continuity $t \in \mathbb{R}^R$ of $F(t, \omega)$, then $F(t, \omega)$ is a bounded random variable in the model space, and the product-space distribution of $\hat{\theta}_\nu$ converges to $F_\theta(t) = \int_\Omega F(t, \omega)\, dP(\omega)$. In particular, if $\hat{\theta}_\nu$ is design-consistent a.s. $\omega \in \Omega$, then it is consistent in the product space.

(iii) If $\lambda_\nu \to \lambda$ in the law of the model (P) and $F_\nu(t, \omega) \to F_\theta(t)$ in probability P as $\nu \to \infty$, for all points of continuity $t \in \mathbb{R}^R$ of $F_\theta(t)$ where $F_\theta(t)$ is a non-stochastic distribution function, then the joint distribution function of $(\hat{\theta}_\nu, \lambda_\nu)$ converges to the product of the two limiting distributions. The random variables $\hat{\theta}_\nu$ and λ_ν are said to be asymptotically independent. Note that when the limiting design-based distribution is normal with mean zero, we only require that the limiting variance be non-stochastic in the model. This last condition would follow from simple conditions in the super-population model. The proof is given in the Appendix

6. Sample estimators derived from an Estimating Equation (EE)

In this section we describe a methodology to derive the asymptotic normality of the sample estimating equation estimator $\hat{\theta}_N \in \mathbb{R}^R$ when referred to a super-population parameter $\theta_0 \in \mathbb{R}^R$. As for example 5.1, it consists of combining existing asymptotic results in both the design and super-population probability spaces, under the umbrella of the product space, and the application of Theorem 5.1.

We first define an estimating equation estimator for the super-population set up. Let $(\Omega, \mathcal{F}, P)$ and

$(Y^{N\nu}, X^{N\nu}, Z^{N\nu})$ represent a super-population as in Definition 3.1. In what follows N_ν denotes the number of stochastically independent vectors in the super-population. Given a design p_ν, the sample size (or first-stage sample size) n_ν denotes the number of stochastically independent units in the sample. Let g represent continuous functions defined on $\mathbb{R}^{p+k+R}$. We consider functions of the form

$$G_{N_\nu}(\theta,\omega) = 1/\alpha(N_\nu) \sum_{i=1}^{N_\nu} g(Y_i(\omega), X_i(\omega), \theta) \tag{6.1}$$

where $\omega \in \Omega, \theta \in \mathbb{R}^R, g \in \mathbb{R}^R, \alpha(N_\gamma)/N_\nu = O(1)$ *as* $\nu \to \infty$. A finite population EE is defined by

$$G_{N_\nu}(\theta,\omega) = 0. \tag{6.2}$$

Definition 6.1 An (finite population) EE estimator is defined as a solution θ_{N_ν} of the finite population estimating equation: $G_{N_\nu}(\theta_{N_\nu},\omega) = 0$. For $\omega \in \Omega$ fixed, θ_{N_ν} is a finite population parameter.

Yuan and Jennrich (1997) set very general conditions that lead to the existence, strong consistency and asymptotic normality of estimating equation estimators. The super-population models used by Yuan and Jennrich require independent but not necessarily identically distributed random vectors $g(Y_i, X_i, \theta)$. We could also apply their results to clustered data models where we can add up the vectors within a cluster (i.e. $g_i(\theta) = \sum_{j=1}^{M_i} g(Y_{ij}, X_{ij}, \theta)$, the cluster totals $g_i(\theta)$ are stochastically independent, and the cluster sizes M_i stay bounded as the number N_ν of clusters increase towards infinite.

Now let $(S_\nu, C(S_\nu), p_\nu)$ be a design probability space where the (fixed) first stage expected sample size is n_ν. Let $\hat{G}_{N_\nu}(\theta,\omega)$ be a design-consistent estimator of $G_{N_\gamma}(\theta,\omega)$. A sample EE is defined by

$$\hat{G}_{N_\nu}(\theta,\omega) = 0. \tag{6.3}$$

Definition 6.2 A sample EE estimator $\hat{\theta}_{N_\nu}$ is defined as a solution of the sample EE in (6.3).

Theorem 6.1 shows that the sample EE estimator (around the model parameter) is asymptotically normal in the law of the product space. Conditions (1) to (3) are the conditions given by Yuan and Jennrich (1998) for the existence, consistency of θ_N and asymptotic normality of

$$\sqrt{N}(\theta_N - \theta_0).$$

We will see that conditions (1), (4) and (5) below imply the existence, design-consistency of $\hat{\theta}_N$ and design-asymptotic normality of $\sqrt{n}(\hat{\theta}_N - \theta_N)$.

Theorem 6.1 Let $(\Omega, \mathcal{F}, P)$ and $(Y^{N_\nu}, X^{N_\nu}, Z^{N_\nu},)$ denote a sequence of super-population composed by N_ν independent random vectors and let $(S_\nu, C(S_\nu), p_\nu)$ be a sequence of design spaces defined on finite populations U_ν generated by the corresponding super-populations defined above. Let $N_\nu \to \infty$ as $\nu \to \infty$, and let n_ν be the first stage fixed sample size (or expected value of the first stage sample size). Note that N_ν, n_ν, S_ν and p_ν depend on the index ν, but we omit it in what follows for the sake of simplicity. Let $f = \lim_n n/N > 0$ as $\nu \to \infty$. Note that we do not require that f=0. We assume the following conditions:

1. $G_N(\theta_0) \to 0$ with probability one.
2. There is a compact neighbourhood $B(\theta_0)$ of θ_0 on which, with probability one, all $G_N(\theta)$ are continuously differentiable and the Jacobians $\dot{G}_N(\theta)$ converge uniformly to a non-stochastic limit $J(\theta)$ which is non-singular at θ_0.
3. $\sqrt{N}\, G_N(\theta_0) \Rightarrow N(0, \Gamma_m)$ in the law of the super-population.
4. There is a compact neighbourhood $B(\theta_0)$ of θ_0 on which $\partial \hat{G}_N(\theta)/\partial\theta$, which are assumed continuous, converge uniformly in design probability to a non-stochastic (in design) limit which coincides with $J(\theta)$ at θ_0 for almost every $\omega \in \Omega$. Note that if the $\hat{G}_N(\theta)$ is linear in the g_i's and all the $G_N(\theta)$ are continuously differentiable then all the $\hat{G}_N(\theta)$ are too.
5. $\sqrt{n}\, \hat{G}_N(\theta_N) \Rightarrow N(0, \Gamma_d)$ in the law of the design as $n \to \infty$ for almost every $\omega \in \Omega$, where the variance matrix Γ_d is non-stochastic in the super-population.

Then

$$\sqrt{n}(\hat{\theta}_N - \theta_0) \Rightarrow N(0, \Gamma) \tag{6.4}$$

in the law of the product space $(S \times \Omega, C(S) \times \mathcal{F}, P_{d,m})$, where for $J = J(\theta_0)$,

$$\Gamma = J^{-1}[\Gamma_d + f\, \Gamma_m]\, J^{-1}. \tag{6.5}$$

Proof: For simplicity we assume that $f = n/N$ for all n.

$$\sqrt{n}(\hat{\theta}_N - \theta_0) = \sqrt{n}(\hat{\theta}_N - \theta_N) + \sqrt{f}\sqrt{N}(\theta_N - \theta_0) \tag{6.6}$$

Condition 1 to 3 imply the asymptotic normality of the second term on the right hand side of (6.6), in the law of the model (see Yuan and Jennrich (1998)). This and Theorem 5.1 (i) , in turn imply convergence in the law of the product space. Next we observe that $\hat{\theta}_N$ exists and $\hat{\theta}_N - \theta_N \to 0$ in design probability, as $n \to \infty$ for almost every $\omega \in \Omega$. Indeed, the $\hat{G}_N(\theta)$ are continuously differentiable and design consistency implies that $\hat{G}_N(\theta)$ converges to $G(\theta)$ (the limit of $G_N(\theta)$ in Yuan & Jennrich) in design probability. Hence we can apply to $\hat{G}_N(\theta)$ the same techniques of Theorems 1 and 2 of Yuan and Jennrich (1998), and thus conditions 1 and 4 imply both the existence of $\hat{\theta}_N$ and $\hat{\theta}_N \to \theta_0$ in design probability. Since the above mentioned Theorems 1 and 2 imply also $\theta_N \to \theta_0$, almost surely in the model probability P, we have $\hat{\theta}_N - \theta_N \to 0$ in design probability a.s. $\omega \in \Omega$. Now , following the reasoning of Yuan and Jennrich (1998), conditions 4 and 5 imply asymptotic normality of the first term in the right hand side of (6.6) (see also Binder (1983)). This in turn implies convergence in the product space, by Theorem 5.1 (ii) However these two terms above are not stochastically independent in general. Theorem 5.1 (iii) and condition 5 imply the "asymptotic independence" of the terms and the asymptotic normality of the sum

Example 6.1 The ratio estimator of the finite population mean. We assume a nested one-stage super-population model composed of L disjoint strata of N_h i.i.d.r.v.'s (Y_{hi}, Z_{hi}), $i = 1,\ldots, N_h$, with mean $\mu_h = E_m(Y_{h1})$ and variance $\sigma_h^2 = V_m(Y_{h1})$, $h = 1,\ldots, L$. Let $\mu_N = \frac{1}{N}\sum_{h=1}^{L} N_h\mu_h$ be the parameter of interest, $N = N_1+\ldots+N_L$, and $\Gamma_N = \frac{1}{N}\sum_{h=1}^{L} N_h\sigma_h^2$. The finite population EE and finite population EE estimator are respectively,

$$G_N(\mu_N, \omega) = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} g_{hi}(\mu_N, \omega) = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h}(Y_{hi}(\omega)-\mu_N), \quad \bar{Y}_N(t) = \frac{1}{N}\sum_{h=1}^{L}\sum_{i=1}^{N_h} Y_{hi}.$$

Under a stratified one-stage p.p.s.-design, units in sample s_h are selected with probabilities $p_{hi} = p(Z_{hi}, i = 1,\ldots, N_h)$ with replacement, $h = 1,\ldots, L$. The sample EE, with $y_{hi} = Y_{hi}(\omega)$, is given by

$$\hat{G}_N(\mu_N, \omega) = (1/N)\sum_{h=1}^{L}\sum_{i\in s_h}(y_{hi}-\mu_N)/n_h p_{hi},$$

and the corresponding sample EE is the ratio estimator,

$$\bar{y}_R = (1/\hat{N})\sum_{h=1}^{L}\sum_{i\in s_h} y_{hi}/n_h p_{hi}, \quad \hat{N} = \sum_{h=1}^{L}\sum_{i\in s_h} 1/n_h p_{hi}.$$

Let $n = n_1+\ldots+n_L$ and $f = \lim_n n/N$. We assume $n \to \infty$ if and only if $N \to \infty$. We aim to obtain the asymptotic normality of $\sqrt{n}(\bar{y}_R - \mu_N)$ as $n \to \infty$. Here we construct a product space with the unconditional model probability measure P rather than $P(\cdot|\sigma(F_z))$ of Example 4.2. We decompose $\sqrt{n}(\bar{y}_R - \mu_N)$ into the two terms in the right hand side of (1.1), and investigate the conditions necessary for asymptotic normality and stochastic independence of the two terms. In the simple case of the ratio estimator of the mean, existence of both the finite population and sample EE estimators follows from the exact solutions of the EE. Their consistence follows from Assumption 1 and the fact that Assumptions 2 and 4 hold for the entire parameter space. We only require to verify Assumptions 1, 3 and 5 of Theorem 6.1.

$G_N(\mu_N)\to 0$ a.s. follows from the Strong Law of Large Numbers (SLLN) for $X_{hi}=Y_{hi}-\mu_N$, which holds if $\sum_{h=1}^{\infty}\sigma_h^2 b_h<\infty$, with $b_h=1/(N_1+...+N_{h-1}+1)^2+...+1/(N_1+...+N_h)^2$ (Theorem 22.4, Billingsley, 1995).

The CLT for $\sqrt{N}(\bar{Y}_N - \mu_N)$ with variance $\Gamma_m = \lim_N \frac{1}{N}\sum_{h=1}^{L} N_h\sigma_h^2$ follows from Liapunov's condition (Theorem 27.3, Billingsley, 1995) and the existence of $\Gamma_m<\infty$:

$$\sum_{h=1}^{L} N_h E_m|Y_{h1} - \mu_h|^{2+\delta} = o(N^{1+\frac{\delta}{2}}\Gamma_N^{1+\delta/2}) \text{ as } n\to\infty, \text{ for some } \delta>0.$$

The CLT for $\sqrt{n}(\bar{Y}_R - \bar{Y}_N)$ with variance $\Gamma_d = f\,\lim_n(1/N)\sum_h(\sum_i e_{hi}^2/n_h p_{hi} - e_h^2/n_h)\approx 0$ constant a.s. ω follows from assuming conditions $C_2^{\prime}, C_3^{\prime}$ below, conditions C_1, C_4 in the Appendix applied to the residuals $e_{hi}=Y_{hi} - \bar{Y}_N$, $e_h=\sum_i e_{hi}$ (see Yung, W. and Rao, J.N.K., 2000) and that Γ_d be constant a.s.:

$C_2^{\prime}$ $(n/N)\max_{h,i}1/n_h p_{hi}=O_P(1)$ as $n\to\infty$.

$C_3^{\prime}$ $\max_h N_h/N = O(L^{-1})$ as $n\to\infty$

Example 6.2 General EE estimator under a stratified two-stage super-population model and design. Binder (1983) established sufficient conditions for the asymptotic normality of the sample EE estimator $\hat{\theta}_N$ in the design probability space (Assumption 5 of Theorem 6.1). These conditions are very general and some depend on the particular design in consideration. We examine here the design conditions for the asymptotic normality of $\sqrt{n}\,\hat{G}_N(\theta_N)$ to hold in the finite population, and the sufficient assumptions in the super-population that imply these design conditions, when the product space is that of Example 4.2. We assume the stratified two stage super-population model of Example 3.1 with the addition of the auxiliary information given by X_{hij}, (h, i, and j as in Example 3.1). Thus the finite population estimating equation is given by

$$G_N(\theta) = \frac{1}{M}\sum_{h=1}^{L}\sum_{i=1}^{N_h} g_{hi}(\theta)$$

where $g_{hi}(\theta) = \sum_{j=1}^{M_{hi}} g(Y_{hij}, X_{hij}, \theta)$, $M_h = \sum_{i=1}^{N_h} M_{hi}$, $N = \sum_{h=1}^{L} N_h$. Let us consider the stratified probability proportional to size design of Example 2.1 and of Krewski and Rao (1981). In addition, we assume condition C_2 in the Appendix. Let $\hat{G}_N(\theta)$ be the sample estimator of $G_N(\theta)$ described in Example 2.1 and assume that conditions 1 and 2 of Theorem 6.1 hold. Since $G_N(\theta_N) = 0$, we have

$$\hat{G}_N(\theta_N) = \hat{G}_N(\theta_N) - G_N(\theta_N) \qquad (6.7)$$

and hence we can express (6.7) as the sum of $n = n_1 + \ldots + n_L$ independent random vectors $Z_{hi}(\theta_N)$ with mean zero:

$$\hat{G}_N(\theta_N) = \sum_{h=1}^{L} W_h \frac{1}{n_h} \sum_{i=1}^{n_h} Z_{hi}(\theta_N)$$

with $W_h = M_h/M$, $Z_{hi}(\theta_N) = \sum_{k=1}^{N_h} \hat{g}_{hk}(\theta_N)\, I_{hk}^{i}/M_{hk} - \sum_{k=1}^{N_h} g_{hk}(\theta_N)/M_h$, where $\hat{g}_{hk}(\theta_N)$ denotes the second stage unbiased sample estimator of $g_{hk}(\theta_N)$.

Let $\Gamma_h = V_d(Z_{hi}(\theta_N))$ $i = 1, \ldots, N_h$ and let $\Gamma_d = \lim n \sum_{h=1}^{L} W_h^2\, \Gamma_h / n_h$, which is assumed to exist. We assume conditions C_1 to C_4 hold for the vector sample mean $\hat{G}_N(\theta_N)$. This implies that Liapunov's condition for the CLT for arrays is satisfied (see Theorem 27.3 for scalars, Billingsley, 1995 and apply it to $\{\lambda' Z_{hi}(\theta_N)\ i = 1, \ldots, n_h, h = 1, \ldots, L\}_{L \geq 1}$ for every $\lambda \in \mathbb{R}^R$, $\lambda \neq 0$). Thus $\sqrt{n}\, \hat{G}_N(\theta_N) \to N(0, \Gamma_d)$ in the law of the design.

Now, Proposition 3.1 gives conditions in the super-population for condition C_1 to hold. Recall that for proposition 3.1 to hold, we assume that the super-population is nested as $N \to \infty$. Note that C_3 translates directly to the super-population space, given that in $(\Omega, \mathcal{F}, P_M)$ the cluster sizes are non-stochastic. Condition C_4 about the limiting variance Γ_d and Theorem 5.1's condition that it be a non-stochastic positive definite matrix, require $n/N \to f > 0$, $M/N \to \ell < \infty$ as $n \to \infty$ and conditions in the super-population that are more complex than those stated in Proposition 3.1, but they can be derived in the same way . We do not spell them out here

APPENDIX

Krewski-Rao (1981) designs conditions for the asymptotic normality of the sample mean.

(C_1) $\quad \sum_{h=1}^{L} W_h E_d|\hat{\theta}_h^i - \bar{Y}_h|^{2+\delta} = O(1)$ as $n \to \infty$, $\hat{\theta}_h^i$ as in Example 2.1.

(C_2) $\quad (n/N)\max_{h,i} m_{hi} w_{hij} = O(1)$ as $n \to \infty$ where m_{hi} are the 2nd stage sample sizes and w_{hij} are the sampling weights.

(C_3) $\max_h W_h = M_h/M = O(L^{-1})$, which implies that no strata is of disproportionate size. Here $M = \sum_{h=1}^{L} M_h$, where M_h is the number of ultimate units in stratum h.

(C_4) $\Gamma_d^N(\omega) = n\ V_d(\hat{\theta}_N) \to \Gamma_d$ as $n \to \infty$, $\hat{\theta}_N$ as in Example 2.1.

Proof of Proposition 3.1

Let $\mathcal{E}_d(\omega) = \sum_{h=1}^{L} W_h E_d|\hat{\theta}_h^i - \bar{Y}_h|^{2+\delta}$, $\delta > 0$. We have to show that $\mathcal{E}_d(\omega)$ stays bounded as $n\to\infty$, for the ω generating the finite population. If $p>1$, then

$$E_d|(1/N)\sum_{k=1}^{N} X_k|^p \le (1/N)\sum_{k=1}^{N} E_d|X_k|^p \qquad \text{(A.I)}$$

(see Chow and Teicher, 1997, p.107). We have, setting N=2 and p=2+δ above,

$$\mathcal{E}_d(\omega) \le 2^{2+\delta}\sum_{h=1}^{L} W_h(E_d|\hat{\theta}_h^i|^{2+\delta} + |\bar{Y}_h|^{2+\delta}). \qquad \text{(A.II)}$$

Now, $\hat{\theta}_h^i$ is as in Example 2.1, thus only one term of $\hat{\theta}_h^i$ is non-zero; hence for any i=1,..., n_h

$$E_d|\hat{\theta}_h^i|^{2+\delta} = \sum_{k=1}^{N_h} |Y_{hk}(\omega)/M_{hk}|^{2+\delta}\ p_{hk} \le (1/N_h)\sum_{k=1}^{N_h} |Y_{hk}(\omega)|^{2+\delta} \qquad \text{(A.III)}$$

since $M_{hk} \ge 1$ and $M_h \ge N_h$, k=1,..., N_h , h=1,...,L. Similarly, by (A.I) with $N=N_h$,

$$|\bar{Y}_h|^{2+\delta} = |(1/M_h)\sum_{k=1}^{N_h} Y_{hk}(\omega)|^{2+\delta} \le (1/N_h)\sum_{k=1}^{N_h} |Y_{hk}(\omega)|^{2+\delta} \qquad \text{(A.IV)}$$

Hence (A.III) and (A.IV) yield

$$\mathcal{E}_d(\omega) = O(1)\cdot\sum_{h=1}^{L} W_h(1/N_h)\sum_{k=1}^{N_h} |Y_{hk}(\omega)|^{2+\delta}. \qquad \text{(A.V)}$$

Now, (A.V) and the Strong Law of Large Numbers (SLLN)for nested arrays imply that $\mathcal{E}_d(\omega)$= O(1) a.s. ω (see for example, Theorem 1.14 Shao, 1999)

Proof of Example 4.1

Consider $\{W_k(s,\omega)\le a\} = \bigcup_{i=1}^{N} S_{i,k}\times F_i$, $S_{i,k} = \{s\in S:\ I_i^k = 1\}$, $F_i = \{\omega\in\Omega:\ Y_i(\omega)\le a\}$ and since $p(s,\omega)$ is constant over the product space for SRS designs, we have $P_{d,m}(W_k(s,\omega)\le a) = \sum_{i=1}^{N} p(S_{i,k})P(F_i)$. Thus under both SRSWOR and SRSWR we have $P_{d,m}(W_k(s,\omega)\le a) = \frac{1}{N}\sum_{i=1}^{N} P(Y_i(\omega)\le a)$, $k = 1,\dots,N$. Under

SRSWOR, $P_{d,m}(W_1(s,\omega)\le a,\ W_2(s,\omega)\le b) = \frac{1}{N(N-1)}\sum_{i=1}^{N}\sum_{j\ne i} P(Y_i(\omega)\le a)P(Y_j(\omega)\le b)$.

Under SRSWR,

$$P_{d,m}(W_1(s,\omega)\le a,\ W_2(s,\omega)\le b) = \frac{1}{N^2}\left[\sum_{i=1}^{N}\sum_{j\ne i} P(Y_i(\omega)\le a)P(Y_j(\omega)\le b) + \sum_{i=1}^{N} P(Y_i(\omega)\le \min(a,b))\right]$$

Hence, for $k\ne \ell$, $P_{d,m}(W_k(s,\omega)\le a,\ W_\ell(s,\omega)\le b)\ne P_{d,m}(W_k(s,\omega)\le a)P_{d,m}(W_\ell(s,\omega)\le b)$, except for the WOR case when the random variables $Y_i(\omega)$ $i=1,...,N$ are identically distributed

Proof of Example 4.5

We use the notation of Example 4.1. For $s_0\in S$ and $k\ne\ell$, under SRS, we have by Proposition 4.1 :

$$P_{m|d}(W_k(s,\omega)\le a,\ s_0) = \sum_{i=1}^{N} P(Y_i\le a)I_i^k(s_0) \text{ and}$$

$$P_{m|d}(W_k(s,\omega)\le a,\ W_\ell(s,\omega)\le b,\ s_0) = \sum_{i=1}^{N}\sum_{j=1}^{N} P(Y_i\le a,\ Y_j\le b)I_i^k(s_0)I_j^\ell(s_0). \qquad \text{(A.VI)}$$

Under SRSWOR, $I_i^k(s_0)I_i^\ell(s_0) = 0$ for every $s_0\in S, k\ne\ell$, and hence these terms disappear in the double sum above. Since the Y^N components are stochastically m-independent, we obtain

$P_{m|d}(W_k(s,\omega)\le a,\ W_\ell(s,\omega)\le b,\ s_0) = P_{m|d}(W_k(s,\omega)\le a,\ s_0)P_{m|d}(W_\ell(s,\omega)\le b,\ s_0)$.

Under SRSWR however, there are samples $s_0\in S$ for which $I_i^k(s_0)I_i^\ell(s_0) = 1$ for some i's, hence the double sum above contains non-zero terms where $i=j$. For samples with repeated $i-$labels, we have:

$P_{m*d}(W_k(s,\omega)\le a,\ W_\ell(s,\omega)\le b,\ s_0) = \sum_{i=1}^{N}\sum_{j\ne i} P(Y_i\le a,\ Y_j\le b)I_i^k(s_0)I_j^\ell(s_0) + \sum_{i=1}^{N} P(Y_i\le\min(a,b))I_i^k(s_0)I_i^\ell(s_0)$.

Which means that we cannot always attain the equality we obtain for the WOR sample and in those cases the projected $W_k(s_0,\omega)$ are model dependent random variables □

Proof of Example 4.6

For simplicity we omit writing the index ν. The W_k are stochastically independent by Example 4.5. Moreover, they are identically distributed random variables with mean zero and variance σ^2, since their distribution is given by (A.VI), given s_0, the $I_i^k(s_0)$ is equal to one for only one i=1,..., N, and the Y_i, $i=1,...,N$ are identically distributed. We apply Theorem 27.2 pp. 359-360 , Billingsley (1995) to this array of i.i.d. r.v., after noting that Lindeberg condition is satisfied for such arrays because the i.i.d. r.v.'s $W_{\nu k}^2$ $k=1,...N_\nu, \nu\ge 1$ are uniformly integrable (see (27.9) of

Billingsley,1995) $\square$

Proof of Proposition 4.2 First we prove (i) : for each $\omega \in \Omega$, $P_{d|m}(\cdot, \omega)$ is a probability measure on the product space. Next we show (ii) : for each measurable set B in the product space, $P_{d|m}(B, \cdot)$ is a version of the conditional probability of B given $S \times \mathcal{F}$, i.e. it is $S \times \mathcal{F}$ -measurable and we have:

$$\int_{S \times F} P_{d|m}(B, \omega)\, d P_{d,m} = P_{d,m}(B \cap S{\times}F), \text{ for any } F \in \mathcal{F}. \qquad \text{(A.VII)}$$

To prove (i), we need to show σ - additivity. From equation (4.6), the σ - additivity follows from the finite additivity of $p_{d,m}$ and the σ - additivity of the indicator functions $\left(I_{\cup_i F_i}(\omega) = \sum_i I_{F_i}(\omega)\right.$ for disjoint F_i). To prove (ii), we note first that $P_{d|m}(B, \cdot)$ is $S \times \mathcal{F}$ - measurable, since $p_{d,m}(s, \cdot)$ is $\mathcal{F}$ - measurable. Next, it suffices to prove (A.VII) on the elementary rectangles $B = \{s_0\} \times F_0$. By definition of $P_{d|m}$ (equation (4.6)), the left hand side of (A.VII) is equal to :

$$\int_{S \times F} p_{d,m}(s_0,\omega)\; I_{F_0}(\omega)\; dP_{d,m} = \sum_{s \in S} \int_{F \cap F_0} p_{d,m}(s_0,\omega)\; p_{d,m}(s,\omega)\; dP,$$

where the equality above holds because over a collection of samples $A \times S$, $dP_{d,m} = \sum_{s \in A} p_{d,m}(s,\omega) dP$. Definition 4.3 and the fact that the design is a probability measure for each ω, imply that the sum above equals:

$$\int_{F \cap F_0} p_{d,m}(s_0,\omega)\; dP = P_{d,m}(\{s_0\} \times F \cap F_0)\; \square$$

Proof of Theorem5.1

(i) $P(\lambda_\nu \leq u) = P_{d,m}(S \times \{\lambda_\nu \leq u\})$ from Definition 4.3 and the fact that $\sum_{s \in S} p(s,\omega) = 1$ for all $\omega \in \Omega$, hence (i) follows.

(ii) $F_\nu(t,\omega)$ converges in probability to $F(t,\omega)$, at points of continuity t, so by Remark 5.1 we can write $P^{\nu}_{d*m}(B_\nu(t),\omega) - F(t,\omega) \to 0$ in probability (P), where $B_\nu(t) = \{(s,\omega) \in S_\nu \times \Omega\text{: } \hat{\theta}_\nu \leq t\}$. By (i) it converges in the law of the product space $(P^{\nu}_{d,m})$. Since $0 \leq P^{\nu}_{d*m}(B_\nu(t),\omega) \leq 1$, the bounded convergence theorem (see Theorem 3 of Chow and Teicher p. 99) implies:

$$\int_{S_\nu \times \Omega} P^{\nu}_{d|m}(B_\nu(t),\, \omega)\; dP^{\nu}_{d,m} - F_\theta(t) \to 0 \text{ as } \nu \to \infty. \qquad \text{(A.VIII)}$$

On the other hand,

$$P^{\nu}_{d,m}\{(s,\omega):\ \theta_\nu(s,\omega)\le t\} = P^{\nu}_{d,m}(B_\nu(t)\cap S_\nu\times\Omega) = \int_{S_\nu\times\Omega} P^{\nu}_{d|m}(B_\nu(t),\omega)\ dP^{\nu}_{d,m}$$

by Proposition 4.2. This last equality and (A.VIII) imply (ii).

(iii)Let $I_\nu(\omega)$ be the indicator function of the measurable set $E_\nu = \{\omega:\ \lambda_\nu(\omega)\le u\}$. Using Proposition 4.2 and the definition of F_ν, we express the distribution function of $(\hat\theta_\nu,\lambda_\nu)$ as:

$$P^{\nu}_{d,m}\{(s,\omega):(\hat\theta_\nu,\lambda_\nu)\le(t,u)\} = P^{\nu}_{d,m}(B_\nu(t)\cap S_\nu\times E_\nu) = \int_{S_\nu\times E_\nu} F_\nu(t,\omega)\ dP^{\nu}_{d,m}.$$

Now, $\int_{S_\nu\times E_\nu} F_\nu(t,\omega)\ dP^{\nu}_{d,m} = \int_{E_\nu} F_\nu(t,\omega) \sum_{s\in S_\nu} p(s,\omega)\ dP_{d,m} = \int_\Omega I_\nu \cdot F_\nu(t,\omega)\ dP,$ and by letting H_λ denote the distribution function of λ, we have for $t,\ u\in\mathbb{R}$:

$$\int_\Omega I_\nu\cdot F_\nu(t,\omega)\ dP - F_\theta(t)\ H_\lambda(u) = \int_\Omega I_\nu\,(F_\nu(t,\omega)-F_\theta)\ dP + F_\theta\int_\Omega (I_\nu - H_\lambda)\ dP.$$

All functions above are bounded by one, so the first term of the right-hand side converges to zero by the bounded convergence theorem since by hypothesis $F_\nu(t,\cdot) - F_\theta(t)$ converges to zero in probability P at points t of continuity of F_θ (see also Remark 5.1). The second term of the right-hand side also converges to zero, since by hypothesis, at points u of continuity of H_λ,

$$\int_\Omega (I_\nu - H_\lambda)\ dP = P(\lambda_\nu(\omega)\le u) - H_\lambda(u)\to 0$$

REFERENCES

Billingsley, P.(1995). *Probability and Measure*, third edition, Wiley, New York.

Binder, D. A.(1983). On the variances of asymptotically normal estimators from complex surveys, *International Statistical Review*, **51**, 279-292.

Binder, D. and Roberts, G.(1999). Designed-based and model-based methods for estimating model parameters. *Presented at the International Conference for Analysis of Survey Data, 24-26 August, 1999,* Southampton, U.K.

Chow, Y.S., Teicher, H. (1997). *Probability Theory,* third edition. Springer-Verlag, New York

Fuller, W.A. (1975). Regression analysis for sample surveys. Sankhyã **37**, 117-132.

Godambe,V.P. and Thompson, M.E. (1986). Parameters of super-populations and survey population: their relationship and estimation, *International Statistical Review*, **94**, 127-137.

Hájek, J. (1960). Limiting distributions in simple random sampling from finite populations, *Publ.*

Math. Inst. Hungarian Acad. Sci., **5**, 361 - 374.
Hájek, J. (1964). Asymptotic theory of rejective sampling with varying probabilities from a finite population. *Ann. Math. Statist*. **35**, 1491-523.
Hájek, J. (1981) (assembled after his death by Vaclav Dupac) *Sampling from a finite population,* M. Dekker, New York.
Hartley, H.O. and Sielken, R.L. (1975). A "super-population viewpoint" for finite population sampling, *Biometrics*,31,411-422.
Korn, E. L. and Graubard, B.I. (1998). Variance estimation for super-population parameters, *Statistica Seneca,* **8**, 1131-1151.
Krewski, D. and Rao, J.N.K.(1981). Inference from stratified samples: Properties of linearization, jackknife and balanced repeated replication methods, *Ann. Stat.* **9**, 1010-1019.
Molina, E.A., Smith, T.M.F. and Sudgen, R.A. (2001). Modelling Overdispersion for Complex Survey Data. *International Statistical Review*, **69**, 3, 373-384.
Pfeffermann, D. and Sverchkov, M. (1999) Parametric and semiparametric estimation of regression models fitted to survey data. *Sankhya*, Series B, **61**, 166-186.
Rodríguez, J. E. (2001) A probabilistic framework for inference in finite population sampling, *2001 Proceedings of the Section on Survey Research Methods of the American Statistical Association*, to appear.
Rosén, B. (1972) Asymptotic theory for successive sampling with varying probabilities without replacement. I and II. *Ann. Math. Statist*. **43**, 373-97, 748-76.
Rosén, B. (1997). On sampling with probability proportional to size. *Journal of Statistical Planning and Inference,* **62**, 159-191.
Rubin Bleuer, S. (1998). Inference for parameters of the super-population, Part I, *Research Sabbatical Report, Internal report*, Statistics Canada.
Rubin Bleuer, S. (2000). Some issues in the analysis of complex survey data. *Statistics Canada Series, Methodology Branch, Business Survey Methods Division,* BSMD- 20-001 E.
Rubin Bleuer, S. (2001). A test for survival distributions using data from a complex sample. *Proceedings of the Survey Methods Section, SSC Annual meeting, 2001* (p.103-110).
Rubin Bleuer, S. and Schiopu Kratina, I. (2000). Some issues in the analysis of complex survey data. *2000 Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 734-739.
Rubin Bleuer, S. and Schiopu Kratina, I. (2001). A probabilistic set-up for model and design-based inference. *Statistics Canada Series, Methodology Branch,* METH 2002 - 002E.
Rubin Bleuer, S. and Schiopu Kratina, I. (2002). On the two-phase framework for joint design and model-based inference.
Särndal, C-E, Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling,* Springer-

Verlag, New York.
Shao, J.(1999). *Mathematical Statistics*, Springer-Verlag, New York.
Yuan,K. And Jennrich, R. (1998). Asymptotics of estimating equations under natural conditions. *Jounal of Multivariate Analysis* **65**, 245-260.
Yung, W. and Rao, J.N.K.(2000) Jackknife variance estimation under imputation for estimators using poststratification information, *Journal of the American Statistical Association, Theory and Methods*, **95**, **No. 451**, 903-915.