A PROBABILISTIC SET-UP FOR MODEL AND DESIGN-BASED INFERENCE

Susana Rubin-Bleuer and Ioana Şchiopu-Kratina, Statistics Canada
Susana Rubin Bleuer, 120 Parkdale Ave., Ottawa, Ontario, Canada, K1A-0T6

Key words: design-based inference, analytic inference.

1. INTRODUCTION

Classical sampling theory concerns inference for finite population parameters. This enables us to work exclusively within a sample probability space, which we design and control. However, there are many situations when we have to resort to postulating a model, for example, when we would like to test or draw conclusions for a more general population than the finite population from which we obtained the sample. Once we incorporate a general population model (or super-population) in our framework, our inference procedures need to account for the design (unequal probabilities, dependent selection indicators, etc.), other survey processes (non-response, poststratification, etc.) and the super-population model. Rubin-Bleuer (1998, 2000) and Rubin-Bleuer & Schiopu-Kratina (2000) have formally defined a probability space that includes both the designed sample space and the super-population model, which was called the “product space”. They described 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.

In this paper, we further explore the structure of the product probability space by exploiting additional information by means of conditioning. We show that the design-based inference and model-based inference fit naturally into the structure of the product space as conditional inference with an appropriately defined probability. We also show that when the design does not depend on the super-population associated with the finite population, the inference can be carried out in the model space and the design can be ignored. This is the approach taken by Fuller (1975). The advantages of considering a super-population that generates then finite population are also explored.

We repeat a few definitions from our previous work to make this paper self-contained. In Section 2 we define the finite population, the design probability and the sample estimator. In these definitions, we assume that the populations of interest are composed by the characteristics of interest, the prior information with which we design the survey, and we acknowledge the possible existence of auxiliary information. We also give an example of a sample estimator that we use in further developments. In Section 3 we define the super-population and we 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. In section 4 we define the product probability space and show how some dependence-independence properties in the model and the design space translate to the product space. Finally, in section 5, we look at some useful conditional probabilities, which reflect the change in the product space probability measure $P_{d,m}$ due to additional information. We also give examples of the use of the projection of $P_{d,m}$ onto the model space.

2. FINITE POPULATIONS AND 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 $y_i \in \mathbb{R}^k$ represent the characteristics of interest, $x_i \in \mathbb{R}^l$ represent the auxiliary information, and $z_i \in \mathbb{R}^m$ contains prior information available at the time of the design of the survey on all units {i}, i = 1, ... N. Here k, l and m are positive integers. We write $y^N = (y_i)_{i=1,..N}$, $x^N = (x_i)_{i=1,..N}$ and $z^N = (z_i)_{i=1,..N}$ ■

As in Särndal et al (1992, p. 25), we assume that a probabilistic (randomized) selection, or sampling scheme is given. A sample is the realization of such a randomized selection. Let N (or M if the scheme is multistage) be the finite population size (i.e. the number of ultimate sampling units in the population). A comprehensive definition of a sample is that of Hajek (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). In order for the set of all possible samples S to be completely determined it is necessary to know a priori all the stratum and cluster sizes and their respective sample sizes or expected sample sizes. If so, S is well defined, since every label in the finite population must have a positive probability of selection. Under a (WOR) scheme, a sample can also 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.

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 Särndal et al. p.27, or Hájek 1981, p. 21). Our 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 2.2 Let N be the number of ultimate sampling units in the population. Given a sampling scheme, let S be the set of all possible samples under the scheme. Let $\mathcal{C}(S)$ consist of all subsets of S. Let $D(z) \subseteq \mathbb{R}^m$ be a subset of values of the prior information. A **sampling design associated to a sampling scheme** is a function $p : \mathcal{C}(S) \times \mathbb{R}^{m\times N} \to [0, 1]$ such that:
(i) p(s, -) is Borel - measurable in $\mathbb{R}^{m\times N}$, $\forall\, s \in S$
(ii) $p(- , z_1, z_2, \dots z_N)$ is a probability measure on $\mathcal{C}(S)$ $\forall\, (z_1, \dots z_N) \in D(z) \subseteq \mathbb{R}_+^{m\times N}$
We say that $(S, \mathcal{C}(S), p)$ is a design space ■

Without loss of generality, in all applications we will take m = 1. Under a two stage design with N primary sampling units (psu's) we can carry on the design with prior information on the N psu's only. But the definition of design can be extended to include prior information on all sampling units.
In what follows, the subscript "d" refers to design randomization.

Example 2.1 Stratified two-stage with probability proportional to size (PPSWR). This type of design is often used in household surveys and could be extended to include several stages of sampling. The population is stratified into L strata, each one containing N_h psu's. Let N be the number of psu's in the population, $N = \sum_{h=1}^{L} N_h$. Each psu {hi} consists of M_{hi} ultimate units, $M_h = \sum_{i=1}^{N_h} M_{hi}$ is the number of ultimate sampling units in stratum h and $M = \sum_{h=1}^{L} M_h$. The "size" of the psu {hi} is $z_{hi} = M_{hi}$, i = 1,...N_h, h = 1,.. L. We set $z^N = M^N$. Suppose $n_h \geq 2$ psu's are selected with replacement from the N_h psu′s in the h*th* stratum with probabilities $p_{hi} = M_{hi} / M_h$, i = 1,...N_h, h = 1,.. L, at each draw. The selection is done independently in each stratum, and independent sub-samples are taken within those psu's selected more than once ■

Definition 2.3 A finite population parameter θ_N is a Borel- measurable function defined on a subset D(y,x,z) of $\mathbb{R}^{(k+l+m)\times N}$ with values in $\mathbb{R}$. An estimator of this finite population parameter associated with a design or **sample estimator** is a function $\theta_d : S \times D(y,x,z) \to \mathbb{R}$, where the domain $D(y,x,z) \subseteq \mathbb{R}^{(k+l+m)\times N}$, $\theta_d\,(s,\cdot)$ is Borel - measurable and $\theta_d(\cdot, y^N, x^N, z^N)$ is C(S) - measurable■

Example 2.2 Stratified two-stage PPSWR design. We define here a sample estimator that we will use later on in the paper. This estimator was also used by Krewski & Rao (1981) for their work on inference from stratified samples. In the context of Example 2.1, let the prior information be given by the psu sizes. The finite population mean is $\theta_N = y/M = \sum_{h=1}^{L} W_h \bar{y}_h$, where $W_h = M_h/M$ is the stratum weight, $\bar{y}_h = \sum_{i=1}^{N_h} y_{hi}/M_h$ is the finite population stratum mean, and y_{hi} is the total of psu {hi} i = 1,.. N_h, h = 1,...L. Let $\hat{y}_{hi}$ be an unbiased estimator of the total y_{hi} for a selected psu based on sampling at the second stage. Then a sample estimator of the stratum mean is $\hat{\theta}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{\theta}_{hi}$, where $\hat{\theta}_{hi} = \sum_{k=1}^{N_h} \frac{\hat{y}_{hk}}{M_{hk}} I_{hk}^{i}$ and $I_{hk}^{i} = 1$ if psu {hk} is selected in the h-*th* stratum sample at the i-*th* draw and 0 otherwise, k = 1,... n_h. Finally, a sample estimator of the mean θ_N is given by $\hat{\theta}_N(s, y^N, M^N) = \sum_{h=1}^{L} W_h \hat{\theta}_h$.
The estimator $\hat{\theta}_N$ is design-unbiased and design-consistent (Krewski & Rao, 1981) ■

3. SUPER-POPULATIONS

The following definition is similar to the definition of super-population given in Hàjek (1981, p. 14).
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 $(\Omega, \mathscr{F}, P)$ and random vectors (Y_i, X_i, Z_i), $Y_i\colon \Omega \to \mathbb{R}^k$, $X_i\colon \Omega \to \mathbb{R}^l$, $Z_i\colon \Omega \to \mathbb{R}^m$, 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 denote by $Y^N = (Y_i)$ i=1,...N. Similarly, we define X^N and Z^N. 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 are given a priori is called a **super-population model**. We note that the outcome ω_0 which generates the finite population need not be unique ■

We give below an application related to the work of Krewski & Rao (1981). Consider a sequence of finite populations with stratified, PPSWR designs (Example 2.2) and associated super-populations, with $N_{\nu h}$ clusters in stratum h and number of strata L_ν , h = 1,...L_ν, $\nu \geq 1$. The cluster totals y_{hi} are the realizations of super-population-random vectors Y_{hi}, i=1...N_h, h=1...L, which are stochastically m-independent within and across strata, and are identically distributed within strata. Krewski & Rao (1981) give conditions for inference on the finite population means $\theta_{N_\nu} = y_\nu / M_\nu$ using the design based estimators $\hat{\theta}_{N_\nu} = \sum_{h=1}^{L_\nu} W_h \hat{\theta}_h$ of Example 2.2. We show that moment conditions in the super-population yield the Liapunov- type conditions of Krewski & Rao for asymptotic normality of the sample mean (in the law of the design). The number of strata $L_\nu \to \infty$, as $\nu \to \infty$ and the number of clusters $N_{\nu h}$ in each stratum remains bounded. In the following, we omit indexing the populations.

Proposition 3.1 We assume

(C_3) of Krewski-Rao (1981) p.1014: $W_h = \frac{M_h}{M} = O(L^{-1})$,

which implies that no strata is of disproportionate size. If, in addition, we assume the following model - based conditions:

(M_1) $L^{-1} \sum_{h=1}^{L} E_m |Y_{h1}|^{2+\delta} = O(1)$ as $\nu \to \infty$, and $\sum_{h=1}^{\infty} \frac{V_m |Y_{h1}|^{2+\delta}}{h^2} < \infty$, then condition (C_1) of Krewski-Rao (1981) p.1014, holds, namely:

$(C_1) \sum_{h=1}^{L} W_h E_d |\hat{\theta}_{hi} - \bar{Y}_h|^{2+\delta} = O(1)$ a.s. ω, where $\hat{\theta}_{hi}$ is defined in Example 2.2.

Proof

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

$$|(1/N) \sum_{k=1}^{N} X_k|^p \leq (1/N) \sum_{k=1}^{N} |X_k|^p, \qquad (3.1)$$

(see Chung, 1974, p.48), we have $|\frac{\hat{\theta}_{hi} - \bar{Y}_h}{2}|^{2+\delta} \leq (1/2)(|\hat{\theta}_{hi}|^{2+\delta} + |\bar{Y}_h|^{2+\delta})$, and hence

$$\mathcal{E}_d(\omega) \leq 2^{1+\delta} \sum_{h=1}^{L} W_h E_d (|\hat{\theta}_{hi}|^{2+\delta} + |\bar{Y}_h|^{2+\delta}). \qquad (3.2)$$

Now, by definition of $\hat{\theta}_{hi}$,

$$E_d |\hat{\theta}_{hi}|^{2+\delta} = \sum_{k=1}^{N_h} |\frac{y_{hk}}{M_{hk}}|^{2+\delta} p_{hk} \leq (1/ N_h) \sum_{k=1}^{N_h} |y_{hk}|^{2+\delta} \qquad (3.3)$$

since $M_{hk} \geq 1$ and $M_h \geq N_h$, k=1,..., N_h, h=1,...,L. Similarly, by (3.1) with $N=N_h$,

$$|\bar{Y}_h|^{2+\delta} = |(1/ M_h) \sum_{k=1}^{N_h} y_{hk}|^{2+\delta} \leq (1/ N_h) \sum_{k=1}^{N_h} |y_{hk}|^{2+\delta} \qquad (3.4)$$

Hence (3.3), (3.4) and (C_3) yield

$$\mathcal{E}_d(\omega) = O(1/L) \sum_{h=1}^{L} ((1/ N_h) \sum_{k=1}^{N_h} |y_{hk}|^{2+\delta}) \qquad (3.5)$$

Now, (3.5) and (M_1) imply that $\mathcal{E}_d(\omega)$= O(1) a.s. ω (see for example, Theorem 1.14, Shao, 1999) ■

4. THE PRODUCT SPACE

Definition 4.1

Consider a finite population U of size N (number of ultimate sampling units), generated by a super-population (Y^N, X^N, Z^N): $(\Omega, \mathscr{F}, P) \to \mathbb{R}^{(k+l+m)\times N}$ as in Definition 3.1. We assume that the size N 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 given a design p, it is necessary to have fixed the number N (or M) of ultimate units of the population, as well as the number and size of strata and p.s.u.'s, secondary sampling units, etc., for the space S of all possible samples to be well defined (before we even know the outcome $\omega \in \Omega$ that will generate the finite population). We define the product space as a measurable space given by $\Omega_{d,m} = S \times \Omega$ with the σ-field $C(S) \times \mathscr{F}$ ■

Remark 4.1 Consider the elementary rectangles of the form $\{s\} \times F$, $s \in S$, $F \in \mathscr{F}$. Since C(S) is a finite collection of sets, the σ-field $C(S) \times \mathscr{F}$ consists of finite unions of elementary rectangles. More precisely, a measurable set $B \neq \emptyset$ in this σ-field can be uniquely written as the disjoint union

$$B = \bigcup_{s \in A} \{s\} \times F_s, \ A \in C(S), \ F_s \in \mathscr{F}, \qquad (4.1)$$

where all s are distinct and the $F_s \neq \emptyset$ ■

The next result shows that a design and sample estimator can be viewed as measurable functions in the product space.

Proposition 4.1 Consider a super-population associated with a finite population U, as in Definition 3.1. Let p be a design on the finite population as in Definition 2.2 with domain D(z). Let us assume that the range of the Z^N of Definition 3.1 is contained in the domain, i.e., $R(Z^N) \subseteq D(z)$. Let $\mathscr{F}(Z^N) = \sigma(Z^N)$ be the sub σ - field of $\mathscr{F}$ generated by Z^N. Then the design can be viewed as a $C(S) \times \mathscr{F}(Z^N)$ - measurable function $p_{d,m}$ defined by:

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

Similarly, consider the sample estimator

$$\theta_d : S \times D(y, x, z) \to \mathbb{R},$$

and let us assume that the range $R(Y^N, X^N, Z^N) \subseteq D(y, x, z)$. Let $\mathscr{F}_N = \mathscr{F}(Y^N, X^N, Z^N) = \sigma(Y^N, X^N, Z^N)$ be the sub σ - field of $\mathscr{F}$ generated by Y^N, X^N and Z^N. Then the sample estimator can be viewed as an $C(S) \times \mathscr{F}_N$ - measurable function given by:

$\theta_{d,m}(s, \omega) = \theta_d(s, Y^N(\omega), X^N(\omega), Z^N(\omega))$, $\omega \in \Omega$, $s \in S$ (4.3)

.**Proof**: Since both $p(s,\cdot)$ and $\theta_d(s,\cdot)$ are Borel functions on $\mathbb{R}^{m\times N}$ and $\mathbb{R}^{(k+l+m)\times N}$ respectively, their composition with random vectors (Y^N, X^N, Z^N) renders them measurable with respect to the minimum sigma fields $\sigma(Z^N)$ and $\sigma(Y^N, X^N, Z^N)$, respectively (see for example Chow and Teicher (1997) Theorem 4, p.17) ■

We next define a probability measure on the product space $(S \times \Omega, C(S) \times \mathscr{F})$.

Definition 4. 2 We define the measure $P_{d,m}$ first on elementary rectangles of the product σ - field, and then on general measurable sets expressed in their reduced form :

$P_{d,m}(\{s\} \times F) \triangleq \int_F p_{d,m}(s, \omega)\, d P$, $s \in S$, $F \in \mathscr{F}$

$P_{d,,m}(\bigcup_{s\in A} \{s\} \times F_s) \triangleq \sum_{s \in A} \int_{F_s} p_{d,m}(s, \omega)\, d P$, $A \in C(S)$, $F_s \in \mathscr{F}$, $F_s \neq \varnothing$.

$P_{d,m}$ is additive by definition, and hence σ-additive because all sets in $C(S) \times \mathscr{F}$ can be expressed as a finite union of elementary rectangles ■

Proposition 4.2 If $B = S \times F$, $F \in \mathscr{F}$, then $P_{d,m}(B) = P(F)$. In particular, $P_{d,m}(S \times \Omega)=1$.

Remark 4.2 The σ-additivity of $P_{d,m}$ and Proposition 4.2 imply that $P_{d,m}$ is a probability measure on the product space.

Example 4.1 Stochastic dependence in the product space. Let Y^N and p denote, respectively, the super-population and design of Definition 4.1with Y^N composed of N independent not necessarily identically distributed random variables and p a design associated with a "Simple Random Sample Without Replacement" (SRSWOR) or a "Simple Random Sample with Replacement" (SRSWR) scheme of sample size n.

Let us denote by $y_s = \{y_i, i \in s\}$the values of y^N associated with the units i in a sample $s \in S$. We define the k-th draw-selection indicators $\{I^k_1,...,I^k_N\}$ by $I^k_i=1$ 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

$$y_s = \left(\sum_{i=1}^{N} y_i I^1_i(s), \sum_{i=1}^{N} y_i I^2_i(s),..., \sum_{i=1}^{N} y_i I^n_i(s)\right).$$

Each coordinate of the sequence (representing a draw) is the y-value of a unit of the finite population. If the design is WR then I^k_i and I^l_j are d-stochastically independent for $k\neq l$ and all i,j, whereas if the design is WOR the I^k_i and I^l_j are d-stochastically dependent and if $I^k_i=1$ then $I^l_i=0$ for all $l\neq k$. The y_s can be viewed as a group of random variables Z_k in $\Omega_{d,m}$: $Z_k(s,\omega) \triangleq \sum_{i=1}^{N} Y_i(\omega) I^k_i(s)$. Then whether the design p is SRSWOR or SRSWR, the Z_k are stochastically dependent random variables in the product space, even though the original super-population random variables Y^N were stochastically independent in $(\Omega, \mathscr{F}, P)$: we have, for $k\neq l$,

$$P_{d,m}(Z_k(s,\omega)\leq x , Z_l(s,\omega)\leq z) \neq P_{d,m}(Z_k(s,\omega)\leq x)P_{d,m}(Z_l(s,\omega)\leq z).$$

Indeed, under WOR, $P_{d,m}(Z_k(s,\omega)\leq x)=\frac{n}{N}\sum_{i=1}^{N} P(Y_i\leq x)$,

$$P_{d,m}(Z_k(s,\omega),Z_l(s,\omega)\leq z)=\frac{n(n-1)}{N(N-1)}\sum_{i=1}^{N}\sum_{j\neq i} P(Y_i\leq x)P(Y_j\leq z).$$

Whereas under WR $P_{d,m}(Z_k(s,\omega)\leq x)=\frac{1}{N}\sum_{i=1}^{N} P(Y_i\leq x)$,

$$P_{d,m}(Z_k(s,\omega),Z_l(s,\omega)\leq z)=$$

$$\frac{1}{N^2}\{\sum_{i=1}^{N}\sum_{j\neq i} P(Y_i\leq x)P(Y_j\leq z)+\sum_{i=1}^{N} P(Y_i\leq \min(x,z))\}$$ ■

5. CONDITIONAL PROBABILITIES

Let $B \in C(S)\times\mathscr{F}$, and $\omega \in \Omega$. We denote by $B(\omega)$ the ω - section of B, i.e., $B(\omega) = \{s \in S \mid (s,\omega) \in B\}$.

Definition 5.1 Let $B=\bigcup_{s \in A} \{s\}\times F_s \in C(S) \times \mathscr{F}_N$ where $\mathscr{F}_N = \sigma(Y^N, X^N, Z^N)$ as in Proposition 4.1 , $s^* \in S$ and $\omega \in \Omega$. We define the set function :

$P_{d|m}(B, (s^*, \omega)) = p_{d,m}(B(\omega), \omega) = \sum_{s\in B(\omega)} p_{d,m}(s, \omega)$

Note that $P_{d|m}$ is constant for $s^* \in S$ ■

The set function $P_{d\,|\,m}$ defined above is the (regular) conditional probability measure on the product space, given the marginal σ - field $S \times \mathscr{F}_N$, as shown by Proposition 5.1. The σ - field $S \times \mathscr{F}_N$ summarizes our knowledge of the model. The conditional density $p_{d\,|\,m}$ (s,ω) turns out to be $p_{d\,,\,m}$, since the sampling design is a probability on the design space (S, *C* (S), p), that is, we "recover" the design probability we started out with. For pertinent definitions and examples the reader is sent to Chow and Teicher (1997) pp. 223-224. One advantage of regular conditional probabilities is that conditional expectations may be then envisaged as ordinary expectations relative to the conditional probability measure (Chow and Teicher (1997), Theorem 1, Section 7.2).

Proposition 5.1. Conditioning on the model. The set function
$P_{d\,|\,m}(.\,,\,.) : (C(S)\times\mathscr{F}_N) \times (S\times\Omega) \rightarrow [0,1]$
of Definition 5.1, is the (regular)conditional probability measure on $C(S) \times \mathscr{F}_N$ given the σ - field $S \times \mathscr{F}_N$, and $\int_{S \times F} P_{d|m}(B,(s_0\,,\omega_0))\, d\,P_{d,m} = P_{d\,,\,m}(B \cap S\times F)$, for any $F \in \mathscr{F}_N$ and $s_0 \in S$, $\omega_0 \in \Omega$.

Remark 5.1 If in Definition 5.1 we replace $\mathscr{F}_N$ by $\mathscr{F}$, such that $\mathscr{F}_N \subset \mathscr{F}$ then we can replace $\mathscr{F}_N$ by $\mathscr{F}$ everywhere in Proposition 5.1■

Remark 5.2 Proposition 5.1 implies that if we know the model, then conditioning on the model (i.e. on the σ - field $S \times \mathscr{F}_N$) yields a probability measure that is equal to the design probability measure applied to the projections of sub-sets of $S\times\Omega$ onto S.

Before we establish a sampling scheme, we usually know what we call "prior" information, i.e., values $z^N=(z_1,...z_N)$ that are used for the design. Say, for example, the z^N are the cluster sizes of a Canadian population existing right now. Let $F_z= \{\omega \in \Omega: \ Z_i(\omega)= z_i\ , \ i = 1,...N\}$ We may select the sample with probability proportional to those sizes, but we might want to learn about a more general population than the finite population living in those clusters now. In this case, we would use the prior information by conditioning on the σ - field $\mathscr{G}$ generated by the event $S\times F_z$ and use the conditional probability measure $P_{d\,,\,m}(\cdot \mid \mathscr{G})$ to do inference.

Proposition 5.2. Conditioning on the prior information. $P_{d\,,\,m}(\cdot \mid \mathscr{G})$ is a regular conditional probability measure. Moreover, it is constant on $S\times F_z$: for $B= \bigcup_{s \in A} \{s\}\times F_s \in C(S)\times\mathscr{F}_N$, and $(s_0\,,\omega_0) \in S\times F_z$ then

$$P_{d\,,\,m}(B|\,\mathscr{G}\,)(s_0\,,\omega_0) = \sum_{s\in A} p_{d,m}(s,z_1,...,z_N)\,P(F_s|F_z)\,,$$

if $P(F_z)>0$ and equal to $P_{d\,,\,m}(B)$ if $P(F_z)=0$.
Proof: $P_{d\,,\,m}(\cdot \mid \mathscr{G})$ is a regular probability measure, and has the above expression, because the σ - field $\mathscr{G}$ is generated by a partition of $S \times \Omega$ (see Example 1, section 7.2, Chow and Teicher (1997)).

Remark 5.3. Other authors also work with conditional probability measures to do analytic inference with survey data. Sverchkov & Pfeffermann (2000) have a different conception of the selection probabilities (depending on Y^N and X^N as well as Z^N). However, their "parametric distribution of the sample data", which they use to do inference, can be thought of as the conditional probability measure given a sample and the auxiliary data, i.e., $P_{d\,,\,m}(\cdot \mid \mathscr{G})$ where the σ - field $\mathscr{G}$ is generated by the event $\{s\}\times F_x$ where s is a fixed sample and
$F_x = \{\omega \in \Omega: \ X_i(\omega)= x_i\ , \ i = 1,...N\}$.

We calculate next the conditional probability given the "orthogonal" marginal field of the product space, namely $C(S) \times \Omega$. Conditioning on the field $C(S) \times \Omega$ represents exploiting the control we have over the design. We denote by $P_{m|d}$ the conditional probability $P_{m,d}$ given the field $C(S) \times \Omega$ ■

Proposition 5.3. Conditioning on the design. $P_{m|d}$ is a (regular) conditional probability measure and for sets $B \in C(S)\times \Omega$, expressed in its reduced form, $B= \bigcup_{s \in A} \{s\}\times F_s$ $A \in C(S)$, $F_s \in \mathscr{F}$,

$$P_{m|d}(B, (s_0\ \omega_0)) = \int_{F_{s_0}} p(s_0, \omega)\, dP \big/ \int_{\Omega} p(s_0, \omega)\, dP, \quad (5.1)$$

if $s_0 \in A$ and 0 otherwise. If, in particular, $p(s, \omega)$ does not depend on ω, then $P_{m|d}(B,(s_0\ \omega_0)) = P(F_{s_0})I_A(s_0)$. Here I_A is the indicator function of the set A.
Proof: as for Proposition 5.2, $P_{m|d}$ is a conditional probability measure and Equation (5.1) holds, because the conditioning field is the field generated by the partition of $S\times \Omega$ given by $\{s \text{ x } \Omega,\ s \in S\}$■

Thus, if the design is independent of the realization of the super-population, inference in this conditional probability coincides with inference in the super-population and the design randomization can be ignored. Such is the case of SRSWOR and SRSWR.

Example 5.1 Projection of the sample onto the model space by means of $\mathbf{P_{m|d}}$. In the context of example 4.1, under SRSWOR, when we project the Z_k- variables back into the model space by means of the conditional probability $P_{m|d}$, they recover the original independence of the Y^N : a WOR design implies that there are no repetitions of the Y_i's in the sample, so the sample is a subset of the original Y^N. But under SRSWR the Z_k's projected onto the model space lose their original independence. Indeed, for $s_0 \in S$ and $k \neq l$, whether under SRSWOR or SRSWR, we have:

$$P_{m|d}(Z_k(s,\omega) \le x; s_0) = \sum_{i=1}^{N} P(Y_i \le x) I^k{}_i(s_0) \text{ and}$$

$$P_{m|d}(Z_k(s,\omega) \le x, Z_l(s,\omega) \le z; s_0) =$$

$$\sum_{i=1}^{N} \sum_{j=1}^{N} P(Y_i \le x, Y_j \le z) I^k{}_i(s_0) I^l{}_j(s_0) \text{ by proposition 5.3.}$$

Now, under SRSWOR, $I^k{}_i(s_0)\ I^l{}_i(s_0) \equiv 0$ for every $s_0 \in S$, and hence the terms in the double sum above with i=j disappear. Since the Y^N are independent, for $i \neq j$ $P(Y_i \le x, Y_j \le z) = P(Y_i \le x)P(Y_j \le z)$, which yields independence of the Z_k's:

$$P_{m|d}(Z_k(s,\omega) \le x, Z_l(s,\omega) \le z; s_0) =$$

$$P_{m|d}(Z_k(s,\omega) \le x; s_0) P_{m|d}(Z_l(s,\omega) \le z; s_0). \qquad (5.2)$$

Under SRSWR however, there are samples $s_0 \in S$ for which $I^k{}_i(s_0)\ I^l{}_i(s_0) = 1$ for some i's, and for those samples, the double sum above contains terms where i=j , for which we have $P(Y_i \le x, Y_i \le z) = P(Y_i \le \min(x,z))$. Thus, for those samples,

$$P_{m|d}(Z_k(s,\omega) \le x, Z_l(s,\omega) \le z; s_0) =$$

$$\sum_{i=1}^{N} \sum_{j \neq i}^{N} P(Y_i \le x, Y_j \le z) I^k{}_i(s_0) I^l{}_j(s_0) +$$

$$\sum_{i=1}^{N} P(Y_i \le \min(x,z)) I^k{}_i(s_0) I^l{}_i(s_0).$$

This means that under SRSWR, for the samples s_0 where we selected repeated i- labels, we cannot attain the equality (5.2) and thus the Z_k (s_0,ω) are m-dependent random variables ■

Example 5.2 Asymptotic normality of the sum of sample units when projected by $\mathbf{P_{m|d}}$. Consider a sequence of super-populations associated with finite populations as in Definition 3.1. We assume that $Y_{\nu i}$, i = 1, ...N_ν, $\nu \ge 1$ are i.i.d. r.v.'s on (Ω, $\mathscr{F}$, P) with 0 mean and finite second moment $\sigma^2 \neq 0$. The design is SRSWOR of size n_ν , ν = 1,.... Then $(\sigma^2 n_\nu)^{-1/2}[\sum_{i \in s_\nu} Y_{\nu i}]$ convergence in the m - distribution to a standard normal r.v. N(0,1) .

Indeed, by Proposition 5.3, the survey design can be ignored and we can view the selection of the sample s_ν as a nonrandom selection of n_ν labels from N_ν, which creates the array:

$$\{Y_{\nu i}\},\ i \in s_\nu, \nu \ge 1.$$

In order to apply Theorem 27.2 pp. 359-360 , Billingsley (1995) to this array of i.i.d. r.v., we note that Lindeberg condition is satisfied for such arrays because the i.i.d. r.v.'s $Y_{\nu i}^2, i = 1,...N_\nu, \nu \ge 1$ are uniformly integrable (see (27.9) of Billingsley (1995) ■

REFERENCES

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

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

Chung, K.L.(1974). *A Course in Probability Theory*, second edition. Academic Press, New York.

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

Hájek, J. (1960). Limiting distributions in simple random sampling from finite populations, *Publ. Math. Inst. Hungarian Acad. Sci*., **5**, 361 - 374.

Hájek, J. (1981). *Sampling from a Finite population.* Series in Textbooks and monographs, vol.37, Marcel Dekker, Inc..New York & Basel.

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.

Sverchkov, M. and Pfeffermann, D. (2000). Prediction of finite population totals under informative sampling utilizing the sample distribution. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 41-46.

Rubin Bleuer, S. (1998). Inference for parameters of the superpopulation, Part 1, *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. and Schiopu Kratina, I. (2000). Some issues in the analysis of complex survey data. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 734-739.

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.