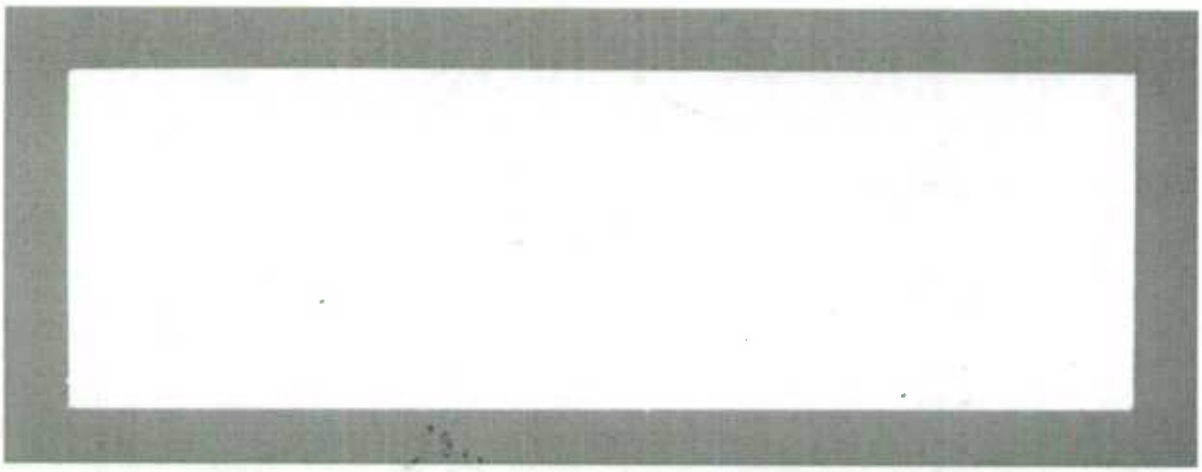

Methodology Branch

Business Survey Methods Division

Direction de la méthodologie

Division des méthodes d'enquêtes auprès des entreprises

Canada

ON THE CONVERGENCE OF SAMPLE EMPIRICAL PROCESSES

By

Susana Rubin-Bleuer

BSMD - 2003 - 003E

ON THE CONVERGENCE OF SAMPLE EMPIRICAL PROCESSES

Susana Rubin Bleuer[1]

[1] Susana Rubin-Bleuer, Statistics Canada, Ottawa, Canada, K1H-0T6, rubisus@statcan.ca

ABSTRACT. We prove a Glivenko-Cantelli property in a "joint" design-model probability space that changes as the sample size increases. We also prove weak convergence of the sample empirical process under selected complex sample designs.

RÉSUMÉ. On établisse le théorème de Givenko-Cantelli pour un espace qui inclut à la fois l'espace plan d'échantillonnage et l'espace modèle. On établisse aussi la convergence faible du procès empirique d'échantillon sous plans d'échantillonnage complexes.

1. INTRODUCTION

Statistical tools like the empirical distribution function and the corresponding empirical process yield many important tests and related results widely used in the practice. The Glivenko-Cantelli theorem, which concerns a statistic of the empirical distribution function is sometimes referred to as the fundamental theorem in statistics. However, until recently, few of these results have been extended to their sample counterparts in an analytic form. In this paper we prove a Glivenko-Cantelli type theorem for the sample space and we also show weak convergence of the sample empirical process, which in turn implies convergence in law of many important processes used, for example, in survival analysis (see, for example, Rubin-Bleuer (2001) and Lin (2000).

The Glivenko-Cantelli theorem states that given a sequence of independent, identically distributed random variables (i.i.d.r.v.) from a distribution function F, all defined in the same probability space, their empirical distribution function converges uniformly (in sup norm) to the distribution function F, with probability one, as $N \to \infty$.

Now we consider a sample estimator of the empirical distribution function. Given a sample design on a finite population of size N, we can construct a sequence of sample probability spaces which change as N changes. Thus, any asymptotic property we might want to consider has to be proved either in probability or in law, rather than with probability one, since there is no concept of almost surely when the spaces change as $N \to \infty$.

In this paper we show convergence in sup norm of the sample empirical distribution function and weak convergence of the sample empirical process for some complex designs, under sufficient design conditions.

The treatment that we give here to the problem consists in acknowledging that there is a super-population which spans the finite population, and using the distribution F of the super-population model in the proof, not unlike the usual proof of the theorem. We define a "joint" design-model space and look at uniform convergence in this space. We take into account all the processes that generate it, and we do so here using the super-population theory approach introduced by Hartley and Sielken (1975). The approach regards the finite population of interest as the outcome of a sample of N independent random variables from an infinite population (the set of these N random variables are often called the super-population). And it regards the stochastic procedure generating the observed sample of size n from the finite population as the second phase sample of a two-phase sampling process. Thus, in terms of the super-population model, the design-probability may be viewed as being based on the conditional distribution given a particular outcome of the first phase process.

Rubin-Bleuer (2000) and Rubin-Bleuer and Schiopu-Kratina (2002) formally defined a general space, which contains both the sampling design and the super-population that generates the finite population. This space, called the product probability space, is the product of the design probability space and the super-population (or model) space, with a sigma field defined by the product of the corresponding sigma fields, and a well-defined probability measure $P_{d,m}$. The "product space" changes with the size of the finite populations and samples, so as the sample size n goes to infinity (we assume $\liminf n/N > 0$ as $N \to \infty$, $n \to \infty$) we have to deal with a sequence of probability spaces.

In Section 2 we establish notation for different sample designs and for the design-model space mentioned above, and set conditions to be used later on in the article.

In Section 3 we prove that the sup norm of the sample empirical distribution function minus the model distribution function converges in probability to zero as $N \to \infty,\ n \to \infty$.

In Section 4 we study the sample empirical process and show weak convergence of the sample uniform empirical process.

2. DESIGN-MODEL SPACE

According to the super-population approach mentioned in the introduction, a sample statistic is a sample estimator of a finite population statistic, and hence subject to a randomization twice. Here we define a new probability space that will enable us to work with both randomizations at the same time. We follow the methodology of Rubin- Bleuer and Schiopu-Kratina (2002).

Let $(X^N,\ Z^N)$, $X^N = (X_j^N),\ j = 1,...,N,\ Z^N = (Z_j^N),\ j = 1,...,N,$ be random vectors (also called super-population) defined on a infinite probability space $(\Omega,\ \Im,\ P)$ and for $\omega \in \Omega$ let $(X^N(\omega),\ Z^N(\omega))$ be the data associated to the labels of a finite population of size N (we say that the finite population is generated by the super-population and $\omega \in \Omega$). We are interested in the sample estimator of the empirical distribution function of the random sample $X^N = (X_j^N),\ j = 1,...,N,$ and how it behaves as $N \to \infty$.

The random vectors Z^N will play a role in defining the design, since for a fixed outcome $\omega \in \Omega$, $Z^N(\omega)$ could be considered as "prior information", or information available at the time of the design.

In order to define the product space (joint model-design space), 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). Let S_N denote the collection of all possible samples under a sample scheme. Let $C(S_N)$ denote the collection of subsets of the sample space S_N. A sampling design p_{dN} is a function on $C(S_N)\times R^N$ such that for a fixed outcome $\omega \in \Omega$, and "prior information" $Z^N(\omega)$, it is a sampling probability distribution $p_{dN}(s,\omega) = p_{dN}(s, Z^N(\omega))$ and for a fixed $s \in S$, it is Borel-measurable in R^N.

Definition 2.1 Product space. We will assume here, without loss of generality, a generic product space derived from one design, and for its properties we refer to Definitions 4.1 to 4.3 in Rubin-Bleuer & Schiopu-Kratina (2002).

Let $(\Omega\times S_N,\ \Im\times C(S_N),\ P_{d,m})$ with $P_{d,m}(s\times F) = \int_F p_{dN}(s,\omega)\ dP(\omega)$ be the product space determined by the super-population $(X^N,\ Z^N)$ and a sampling design p_{dN} defined on $C(S_N)$.

Note that this space changes as the respective population sizes change. In what follows, we denote by E_d, E_m, $E_{d,m}$ the expectation with respect to the design probability space, the model space and the product space respectively.

Next we assume that we have a sample scheme with expected sample size equal to n and define a few of the more common some sample designs and conditions sufficient for the asymptotic properties to hold. For a "without replacement design" (WOR), let $I_j(s) \quad j = 1,..., N$ denote the sample s selection indicators. Let π_j denote the probability that unit j is selected to sample s, then I_j follows a binomial distribution $I_j \sim B(1, \pi_j) \quad j = 1,..., N$.

For a Probability Proportional to Size with Replacement (pps) scheme with units "sizes" $Z_j(\omega) \quad j = 1,..., N$, and selection probabilities $p_j = Z_j(\omega) / \sum_{l=1}^{N} Z_l(\omega) \quad j = 1,..., N$, let $0 \le J_j(s) \le n \quad j = 1,...,N$, denote the number of times unit j is selected to sample s. Thus $(J_1(s),...,J_N(s))$ follows a multinomial distribution $(J_1(s),...,J_N(s)) \sim MN(n, p_1,..., p_N)$.

For a two-stage design with (pps) in the first stage, let us first assume a joint model-design space defined in a super-population space where the sizes $M_j = Z_j(\omega), \quad j = 1,...,N$ are known a priori, i.e., a super-population space where the measure is the conditional probability

$P_M = P(\cdot \mid F_M), \quad F_M = \{\omega : M_j = Z_j(\omega), \quad j = 1,...,N\}$ (for more detail on this space, see Example 4.2 in Rubin-Bleuer & Schiopu-Kratina (2002)).

The design assumes independent selection in each in primary sampling unit (psu) under a WOR design with probability $\pi_{l|j}$ of selecting unit l in psu j, given that psu j was selected in the first stage, $l = 1,...,M_j, \quad j = 1,...,N$. The sample selection indicators of the second stage are defined conditionally, by $I_{jl}(s) = 1$ if unit l in psu j is selected to the sample given that psu j was selected to the sample in the first stage.

We consider the following conditions for the designs:

$C_0: \quad f = \lim_n n/N > 0 \ as \ n \to \infty$.

$C_1: \quad \lim_N E_m\left(\frac{1}{N}\sum_{j=1}^{N}\frac{1}{np_j}\right) < \infty \ , \ E_m\left(\max_{1\le i\le N}\frac{1}{np_j}\right) < \infty \ as \ n \to \infty$, and

$E_m(Z_j)^{2+\alpha} \le C, \quad j = 1,...,N, \quad N \ge 1$, for pps one-stage designs.

$C_2: \quad \lim_N E_m\left(\frac{1}{N}\sum_{j=1}^{N}\frac{1}{\pi_j}\right) < \infty$ for SRSWOR, Poisson or πps one-stage designs.

$$C_3: \ \frac{1}{M}\sum_{j=1}^{N}\frac{1}{np_j}\left(\sum_{l=1}^{M_j}\frac{1}{\pi_{l|j}}\right) = O(1) \ \ as \ \ n \to \infty.$$

Remark 2.1. We remark that the probability in conditions C_1 and C_2 are on expectations in the law of the super-population (P) to account for the cases where the design probabilities depend on $\omega \in \Omega$.. Note that for one-stage designs, we can define the sample space S_N and the product space $S_N \times \Omega$ before we know the "prior information" or outcome $Z^N(\omega)$ which completely define the design probabilities, since every sequence of (expected) n labels from the finite population has positive probability of being selected. However, for a two-stage design where the first–stage selection probabilities depend on "sizes" or the number $Z_j(\omega)$ of ultimate units in *psu* j, $j = 1,...,N$, the sizes must be known a priori for the sample space S_N (and hence the product space $S_N \times \Omega$), to be well-defined: a two-stage sample is a sequence of blocks of second stage sub-samples and we need to know how many labels are in each *psu* j to define all possible sequences of labels from it (see also Rubin-Bleuer & Schiopu-Kratina (2002), Example 4.2). In C_3 above, we stated an absolute bound (not in expectation), since once the sizes M_j are considered non-stochastic, the second stage selection probabilities from a SRSWOR, $\pi_{l|j} = m_j / M_j$, $l = 1,...,M_j$, $j = 1,...,N$ are also non-stochastic in Ω .

Remark 2.2 Condition C_0 ensures that the relationship between the sample and the population sizes (and its impact on the statistics considered) remains the same as we increase the population size towards infinity.

For SRSWOR designs, condition C_1 follows directly from C_0. For Poisson one-stage designs condition C_1 means that no probability is disproportionate as $N \to \infty$. For πps and pps one-stage designs, conditions C_1 and C_2 mean, respectively, that, as $N \to \infty$, the sizes are of the same magnitude in average. Indeed, if $\pi_j \approx nZ_j / \sum_{i=1}^{N} Z_i$, C_1 and C_2 will follow respectively, from C_0 and $\sum_{j=1}^{N} Z_j / N \xrightarrow{P} \mu$ (μ is non-stochastic in Ω) as $N \to \infty$. Similarly, for two-stage design, C_3 follows from C_0 and $\sum_{j=1}^{N} Z_j(\omega) / N \to \mu$ (μ is non-stochastic in Ω) as $N \to \infty$.

Definition 2.2. Empirical and Sample Empirical Distribution Functions. Let us assume the design-model space described above with $X^N = (X_j^N)$, $j = 1,...,N$, i.i.d.r.v.'s from a distribution function F .

Let $F_N(t,\omega) = \frac{1}{N}\sum_{j=1}^{N} I(X_j(\omega) \le t)$, $0 \le t \le \infty$, $\omega \in \Omega$, be the corresponding empirical distribution function and let a sample estimator of $F_N(t,\omega)$ be given by

$$\hat{F}_N(t,s,\omega) = \frac{1}{\hat{N}}\sum_{j=1}^{N} I(X_j(\omega) \le t)\delta_j(s), \quad \hat{N} = \sum_{j=1}^{N} \delta_j(s), \text{ where } \delta_j = I_j(s)/\pi_j \text{ for WOR one-stage}$$

designs and $\delta_j = J_j(s)/np_j$ for *pps* one-stage designs, $j = 1,...,N$. In the case of a *pps* two-stage design (with second stage SRSWOR) we have $F_N(t,\omega)) = \frac{1}{M}\sum_{j=1}^{N}\sum_{l=1}^{M_j} I(X_{jl}(\omega) \le t)$.

We define the sample empirical distribution function by

$\hat{F}_N(t,s,\omega) = \frac{1}{\hat{M}}\sum_{j=1}^{N}\sum_{l=1}^{M_j} I(X_{jl}(\omega) \le t)\delta_{jl}(s)$, $\hat{M} = \sum_{j=1}^{N}\sum_{l=1}^{M_j} \delta_{jl}(s)$, and the sample selection coefficients are $\delta_{jl} = (J_j(s)/np_j)I_{jl}(s)/\pi_{l|j}$, $l = 1,...,M_j$, $j = 1,...,N$.

3. A GLIVENKO-CANTELLI TYPE THEOREM

In this section we develop the product space version of the Glivenko-Cantelli Theorem. For this theorem the only design requirement is that the empirical distribution function be a design-consistent estimator of the finite population empirical distribution function.

Theorem 3.1. If the sample empirical distribution function is design-consistent, i.e., if

$$\hat{F}_N(t,s,\omega) - F_N(t,\omega) \to 0 \text{ in } p_{dN} \text{ as } N \to \infty, \quad \text{for all } \omega \in \Omega,\ 0 \le t < \infty, \text{ then}$$

$$\sup_{0 \le t < \infty} |\hat{F}_N(t,s,\omega) - F(t)| \to 0 \text{ in } P_{d,m} \text{ as } N \to \infty. \tag{3.1}$$

Proof: We define

$$F_N(t-,\omega) = \frac{1}{N}\sum_{j=1}^{N} I(X_j(\omega) < t) \text{ and } \hat{F}_N(t-,s,\omega) = \frac{1}{\hat{N}}\sum_{j=1}^{N} I(X_j(\omega) < t)\delta_j(s)$$

By the Glivenko-Cantelli theorem (see below for a reference) we have,

$$F_N(t,\omega) - F(t) \to 0 \text{ a.s. } (P) \text{ and } F_N(t-,\omega) - F(t-) \to 0 \text{ a.s. } (P) \text{ as } N \to \infty.$$

Now $\hat{F}_N(t,s,\omega)$ is design-consistent for all $\omega \in \Omega$, and $F_N(t,\omega)$ is model consistent, hence by Theorem 5.1 in Rubin-Bleuer and Schiopu-Kratina (2002), both $F_N(t,\omega)$ and $\hat{F}_N(t,s,\omega)$ are consistent in the product space. Thus

$$\hat{F}_N(t,s,\omega) - F(t) = \hat{F}_N(t,s,\omega) - F_N(t,\omega) + F_N(t,\omega) - F(t) \to 0 \text{ in } P_{d,m} \text{ as } N \to \infty. \tag{3.2}$$

A similar argument yields

$$\hat{F}_N(t-,s,\omega) - F(t-) \to 0 \ \ in \ \ P_{d,m} \ \ as \ \ N \to \infty. \tag{3.3}$$

We now follow, almost in its entirety, the proof of the Glivenko-Cantelli theorem given in Billingsley (1979) (Theorem 20.6, p.232).

Let $r \geq 2$ be an integer, and for $k = 1,2,...,r-1$, define $t_{r,k} = \min\{t : k/r \leq F(t)\}$, $t_{r,0} = -\infty$, $t_{r,r} = \infty$.

Note that the minimum exists and $F(t_{r,k}) \geq k/r$ because the distribution function is right continuous.

Now let t be any real number. Thus, there exist integers r, k such that $t_{r,k} \leq t < t_{r,k+1}$.

We have,

$\hat{F}_N(t,s,\omega) - F\ (t) \leq \hat{F}_N(t_{r,k+1}-,s,\omega) - F\ (t_{r,k})$, by the monotonicity of both $\hat{F}_N$ and F.

$$= \hat{F}_N(t_{r,k+1}-,s,\omega) - F\ (t_{r,k+1}-) + F(t_{r,k+1}-) - F(t_{r,k})$$

$$\leq \hat{F}_N(t_{r,k+1}-,s,\omega) - F\ (t_{r,k+1}-) + 1/r.$$

The above inequality follows from $F(t_{r,k}) \geq k/r$ and $F(t_{r,k+1}-) \leq (k+1)/r$. Indeed if $u < t_{r,k+1} \Rightarrow F(u) < (k+1)/r$, hence $F(t_{r,k+1}-) = \lim_{u \uparrow t_{r,k+1}} F(u) \leq (k+1)/r$.

By the same token, $-(F(t_{r,k+1}-) - F(t_{r,k})) \geq -1/r$ and

$$\hat{F}_N(t,s,\omega) - F\ (t) \geq \hat{F}_N(t_{r,k},s,\omega) - F\ (t_{r,k+1}-)$$

$$= \hat{F}_N(t_{r,k},s,\omega) - F\ (t_{r,k}) + F(t_{r,k}) - F(t_{r,k+1}-)$$

$$\geq \hat{F}_N(t_{r,k},s,\omega) - F\ (t_{r,k}) - 1/r.$$

Similarly, for $-\infty < t \leq t_{r,1}$ we have $|\hat{F}_N(t,s,\omega) - F\ (t)| \leq |\hat{F}_N(t_{r,1}-,s,\omega) - F\ (t_{r,1}-0)| + 1/r$, and

for $t_{r,r-1} \leq t < \infty$ we have $|\hat{F}_N(t,s,\omega) - F\ (t)| \leq |\hat{F}_N(t_{r,r-1},s,\omega) - F\ (t_{r,r-1})| + 1/r$.

Thus, for any real number t and all $r \geq 2$,

$|\hat{F}_N(t,s,\omega) - F\ (t)| \leq D_{r,N}(s,\omega) + 1/r$, where

$$D_{r,N}(s,\omega) = \max_{1 \leq k \leq r,\ 1 \leq j \leq r} \{|\hat{F}_N(t_{r,k},s,\omega) - F\ (t_{r,k})|; |\hat{F}_N(t_{r,j}-,s,\omega) - F\ (t_{r,j}-)|\}.$$

Hence

$$\sup_{-\infty<t<\infty} |\hat{F}_N(t,s,\omega) - F(t)| \le D_{r,N}(s,\omega) + 1/r, \text{ for all } r \ge 2,\ N \ge 1.$$

Now we first show that $D_{r,N}(s,\omega) \to 0$ *in* $P_{d,m}$ *as* $N \to \infty$, for all $r \ge 2$. Indeed, for a fixed $r \ge 2$, we apply (3.2) and (3.3) to the finite number of sequences converging to zero in the product space:

$$|\hat{F}_N(t_{r,k},s,\omega) - F(t_{r,k})| \overset{P_{d,m}}{\to} 0 \text{ and } |\hat{F}_N(t_{r,k}-,s,\omega) - F(t_{r,k}-)| \overset{P_{d,m}}{\to} 0 \text{ as } N \to \infty\ 1 \le k \le r.$$

Hence the maximum $D_{r,N}(s,\omega)$ of these sequences also converges to zero in probability. This implies the thesis of the theorem, since for any positive ε and δ, there exists an r with $1/r < \varepsilon/2$ and an $N_0(r,\varepsilon,\delta)$ and a such that for $N \ge N_0$,

$$P_{d,m}(\sup_{-\infty<t<\infty} |\hat{F}_N(t,s,\omega) - F(t)| \ge \varepsilon) \le P_{d,m}(D_{r,N}(s,\omega) + 1/r \ge \varepsilon) \le P_{d,m}(D_{r,N}(s,\omega) \ge \varepsilon/2) \le \delta \bullet$$

4. TIGHTNESS OF THE SAMPLE EMPIRICAL PROCESS

In the super-population or model space, the uniform empirical process converges weakly to a Brownian Bridge (tied down Wiener process) which is a Gaussian process with continuous sample paths and covariance function $E_m(B(u)B(r)) = u \wedge r - ur$.

In this section we are concerned with the sample counterpart of the uniform empirical process, that is, with the sample uniform empirical process, and with its asymptotic properties. We shall see that it also converges weakly to a Gaussian process, but its covariance function will depend on the design on which the sample process is based.

We consider for now the one-stage designs defined in Section 2, where conditions C_0 and C_1 or C_2 hold., and we will show convergence of the finite dimensional distributions and tightness of the sample uniform empirical process under the corresponding design.

Definition 4.1 Sample Uniform Empirical Process. The model uniform empirical process is given by

$$\alpha_N(t,\omega) = \sqrt{n}(F_N(t,\omega) - t), \quad -\infty < t < \infty, \quad \omega \in \Omega,$$

where the empirical distribution function $F_N(t,s,\omega)$ is based on an uniform random sample, i.e., random variables $X_1,...,X_N \approx$ uniform $U(0,1)$ which are conditionally independent given the prior information (which is, for example, $Z_1(\omega),...,Z_N(\omega)$ for the p.p.s. design described in Section 2 and none for SRSWOR).

The sample uniform empirical process is defined by

$$\hat{\alpha}_N(t,s,\omega) = \sqrt{n}(\hat{F}_N(t,s,\omega) - t), \quad -\infty < t < \infty, \quad s \in S,\ \omega \in \Omega, \tag{4.1}$$

where the sample empirical distribution function $\hat{F}_N(t,s,\omega)$ is a consistent estimator of the empirical distribution function.

Theorem 4.1. Covergence of the Finite Dimensional Distributions. The finite dimensional distributions of the sample uniform empirical process, under designs for which the sample ratio

estimator of the mean is asymptotically normal, converge in distribution to those of a Gaussian Process $\{B_d(t): \ 0 \le t \le 1\}$ with covariance function depending on the corresponding sample design. When the design is one-stage p.p.s as defined in Section 2, under conditions C_0 and C_1 the covariance function is given by

$$E_m(B_d(u)B_d(r)\} = f \cdot (u \wedge r - ur) + f \cdot \lim_N E_m(\frac{1}{N}\sum_i \frac{1}{np_i}) \cdot u \wedge r - ur .$$

Proof:

$\alpha_N(t,\omega)/\sqrt{n} = \frac{1}{N}\sum_i (I_i(t) - t)$ is a finite population mean of which $\hat{\alpha}_N(t,s,\omega)/\sqrt{n}$ is the ratio estimator. Hence $\hat{\alpha}_N(t,s,\omega)$ converges in the law of the product space to the random variable $B_d(t) = \sigma(t)\aleph(0,1)$ where $\aleph(0,1)$ is a standard normal random variable and

$$\sigma^2(t) = ft(1-t) + f \lim_N E_m(\frac{1}{N}\sum_i \frac{1}{np_i})t - t^2 . \qquad (4.2)$$

The asymptotic normality of the sample uniform empirical distribution function for every real number t follows from Example 6.1 of Rubin-Bleuer & Schiopu-Kratina (2002). The variance and covariances can be calculated from $\hat{\alpha}_N(t,s,\omega)$ directly, using Slutzky's theorem and letting $N \to \infty$ •

Theorem 4.2. Tightness. Under one-stage designs satisfying conditions C_0 and C_1 or C_2 respectively, the sample uniform empirical process is tight, i.e., for each positive ε and η, there exists a δ, $0 < \delta < 1$, and a N_0 such that for $N \ge N_0$,

$$P_{d,m}\{\sup_{u \le t \le u+\delta} |\hat{\alpha}_N(t,s,\omega) - \hat{\alpha}_N(u,s,\omega)| \ge \varepsilon\} \le \delta\eta . \qquad (4.3)$$

Proof: For notational convenience we take $u = 0$, and this is no restriction since the increments of $\hat{\alpha}_N(t,s,\omega)$ are stationary. Thus we will prove

$$P_{d,m}\{\sup_{t \le \delta} |\hat{\alpha}_N(t,s,\omega)| \ge \varepsilon\} \le \delta\eta \qquad (4.4)$$

The proof follows closely the techniques used to prove tightness of the uniform empirical process in Billingsley (1968), Theorem 13.1, p.105-108.
Suppose that for all $0 \le r \le 1$, $0 \le t + r \le 1$ and $0 \le t + r + u \le 1$, there exists a constant C independent of N, t, r, u such that we have:

$$E_{d,m}\{|\hat{\alpha}_N(t+r) - \hat{\alpha}_N(t)|^2 \cdot |\hat{\alpha}_N(t+r+u) - \hat{\alpha}_N(t+r)|^2\} \le C \cdot r \cdot u . \qquad (4.5)$$

Then, for fixed δ, we deduce, following the same steps of Billingsley (1968) p.107, that

$$P_{d,m}\{\max_{1 \le i \le m} |\hat{\alpha}_N(i\delta/m)| \ge \varepsilon\} \le K \cdot C \cdot \delta^2/\varepsilon^4 + P_{d,m}\{|\hat{\alpha}_N(\delta)| \ge \varepsilon/2\}, \qquad (4.6)$$

where C is the constant in (4.5), and K is another constant independent of ε, δ, N. We continue as in Billingsley, to state that since for each $s \in S$, $\omega \in \Omega$, $\hat{\alpha}_N(t,s,\omega)$ is right continuous in t, as $m \to \infty$,

we have $\max_{1 \le i \le m} |\hat{\alpha}_N(i\delta/m)| \to \sup_{t \le \delta} |\hat{\alpha}_N(t,s,\omega)|$ for each $s \in S$, $\omega \in \Omega$. (4.7)

Now, looking at the second term in (4.6) we observe that the sample uniform empirical process evaluated at $t = \delta$, approaches $B_d(\delta)$ as $N \to \infty$, where $B_d(\delta)$ is defined in Theorem 4.1.

Hence $P_{d,m}\{|\hat{\alpha}_N(\delta)| \ge \varepsilon/2\} \to P\{B(\delta) \ge \varepsilon/2\} \le \frac{E(\aleph^4(0,1))2^4\delta^2}{\varepsilon^4} = \frac{42\delta^2}{\varepsilon^4}$ as $N \to \infty$. (4.8)

Thus given ε and η, choose δ so that $\frac{K \cdot C + 42}{\varepsilon^4}\delta^2 < \delta\eta$. Let now N_δ be such that the inequality in (4.8) holds for $N \ge N_\delta$. Hence together with (4.6) (4.7) and (4.8) we have for $N \ge N_\delta$:

$$P_{d,m}\{\sup_{t \le \delta} |\hat{\alpha}_N(t,s,\omega)| \ge \varepsilon\} \le \frac{K \cdot C + 42}{\varepsilon^4}\delta^2 < \delta\eta,$$

which is (4.4), the statement of the theorem.

To complete the theorem we have to prove inequality (4.5) which we previously assumed to hold. For this we have to account for the differences between the sample and the finite population empirical processes and the fact that the sample counterpart is not a sum of stochastically independent random variables in the product space. We first change the notation to simplify the look of the long equations we will develop. We set:

$$\alpha_i = I(t < X_i(\omega) \le t + r)\delta_i(s) - r, \quad \beta_i = I(t + r < X_i(\omega) \le t + r + u)\delta_i(s) - u, \ i = 1,...,N.$$

We have: $E_{d,m}(\alpha_i) = 0$, $E_{d,m}(\beta_i) = 0$, $i = 1,...,N$.

By condition C_0 there exists N_0 such that for $N \ge N_0$, $N/n \le 2/f$ and hence for $N \ge N_0$ inequality (4.5) is equivalent to:

$$E_{d,m}((\sum_i \alpha_i)^2 \cdot (\sum_j \beta_j)^2) \le N^2(2C/f) \cdot ru \qquad (4.9)$$

Unless the design is Poisson, the α_i, $i = 1,...,N$ are not stochastically independent in the product space. Similarly with the β_i $i = 1,...,N$. This implies that many terms in (4.9) do not cancel out as in the proof for the super-population uniform empirical distribution process. But we will see that those terms are small enough to make inequality (4.9) hold.

Let us further simplify the look of the equations by setting:

$I_i(r) = I(t < X_i(\omega) \le t + r)$, $I_i(u) = I(t + r < X_i(\omega) \le t + r + u)$, $i = 1,...,N$.

Now we have:

$$(\sum_i \alpha_i)^2 \cdot (\sum_j \beta_j)^2 = (\sum_i \alpha_i^2 + \sum_i \sum_{j \ne i} \alpha_i\alpha_j) \cdot (\sum_i \beta_i^2 + \sum_i \sum_{j \ne i} \beta_i\beta_j)$$

$$= (\sum_i \alpha_i^2) \cdot (\sum_k \beta_k^2) + (\sum_k \sum_i \sum_{j \neq i} \alpha_i \alpha_j \beta_k^2) + (\sum_k \sum_i \sum_{j \neq i} \beta_i \beta_j \alpha_k^2) + (\sum_i \sum_{j \neq i} \sum_k \sum_{k \neq l} \alpha_i \alpha_j \beta_k \beta_l)$$

$$= S_1 + S_2 + S_3 + S_4.$$

$$S_1 = \sum_i \alpha_i^2 \beta_i^2 + \sum_i \sum_{i \neq k} \alpha_i^2 \beta_k^2 \text{ and}$$

$$\alpha_i^2 \beta_i^2 = (I_i(r)\delta_i^2(s) - 2rI_i(r)\delta_i(s) + r^2) \cdot (I_i(u)\delta_i^2(s) - 2uI_i(u)\delta_i(s) + u^2)$$

$$= u^2 I_i(r)\delta_i^2(s) - 2ru^2 I_i(r)\delta_i(s) + r^2 I_i(u)\delta_i^2(s) - 2ur^2 I_i(u)\delta_i(s) + r^2u^2,$$

since $I_i(r) \cdot I_i(u) = 0$. Now

$$E_{d,m}(\alpha_i^2 \beta_i^2) = u^2 r E_d(\delta_i^2(s)) - 2r^2u^2 E_d(\delta_i(s)) + r^2 u E_d(\delta_i^2(s)) - 2u^2 r^2 E_d(\delta_i(s)) + r^2u^2$$

For SRSWOR and Poisson designs the design expectations of the sample indicators are bounded as $N \to \infty$. For the $p.p.s.$ design, the expectations of the $\delta_i(s) = J_i(s)/np_i$ follow from the respective expectations of $J_i(s)$ given in the appendix. We have:

$$E_d(\delta_i(s)) = 1, \; E_d(\delta_i^2(s)) = 1 - 1/n + 1/np_i, \; i = 1,\ldots,N.$$

and by condition C_1, $1/np_i$ is uniformly bounded in i as $n \to \infty$. Hence

$$E_{d,m}(\sum_i \alpha_i^2 \beta_i^2) \leq N \cdot B_1 \cdot ru \qquad (4.10)$$

Now $\alpha_i^2 \beta_k^2$ is a finite sum of terms which are products of ru times expectations of product combinations of δ_i's and for all three one-stage designs considered here, these are uniformly bounded in i as $n \to \infty$.

Since there are about $N(N-1)$ terms of the form $\alpha_i^2 \beta_k^2$, there is a constant B_2 independent on N, r or u such that

$$E_{d,m}(\sum_i \sum_{i \neq k} \alpha_i^2 \beta_k^2) \leq N^2 \cdot B_2 \cdot ru \qquad (4.11)$$

From (4.10) and (4.11) we get

$$E_{d,m}(S_1) \leq N^2 \cdot C_1 \cdot ru \qquad (4.12)$$

Now consider $S_2 = (\sum_k \sum_i \sum_{j \neq i} \alpha_i \alpha_j \beta_k^2)$. In the proof for the super-population uniform empirical process advantage is taken on the independence of the α_i's and of the β_i's and some terms cancel out. But here we have a number of terms of the order of N^3, so having each term bounded as $n \to \infty$ is not enough to obtain a bound in $O(N^2) \cdot ru$.

We will show that

$$E_{d,m}(\beta_k^2 \alpha_i \alpha_j) = O(1/N) \cdot ru \text{ for all } k,\ i \neq j,\ i,j,k = 1,...,N. \qquad (4.13)$$

Indeed, assume first that $k \neq i,\ k \neq j$. Note that for a Poisson design, the β_k, α_j and $\alpha_i\ i \neq j,\ k \neq i,\ k \neq j$ are stochastically independent and

$E_{d,m}(\alpha_i) = 0,\ E_{d,m}(\beta_i) = 0,\ i = 1,...,N.$ Hence $E_{d,m}(\beta_k^2 \alpha_i \alpha_j) = 0.$

Now we look at *p.p.s.* and SRSWOR designs.

$$\beta_k^2 \alpha_i \alpha_j = (I_k(u)\delta_k^2 - 2uI_k(u)\delta_k + u^2)(I_i(r)I_j(r)\delta_i\delta_j - r(I_j(r)\delta_j + I_i(r)\delta_i) + r^2)$$

Hence $E_{d,m}(\beta_k^2 \alpha_i \alpha_j) = E_Z(E\ (E_d(\beta_k^2 \alpha_i \alpha_j) \mid Z)) =$

$$= E_Z\{r^2 u E_d(\delta_k^2\delta_i\delta_j) - 2r^2u^2E_d(\delta_k\delta_i\delta_j) + r^2u^2E_d(\delta_i\delta_j) - r^2uE_d(\delta_k^2\delta_j) - r^2uE_d(\delta_k^2\delta_i)$$

$$+ 2r^2u^2E_d(\delta_k\delta_j) + 2r^2u^2E_d(\delta_k\delta_i) - r^2u^2E_d(\delta_j) - r^2u^2E_d(\delta_i) + r^2uE_d(\delta_k^2) - 2r^2u^2E_d(\delta_k) + r^2u^2\}$$

$$= E_Z\{r^2u \cdot E_d(\delta_k^2\delta_i\delta_j - \delta_k^2\delta_j - \delta_k^2\delta_i + \delta_k^2) + r^2u^2E_d(-2\delta_k\delta_i\delta_j + \delta_i\delta_j + 2\delta_k\delta_j + 2\delta_k\delta_i - 2\delta_k - \delta_j - \delta_i + 1)\}$$

Now the expectation in the first term above for a *p.p.s.* design is

$$E_d(\delta_k^2\delta_i\delta_j - \delta_k^2\delta_j - \delta_k^2\delta_i + \delta_k^2) = (1-\frac{1}{n})(1-\frac{2}{n})(1-\frac{3}{n}) + (1-\frac{1}{n})(1-\frac{2}{n})\frac{1}{np_k} - 2(1-\frac{1}{n})(1-\frac{2}{n}) - 2(1-\frac{1}{n})\frac{1}{np_k}$$

$$-(1-\frac{1}{n}) + \frac{1}{np_k} = O(\frac{1}{n}).$$

Indeed, the terms equal to 1 above all cancel each other since there are same number of positive and negative 1's. There are two terms of order $\frac{1}{np_k}$ with coefficients 1 and -1 respectively, so they cancel each other. Thus we have left with terms of order $O(\frac{1}{n})$, some of the with coefficients $\frac{1}{np_k}$, but these are bounded as $n \to \infty$ by condition C_1. And condition C_0 yields

$O(\frac{1}{n}) = O(\frac{1}{N})$ as $n \to \infty$. Similarly, for SRSWOR we also obtain an $O(\frac{1}{N})$ under condition C_0.

The expectation of the second term above for a *p.p.s.* design is

$$E_d(-2\delta_k\delta_i\delta_j + \delta_i\delta_j + 2\delta_k\delta_j + 2\delta_k\delta_i - 2\delta_k - \delta_j - \delta_i + 1) =$$

$$= -2(1-\frac{1}{n})(1-\frac{2}{n}) + (1-\frac{1}{n}) + 2(1-\frac{1}{n}) + 2(1-\frac{1}{n}) - 2 - 1 - 1 + 1 = \frac{1}{n} + \frac{4}{n^2} = O(\frac{1}{n}).$$

And similarly we obtain the same rate of convergence for SRSWOR designs. Hence (4.13) is verified for these terms.

If the index $k = i$ or $k = j$ then some terms disappear (since $I_i(r) \cdot I_i(u) = 0$) and the terms left out are bounded in expectation by a constant times ru. Now there are only $2N(N-1)$ of these terms, and $N(N-1)(N-2)$ terms that verify (4.13). Thus there exists a constant C_2 independent on N, r, u, such that

$$E_{d,m}(S_2) \le N^2 \cdot C_2 \cdot ru \qquad (4.14)$$

Similarly, we obtain

$$E_{d,m}(S_3) \le N^2 \cdot C_3 \cdot ru \qquad (4.15)$$

Now the term $S_4 = \sum_i \sum_{j \neq i} \sum_k \sum_{k \neq l} \alpha_i \alpha_j \beta_k \beta_l$ consists of about N^4 terms. We will show that each term

$E_{d,m}(\alpha_i \alpha_j \beta_k \beta_l) = O(\frac{1}{N^2})$ as $N \to \infty$ for all distinct indexes i, j, k, $l = 1,...,N$. As we found with S_3, Poisson designs imply that these terms are all zero. For the other designs we note that there are $N(N-1)(N-2)(N-3)$ of these terms and at most $4N(N-1)$ terms with a repeated index, which are at most

$$\beta_k \beta_l \alpha_i \alpha_j = I_i(r) I_j(r) I_k(u) I_l(u) \delta_k \delta_j \delta_i \delta_l - r I_j(r) I_k(u) I_l(u) \delta_k \delta_j \delta_l - r I_i(r) I_k(u) I_l(u) \delta_k \delta_l \delta_i \quad O(1).$$

$$+ r^2 I_k(u) I_l(u) \delta_k \delta_l - u I_i(r) I_j(r) I_l(u) \delta_i \delta_j \delta_l + ru I_j(r) I_l(u) \delta_j \delta_l + ru I_i(r) I_l(u) \delta_i \delta_l - r^2 u I_l(u) \delta_l$$

$$- 2u I_i(r) I_j(r) I_k(u) \delta_i \delta j \delta_k + ur I_j(r) I_k(u) \delta_j \delta_k + ur I_i(r) I_k(u) \delta_i \delta_k - r^2 u I_k(u) \delta_k + u^2 I_i(r) I_j(r) \delta_i \delta_j$$

$$- ur^2 I_j(r) \delta_j - ur^2 I_i(r) \delta_i + u^2 r^2.$$

Thus for a *p.p.s.* sample design we have:

$$E_{d,m}(\beta_k \beta_l \alpha_i \alpha_j) = r^2 u^2 E_d \{\delta_i \delta_j \delta_k \delta_l - \delta_j \delta_k \delta_l - \delta_i \delta_k \delta_l + \delta_k \delta_l - \delta_i \delta_j \delta_l + \delta_j \delta_l + \delta_i \delta_l - \delta_l$$
$$- \delta_i \delta_j \delta_k + \delta_j \delta_k + \delta_i \delta_k - \delta_k + \delta_i \delta_j - \delta_i - \delta_j + 1\}$$

$$= r^2 u^2 \{\frac{n(n-1)(n-2)(n-3)}{n^4} - 4\frac{n(n-1)(n-2)}{n^3} + 6\frac{n(n-1)}{n^2} - 4 + 1 = O(\frac{1}{n^2}).$$

And we obtain the same rate of convergence for a SRSWOR design. Hence, there is a constant C_4 such that

$$E_{d,m}(S_4) \le N^2 \cdot C_4 \cdot ru \qquad (4.16)$$

Thus (4.10), (4.14), (4.15 and (4.16) together imply (4.9) which is equivalent to (4.5) and the theorem is proved •

APPENDIX

We state some expectations in the design space of the multinomial random vector

$(J_1(s),...,J_N(s)) \sim MN(n, p_1,..., p_N)$.

$E_d(J_i) = np_i, \ \ i = 1,2,...,N$

$E_d(J_i^2) = n(n-1)p_i^2 + np_i, \ \ i = 1,2,...,N$

$E_d(J_i J_k) = n(n-1)p_i\, p_k, \ \ i,k = 1,2,...,N$

$E_d(J_i J_k J_h) = n(n-1)(n-2)p_i\, p_k p_h, \ \ i,k,h = 1,2,...,N$

$E_d(J_i^2 J_k) = n(n-1)(n-2)p_i^2 p_k + n(n-1)p_i p_k, \ \ i,k = 1,2,...,N$

$E_d(J_i^2 J_k^2) = n(n-1)(n-2)(n-3)p_i^2 p_k^2 + n(n-1)(n-2)(p_i^2 p_k + p_i p_k^2) + n(n-1)p_i p_k, \ \ i,k = 1,2,...,N$

$E_d(J_i^2 J_k J_h) = n(n-1)(n-2)(n-3)p_i^2 p_k p_h + n(n-1)(n-2)p_i p_k p_h, \ \ i,k,h = 1,2,...,N$

$E_d(J_i J_k J_h J_l) = n(n-1)(n-2)(n-3)p_i\, p_k p_h p_l \ \ i,k,h,l = 1,2,...,N$

REFERENCES

Billingsley, P.(1968). *Covergence of Probability Measures*, Wiley, New York.

Billingsley, P.(1979). *Probability and Measure*, Wiley, New York.

Csörgö, M. and Révész, P. (1981). *Strong Approximations in Probability and Statistics*. Academic Press, Inc., New York, and Akadémiai Kiado, Budapest.

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.

Lin, D.Y. (2000). On fitting Cox's proportional hazards models to survey data. Biometrika, 87, 1.

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, June 2001*

Rubin Bleuer, S. and Schiopu Kratina, I. (2002). On the two-phase framework for joint model and design-based inference. *Technical Report Series of the Laboratory for Research in Statistics and Probability, Number 382, Carleton University-University of Ottawa.*

Särndal, C-E, Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*, Springer-Verlag, New York.