An empirical-process central limit theorem for complex sampling under bounds on the design effect

Thomas Lumley[1]

[1] *Department of Statistics, Private Bag 92019, Auckland 1142, New Zealand. e-mail:* t.lumley@auckland.ac.nz

Abstract: Uniform central limit theorems ('Donsker theorems') have been widely useful in semiparametric statistics, both under iid sampling and for stationary sequences and random fields. Only limited results have been available under complex sampling, especially multistage sampling. In this note we derive a complex-sampling analogue of Ossiander's bracketing-entropy conditions for a uniform central limit theorem, under the assumption that certain design effects are uniformly bounded. We discuss the plausibility of this assumption in realistic surveys.

Keywords and phrases: Donsker theorem, uniform central limit theorem, survey sampling, bracketing entropy.

1. Introduction

Data from complex survey samples such as the NHANES series in the US or the British Household Panel Survey in the UK are increasingly being used for a wide range of secondary statistical analyses. Observations in these surveys are sampled with unequal (but known) sampling probabilities. The sampling is typically correlated, with negative correlations induced by stratified sampling and positive correlations induced by cluster sampling.

Data analysts want to use all the same statistical methods for complex samples that they use for independently-sampled data. For some of these statistical methods, the only convenient theoretical techniques involve uniform central limit theorems either for the joint empirical distribution of the data or for other classes of functions. Two examples are the Cox proportional hazards model[17, 3], and rank tests[18]. In iid data such results have been available for some time[22, 21], and there are many extensions to dependent sequences[eg 1, 6, 8], but results for complex samples have been very limited.

A uniform CLT for a one-dimensional empirical cumulative distribution function under simple random sampling without replacement is classical; Shorack [25, Theorem 16.2.3] gives a simple proof. There is an immediate extension to stratified random sampling with a fixed number of strata and stratum sizes growing without limit. Breslow and Wellner [3], using the exchangeable bootstrap [23], showed that the uniform CLT for stratified random sampling extended to classes of functions controlled by uniform entropy or bracketing entropy, giving the same bounds as under iid sampling. Saegusa and Wellner [24] derive the limits after calibration of weights [7].

Wang [28] gave conditions on fourth-order joint sampling probabilities for the one-dimensional empirical CDF, based on conditions of Breidt and Opsomer [2]. These conditions are satisfied by reasonable stratified or PPS samples of individuals, but are not satisfied under single-stage or multi-stage cluster sampling. Extensions to higher dimensions or to other classes of functions are not immediate. Under similar assumptions about sampling, Cardot and Josserand [4] proved a CLT for functional data satisfying Hölder-type conditions, using a maximal inequality based on packing numbers.

In this note we give a simple proof of a bracketing-entropy Donsker theorem for data from complex sampling designs, controlling the impact of dependence by the design effect. We work in a two-phase or superpopulation context, where a sequence of finite populations is drawn from a superpopulation distribution and samples are then drawn from the finite populations. We expect that the same approach could be used for inference conditional on the sequence of finite populations, but strict finite-population inference is typically of less interest for tests and estimators complicated enough to need empirical-process arguments to justify them. The proof uses an existing moment bound for increments of the empirical process in iid data, which follows from the same adaptive truncation argument as Ossiander's theorem. We then translate the bound into a bound on second moments under complex sampling, implying stochastic equicontinuity.

The design effect is a concept introduced by Kish [14, 15] to summarise the efficiency of a complex survey design for estimation of a particular population quantity, analogous to the Pitman relative efficiency. If $\hat{\theta}$ estimates a population quantity θ under a complex survey design, and $\tilde{\theta}$ estimates the same population quantity in a simple random sample with the same sample size, the design effect is $\mathrm{var}[\hat{\theta}]/\mathrm{var}[\tilde{\theta}]$. Two variants are in routine use: DEFF, comparing to a simple random sample without replacement, and DEFT, comparing to iid sampling; we will use the latter.

In this paper we assume that the design effect is uniformly bounded for all differences of functions in the class at all sample sizes. This is a strong assumption mathematically, on the other hand, a bounded design effect is a familiar and plausible assumption from the viewpoint of survey practitioners. The only reason to have design effects exceeding unity in complex surveys is to reduce overall cost, so tradeoffs between design effects for a variety of estimands and survey costs are explicitly considered in the design of large surveys. Choosing a survey design with very large design effects for any statistic of interest would only be sensible if the cost per unit sampled were extremely low compared to simple random sampling. Since the cost of performing an interview gives a non-negligible lower bound on the cost per unit sampled, very large design effects are not plausible in well-designed surveys. For example, in cluster sampling suppose that the cost of an additional unit in the same cluster is lower than the cost of a new cluster by a factor γ. The design effect of a design that minimizes cost for a given precision will then be less than $\gamma^{-1/2}$[5].

In section 3 we give some sufficient conditions for a design-effect bound, compare it to conditions for β-mixing sequences, and also evaluate the assumption in survey data examples.

2. Theorem and Proofs

Each population in the sequence of populations of size N_n is an iid sample from a multivariate superpopulation distribution P. A sample (of expected size n) is drawn, where the probability of

sampling unit i, conditional on the realized finite population, is π_i and the probability of sampling both unit i and unit j is π_{ij}. The sampling probabilities π_i and π_{ij} must be non-zero for all i, j in the population, and they must be known for all i, j in the sample.

The key tool is the Horvitz–Thompson estimator of a population mean[12], using sampling weights $1/\pi_i$ to remove the bias from unequal-probability sampling

$$\hat{T}_X = \frac{1}{N_n} \sum_{i \in \text{sample}} \frac{1}{\pi_i} X_i.$$

The variance of $\hat{T}_X$ conditional on the finite population is estimated by

$$\widehat{\text{var}}[\hat{T}_X] = \frac{1}{N_n^2} \sum_{i,\, j \in \text{sample}} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{x_i}{\pi_i} \frac{x_j}{\pi_j}.$$

We assume the sequence of sampling designs giving samples of expected size n is such that

Assumption D A central limit theorem holds (pointwise) at $\sqrt{n}$ rate: ie, for every function $f \in \mathcal{F}$

$$\frac{\sqrt{n}}{N} \sum_{i \in \text{sample}} (f(X_i) - Pf(X)) \xrightarrow{d} N(0, \sigma_f^2)$$

This assumption is routine in survey statistics, and Fuller [9, Chapter 1] collects some sets of conditions. We also assume the superpopulation distribution satisifies

Assumption P The envelope function F of $\mathcal{F}$ has a finite fourth moment: $\|F\|_{P,4} < \infty$

The technical results we need, and their proofs, are conveniently collected in section 2.14 of van der Vaart and Wellner [27], so we follow their notation. Define the bracketing entropy integral of a class of functions $\mathcal{F}$ from an index set T to $\mathbb{R}$ as

$$J_{[]}(\delta, \mathcal{F}, \|\cdot\|) = \int_0^\delta \sqrt{1 + \log N_{[]}(\epsilon \|F\|,\, \mathcal{F},\, \|\cdot\|}\, d\epsilon$$

where $F(x) = \sup_{f \in \mathcal{F}} f(x)$ is the envelope function and $N_{[]}()$ are the bracketing numbers of $\mathcal{F}$ with respect to the norm $\|\cdot\|$.

The empirical process under iid sampling from the superpopulation is

$$\mathbb{G}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (f(X_i) - Pf)$$

and its complex-sampling analogue is the the Horvitz–Thompson process[17], augmented with sampling weights

$$\mathbb{G}_n^\pi f = \frac{\sqrt{n}}{N} \sum_{i \in \text{sample}} \frac{1}{\pi_i} (f(X_i) - Pf).$$

The design effect for f relative to superpopulation simple random sampling is thus $\text{var}[\mathbb{G}_n^\pi f]/\text{var}[\mathbb{G}_n f]$.

We quote two results proved by van der Vaart and Wellner [27]. Proposition 1 is similar to maximal inequalities used by Kim and Pollard [13], and Proposition 2 is based on results for Orlicz norms in Talagrand [26]. The notation $a \lesssim b$ means there is a constant C such that $a \le Cb$; the constant is universal in Proposition 1 and depends only on q in Proposition 2.

Proposition 1 (Theorem 2.14.2, van der Vaart and Wellner)**.** *Let*

$$a(\eta) = \frac{\eta \|F\|_{P,2}}{\sqrt{1 + \log N_{[]}(\eta \|F\|_{P,2},\, \mathcal{F},\, L_2(P))}}$$

If $\|f\|_{P,2} < \delta \|F\|_{P,2}$ *for all* $f \in \mathcal{F}$ *then*

$$\left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_n f\|^* \right\|_{P,1} \lesssim J_{[]}(\delta, \mathcal{F}, L_2(P)) \|F\|_{P,2} + \sqrt{n} P\left[F\{F > \sqrt{n} a(\delta)\}\right]$$

□

Proposition 2 (Theorem 2.14.5, van der Vaart and Wellner)**.** *For any* $q \ge 2$

$$\left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_n f\|^* \right\|_{P,q} \lesssim \left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_n f\|^* \right\|_{P,1} + n^{-1/2+1/q} \|F\|_{P,q}$$

□

We now state our main result.

Theorem 3. *Let* $\mathcal{F}$ *be a class of functions with finite bracketing entropy integral, satisfying assumptions D and P*

Define

$$\mathcal{F}_\delta = \{f - g : f,\, g \in \mathcal{F},\, \rho_2(f,g) \le \delta\}$$

If $\mathcal{F}_\delta$ *has uniformly bounded design effects for the mean, relative to superpopulation simple random sampling, then* $\mathbb{G}_n^\pi$ *converges weakly to a mean-zero Gaussian process indexed by* $\mathcal{F}$*.*

By Lyapunov's inequality, Proposition 2 implies for any $q \ge 2$

$$\left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_n f\|^* \right\|_{P,2} \lesssim \left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_n f\|^* \right\|_{P,q} \lesssim \left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_n f\|^* \right\|_{P,1} + n^{-1/2+1/q} \|F\|_{P,q}$$

Combining with Proposition 1 we obtain

$$\left\| \sup_{f\in\mathcal{F}} |\mathbb{G}_n f\|^* \right\|_{P,2} \lesssim J_{[]}(\delta, \mathcal{F}, L_2(P)) \|F\|_{P,2}) + \sqrt{n} P\left[F\{F > \sqrt{n} a(\delta)\}\right] + n^{-1/2+1/q} \|F\|_{P,q} \qquad (1)$$

Let $\rho_2(f,g) = \|Gf - Gg\|_{P,2}$ be the standard deviation semimetric on the index set induced by the limiting Gaussian process. $\mathbb{G}_n$ is asymptotically ρ_2-equicontinuous in probability, ie, for any ϵ, $\eta > 0$ there exists $\delta > 0$ such that

$$\limsup_{n\to\infty} P^* \left[\sup_{\rho_2(f,g)<\delta} |\mathbb{G}_n f - \mathbb{G}_n g| > \epsilon \right] < \eta.$$

We can thus construct, for any $\epsilon, \eta > 0$, a partition $\{T_i\}_{i=1}^k$ of the index set with $\rho_2(f,g) < \delta$ for $f, g \in T_i$ and

$$\limsup_{n\to\infty} P^* \left[\sup_i \sup_{f,g\in T_i} |\mathbb{G}_n f - \mathbb{G}_n g| > \epsilon \right] < \eta.$$

Now, we use the bounds in equation 1 applied to the class of increments

$$\mathcal{F}_\delta = \{f - g : f,\, g \in \mathcal{F},\, \rho_2(f,g) \le \delta\}$$

By assumption P we may take $q = 4$, so the last two terms go to zero as $n \to \infty$ for any fixed δ, and the first term goes to zero as $n \to \infty$ followed by $\delta \to 0$.

The first two terms go to zero as $n \to \infty$ followed by $\delta \to 0$. The third term also goes to zero for $q = 4$ by assumption P1. Thus

$$\limsup_{n\to\infty} \left\| \sup_i \sup_{f,g\in T_i} |\mathbb{G}_n f - \mathbb{G}_n g| \right\|_{P,2} = 0.$$

Since this is a bound for the $L_2(P)$ norm, if we assume the design effects vs iid sampling (superpopulation simple random sampling with replacement) for means of $\mathcal{F}_\delta$ are uniformly bounded, we have the same limit

$$\limsup_{n\to\infty} E^* \left[\left(\sup_i \sup_{f,g\in T_i} |\mathbb{G}_n^\pi f - \mathbb{G}_n^\pi g| \right)^2 \right] = 0$$

where the expectation is taken over superpopulation and population sampling, giving

$$\limsup_{n\to\infty} P^* \left[\sup_i \sup_{f,g\in T_i} |\mathbb{G}_n^\pi f - \mathbb{G}_n^\pi g| > \epsilon \right] < \eta.$$

and together with assumption D of a marginal CLT this implies weak convergence of $\mathbb{G}_n^\pi$ to a continuous mean-zero Gaussian process (Thm 1.5.6, van der Vaart & Wellner). □

Remark: Assumption P could clearly be relaxed to require just $2 + \delta$ moments for the envelope, but for this result to be useful we also need the Horvitz–Thompson variance estimator to be consistent for the variance conditional on the sequence of populations, and typical consistency proofs require four moments. For example, Fuller [9, chapter 1] show consistency of the variance estimator together with the central limit theorems.

3. Assessing the design-effect assumption

3.1. Examples of sufficient conditions

The simplest sufficient condition for uniformly bounded design effects in multistage stratified cluster sampling is that the cluster size is bounded. If $\{m_k\}_{k=k}^K$ is the number of sampled observations in the kth stage-one cluster, the design effect is bounded by

$$\Delta_{\max} = \frac{\sum_{k=1}^K m_k^2}{\sum_{k=1}^K m_k} \leq \max_k m_k.$$

Examples of surveys where this bound is small include the Scottish Household Survey ($\max_k m_k = 11$)[11], and the New Zealand Survey of Family Income and Employment, which has m_k roughly constant and averaging 6.67.

For surveys with large clusters, such as those conducted by the US National Center for Health Statistics [19, 20], the design-effect bound requires conditions on the correlation within clusters. Large clusters can still lead to small design effects if the correlation within clusters is weak. First consider a univariate scenario. Suppose the superpopulation distribution is multivariate Gaussian, generated by a random effects model, so that $X_{hi_1i_2} = \alpha_h + b_{i_1} + e_{i_1i_2}$ is the value of the X in the population for unit i_2 in cluster i_1 in stratum h, $\{\alpha_h\}_{h=1}^H$ are vectors of constants, and b_{i_1} and e_{i_1,i_2} are iid Normals with mean zero and variances σ_b^2 and σ_e^2 respectively. Further suppose that the sampling design is stratified random sampling of clusters, so that X is independent of sampling conditional on stratum.

Within each stratum, the intra-cluster correlation of X is $\rho = \sigma_b^2/(\sigma_b^2 + \sigma_e^2)$. As X is multivariate Normal, this is also the maximal correlation coefficient[16], that is, for any square-integrable function f, $|\text{corr}[f(X_{hi_1,i_2}), f(X_{hi_1,j_2})]| \leq \rho$, and the design effect for $f(X)$ is bounded above by

$$\Delta_{\max} = 1 + \rho(\max_k m_k - 1).$$

If the correlations between values of X in the superpopulation are non-negative, a tighter bound is that the design effect for the mean of $f(X)$ is at most equal to the design effect for the mean of X. When X is not multivariate Normal in the superpopulation, the maximal correlation coefficient will typically be greater than ρ. In particular, when $\mathcal{F}$ is the set of indicators of half-lines that characterize the cumulative distribution function, the design-effect bound could fail for continuous variables with sufficiently strong tail dependency.

Alternatively, suppose the clusters are contiguous spatial regions, as is typically the case, and the correlation can be modelled as due to a latent spatial random field. That is, $X_i = f_i(U_i, Z(s(i)))$ where U_i are iid $U[0,1]$, $s(i)$ is the spatial location of individual i, and $Z(s)$ is a spatial field.

If $Z(s)$ is a Markov random field satisfying Dobrushin's unicity condition, it is exponentially ϕ-mixing[eg 10, Theorem 2.1.3], and therefore exponentially ρ-mixing: that is, the maximal correlation coefficient in any function of X falls off exponentially with distance. The design effect is bounded above by

$$\Delta_{\max} = \max_k \frac{1}{m_k} \sum_{i,j=1}^{m_k} \rho_{ij} = \max_k \frac{1}{m_k} \sum_{i,j} e^{-c \cdot d(i,j)}$$

and if the spatial density of observations is uniform this has a finite limit as $m_k \to \infty$.

3.2. Comparison to β-mixing conditions

Doukhan et al. [8] proved a Donsker theorem under β-mixing assumptions with bracketing in the so-called $2,\beta$-norm based on the β-mixing coefficients. A key result is their Proposition 1: for any $f \in L_{2,\beta}(P)$

$$\sum_{i=1}^{\infty} |\mathrm{cov}[f(X_1), f(X_i)]| \le 4\|f\|^2_{P,2,\beta}.$$

This implies

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i) \right\|_2^2 \le \|f\|^2_{P,2} + 4\|f\|^2_{P,2,\beta} \le 5\|f\|^2_{P,2,\beta}.$$

which is (up to constants) weaker than a multiplicative design-effect bound, since for $\Delta > 1$

$$\mathrm{var}\left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i) \right] \le \Delta \mathrm{var}[f(X_1)]$$

implies

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i) \right\|_2^2 \le \Delta\|f\|^2_{P,2} \le |f\|^2_{P,2} + (\Delta - 1)\|f\|^2_{P,2,\beta}.$$

3.3. Design effects for $\mathcal{F}_\delta$ in data

The primary need for the Donsker theorem in our research is for the multivariate cumulative distribution function, so that $\mathcal{F}$ is the class of indicator functions of half-lines, quadrants, octants, etc; and $\mathcal{F}_\delta$ is the class of differences of nearby half-lines, quadrants, octants. As a demonstration of the plausibility of the bound on design effects we estimated the design effect on a grid of functions in $\mathcal{F}_\delta$ in examples of survey data.

We used two teaching samples distributed by UCLA Academic Technology Services. The samples come from a population of standardized test results and related socioeconomic variables for California schools: a one-stage cluster sample and and a two-stage sample, both using school districts as primary sampling units. Since the socioeconomic variables differ strongly between school districts, the design effects for means are larger than is typical. Even so, the design effects for the functions in $\mathcal{F}_\delta$ are moderate.

Table 1 shows the design effect for the means of four variables in the California schools data, and the maximum design effect for means of $I(x < c_1) - I(x < c_2)$ where c_1, c_2 are from a grid of approximately 30 equally spaced values. The four variables are the standardised Academic Performance Index test results in 1999 and 2000, and the proportion of students who are 'English

TABLE 1
Design effects for the mean (DEFF) and the maximum design effect over a grid differences of adjacent indicator functions, for four variables in two samples of California schools.

	1999 API		2000 API		ELL		meals	
	DEFF	max	DEFF	max	DEFF	max	DEFF	max
cluster sample	8.7	3.29	9.3	2.15	2.7	2.2	10.5	2.5
two-stage sample	6.0	3.7	6.3	4.0	9.8	5.6	11.9	3.2

TABLE 2
Design effects for the mean (DEFF) and the maximum design effect over a grid differences of adjacent bivariate indicator functions, for four variables in two samples of California schools.

	1999 API	2000 API		ELL		meals	
	DEFF	DEFF	max	DEFF	max	DEFF	max
cluster sample	8.7	9.3	2.6	2.7	2.3	10.5	5.2
two-stage sample	6.0	6.3	5.8	9.8	5.8	11.9	13.7

language learners' (ELL) and who receive subsidised school meals. In all cases, the maximum design effect for the difference of indicators is less than for the mean of the raw variables.

Table 2 shows similar computations for a bivariate cumulative distribution function: the differences of indicator functions are now $I(x < c_1 \cap y < d_1) - I(x < c_2 \cap y < d_2)$, and the table uses 1999 API as x and each of the other three variables as y. The grid has approximately 15 points for each margin. In all cases but one, the maximum design effect for the difference of indicators is less than the larger of the two design effects for the raw variables.

These results confirm that, at least in this example, the design effects needed for the uniform central limit theorem are comparable to the familiar design effects for means of the raw variables.

4. Summary

Under an assumption that is mathematically strong, but plausible in practice and possible to assess in data, we have shown that data from complex survey samples satisfies a uniform central limit theorem for much the same classes of functions as in independent data. It would be technically interesting to develop more elementary sufficient conditions, and to extend the result from superpopulation sampling to strict finite-population sampling, but two-phase or superpopulation sampling covers the problems where empirical process methods are of the most importance. In addition to generalizing arguments that already rely on empirical process theory, this result may provide an alternative approach to problems such as estimators based on estimating functions not differentiable in their parameters[29].

References

[1] Donald W. K. Andrews and David Pollard. An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review*, 61:119–132, 1994.

[2] F. Jay Breidt and Jean D. Opsomer. Local polynomial regresssion estimators in survey sampling. *Annals of Statistics*, 28:1026–1053, 2000.

[3] Norman E Breslow and Jon A Wellner. Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scand J Statist*, 34:86–102, 2007.

[4] Hervé Cardot and Etienne Josserand. Horvitz–Thompson estimators for functional data: asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika*, 98: 107–118, 2011.

[5] William G. Cochran. *Sampling Techniques.* John Wiley and Sons, 3rd edition, 1977.

[6] Herold Dehling, Thomas Mikosch, and Michael Sørensen, editors. *Empirical process techniques for dependent data*. Birkhäuser, Boston, 2002.

[7] Jean-Claude Deville and Carl-Erik Särndal. Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87:376–382, 1992.

[8] P. Doukhan, P. Massart, and E. Rio. Invariance principles for absolutely regular empirical processes. *Annales de l'Institute Henri Poincare, Section B*, 31(2):393–427, 1995.

[9] Wayne A. Fuller. *Sampling Statistics*. John Wiley and Sons, Hoboken, NJ, 2009.

[10] Xavier Guyon. *Random Fields on a Network: Modeling, Statistics, and Applications*. Springer-Verlag, 1995.

[11] Steven Hope. *Scotland's People: Scottish Household Survey Methodology 2005/2006*. A Scottish Executive National Statistics Publication, 2007.

[12] Daniel G. Horvitz and Donovan J. Thompson. A generalization of sampling without replacement from a finite universe. *J. Amer. Statist. Assoc.*, 47:663–685, 1952.

[13] Jeankyung Kim and David Pollard. Cube root asymptotics. *Annals of Statistics*, 18(191-219), 1990.

[14] Leslie Kish. *Survey Sampling*. John Wiley and Sons, 1965.

[15] Leslie Kish. Design effect. In S. Kotz and N. L. Johnson, editors, *Encyclopedia of Statistical Sciences*, volume 2, pages 347–348. Wiley, New York, 1982.

[16] H. O. Lancaster. Some properties of the bivariate normal distribution considered in the form of a contingency table. *Biometrika*, 44:289–292, 1957.

[17] DY Lin. On fitting Cox's proportional hazards models to survey data. *Biometrika*, 87(1): 37–47, 2000.

[18] Thomas Lumley and Alastair J. Scott. Two-sample rank tests under complex sampling. Technical report, Department of Statistics, University of Auckland, 2012.

[19] National Center for Health Statistics. *Plan and operation of the second National Health and Nutrition Examination Survey, 1976–1980.* Number 15 in Series 1, Programs and collection procedures. National Center for Health Statistics, 1981.

[20] National Center for Health Statistics. *Plan and Operation of the Third National Health and Nutrition Examination Survey, 1976–1980.* Number 32 in Series 1, Programs and collection procedures. National Center for Health Statistics, 1994.

[21] Mina Ossiander. A central limit theorem under metric entropy with L_2 bracketing. *Annals of Probability*, 15:897–919, 1987.

[22] David Pollard. A central limit theorem for empirical processes. *Journal of the Australian Mathematical Society (Series A)*, 33:235–248, 1981.

[23] Jens Praestgaard and Jon A. Wellner. Exchangeably weighted bootstraps of the general empirical process. *Annals of Probability*, 21:2053–2086, 1993.

[24] Takumi Saegusa and Jon A. Wellner. Weighted likelihood estimation under two-phase sampling. arXiv 1112.4951, 2012.

[25] Galen R. Shorack. *Probability for Statisticians*. Springer-Verlag, New York, 2000.

[26] Michel Talagrand. Isoperimetry and integrability of the sum of independent Banach-space valued random variables. *Annals of Probability*, 17:1546–1570, 1989.

[27] Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer, 1996.

[28] Jianqiang C. Wang. Sample distribution function based goodness-of-fit test for complex surveys. *Computational Statistics and Data Analysis*, 56(3):664–679, 2012.

[29] Jianqiang C. Wang and Jean D. Opsomer. On the asymptotic normality and variance estimation of nondifferentiable survey estimators. *Biometrika*, 98:91–106, 2011.