

Empirical processes in survey sampling

Supplementary Materials

Patrice Bertail^{a,b}, Emilie Chautruc, and Stéphane Cléménçon^d

^aMODAL'X - Université Paris Ouest

^bLaboratoire de Statistique - CREST

^cLaboratoire AGM - Université de Cergy-Pontoise

^dInstitut Télécom - LTCI UMR Télécom ParisTech/CNRS No. 5141

S1 Appendix - Some applications to nonparametric statistics

For illustration purposes, we consider here several statistical applications of the asymptotic results established in the main paper.

S1.1 Hadamard differentiable functionals

We first highlight that the FCLT stated above permits to establish the asymptotic normality of any statistic that can be expressed as the empirical version of some Hadamard differentiable functional, see Shorack and Wellner (1986). For the sake of clarity, we recall the definition of *uniform Hadamard differentiability* in Definition S1.1, adapted from Pons and de Turkheim (1991). Our results apply to many situations considered in their paper, related in particular to certain functionals of censored data. Other examples are treated in Gill (1989), van der Vaart and Wellner (1996) (see Chapter 3.9 p. 379 therein, in particular refer to

the discussion about the validity of the bootstrap for uniform Hadamard differentiable functionals). Define $B(\mathcal{F}, \mathbb{P})$ as the set of measures \mathbb{Q} in $\ell^\infty(\mathcal{F})$ whose paths $f \in \mathcal{F} \mapsto \mathbb{Q}f := \int f d\mathbb{Q}$ are $\|\cdot\|_{2, \mathbb{P}}$ -uniformly continuous and bounded. This is the smallest natural space containing \mathbb{G} . We consider the uniform Hadamard differentiability tangentially to the subspace $B(\mathcal{F}, \mathbb{P})$ because it weakens the notion of differentiation and is easier to check in practice.

Definition S1.1 A functional $T : \ell^\infty(\mathcal{F}) \rightarrow \mathbb{R}^q$ is said to be uniformly *Hadamard differentiable* at \mathbb{P} tangentially to $B(\mathcal{F}, \mathbb{P})$, if and only if there exists a continuous linear mapping $dT_{\mathbb{P}}$ such that for any sequence \mathbb{P}_N converging to \mathbb{P} , any \mathbf{h}_N converging to $\mathbf{h} \in B(\mathcal{F}, \mathbb{P})$ and every t_N converging to 0 such that $\mathbb{P}_N + t_N \cdot \mathbf{h}_N \in \ell^\infty(\mathcal{F})$, we have:

$$\frac{T(\mathbb{P}_N + t_N \cdot \mathbf{h}_N) - T(\mathbb{P}_N)}{t_N} - dT_{\mathbb{P}} \cdot \mathbf{h} \xrightarrow{t_N \rightarrow 0} 0.$$

Notice that T may be defined not on the entire space $\ell^\infty(\mathcal{F})$ but on a subset \mathcal{L} only. In this case, one must check that $\mathbb{P}_N + t_N \cdot \mathbf{h}_N \in \mathcal{L}$.

Remark S1.1 We may in addition assume that the differential $dT_{\mathbb{P}}$ admits an integral representation, *i.e.*

$$dT_{\mathbb{P}} \cdot \mathbf{h} = \int T^{(1)}(\mathbf{x}, \mathbb{P}) \mathbf{h}(d\mathbf{x}),$$

where $T^{(1)}(\cdot, \mathbb{P})$ is the influence function defined from \mathcal{X} to \mathcal{B}_1 such that we have

$$\mathbb{E}_{\mathbb{P}} \left(T^{(1)}(\mathbf{X}, \mathbb{P}) \right) = 0.$$

We recall that in the robustness terminology (Hampel et al., 1986), the influence function of the parameter $T(\mathbb{P})$ may be calculated directly by computing the derivative of the functional taken at the contaminated distribution $(1-t)\mathbb{P} + t\delta_{\mathbf{x}}$, *i.e.*

$$T^{(1)}(\mathbf{x}, \mathbb{P}) := \lim_{t \rightarrow 0} \frac{T((1-t)\mathbb{P} + t\delta_{\mathbf{x}}) - T(\mathbb{P})}{t}.$$

In this case, the limiting distribution may be calculated more easily.

Theorem S1.1 – CLT for Hadamard differentiable functionals. *Suppose that the assumptions of Theorem 4.2 hold and that functional $T : \mathcal{L} \subset \ell^\infty(\mathcal{F}) \rightarrow \mathbb{R}^q$ is Hadamard differentiable at \mathbb{P} with differential $dT_{\mathbb{P}}$ and influence function $T^{(1)}(\mathbf{x}, \mathbb{P})$. Then, as $N \rightarrow +\infty$, we have:*

$$\sqrt{N} \left(T(\mathbb{P}_{\mathbb{R}^N}^{\pi(\mathbb{R}^N)}) - T(\mathbb{P}_N) \right) \Rightarrow dT_{\mathbb{P}}.\mathbb{G},$$

where \mathbb{G} is a Gaussian process with covariance operator Σ , as in Eq. (4).

The result above, the proof of which is available in the Supplementary Materials, applies in particular to the following statistics.

Example S1.1 – Expectation and variance. It is well-known that the functionals $T(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}(\mathbf{X})$ and $T(\mathbb{P}) = \mathbb{V}_{\mathbb{P}}(\mathbf{X})$ are uniformly Hadamard-differentiable when considering appropriate classes of function \mathcal{F} . When $T(\mathbb{P}) = \mathbb{E}_{\mathbb{P}}(\mathbf{X})$, Theorem S1.1 exactly reduces to the Central Limit Theorem established in Hájek (1964).

Example S1.2 – Cumulative distribution function. In a univariate setting, the functional $T(\mathbb{P}) = F(x) := \mathbb{P}(X \in (-\infty, x])$ can be dealt with by simply considering the class of indicator functions $\mathbf{u} \mapsto \mathbb{I}\{\mathbf{u} \leq x\}$ with $x \in \mathbb{R}$ and applying next ?? and Corollary 4.1. We provide illustrations of this specific example in Section S1.3.

S1.2 Fréchet differentiable functionals

Hadamard differentiability is sometimes difficult to prove and it does not yield a precise control of the remainder for further approximations like Edgeworth expansions. Another approach followed by Dudley (1990) and Barbe and Bertail (1995) is to assume Fréchet differentiability with respect to a metric $d_{\mathcal{F}}$ indexed by a class of function \mathcal{F} , for which some uniform entropy conditions hold. A functional is said to be *Fréchet differentiable* at \mathbb{P} for such a metric if there exists a gradient (for instance the influence function $T^{(1)}(\mathbf{x}, \mathbb{P})$, which fulfills

$\mathbb{E}_{\mathbb{P}}(\mathsf{T}^{(1)}(\mathbf{x}, \mathbb{P})) = 0$) and a continuous function $\varepsilon(\cdot)$, null at 0, such that for any probability \mathbb{Q} ,

$$\mathsf{T}(\mathbb{Q}) - \mathsf{T}(\mathbb{P}) = \int \mathsf{T}^{(1)}(\mathbf{x}, \mathbb{P})(\mathbb{Q} - \mathbb{P})(d\mathbf{x}) + d_{\mathcal{F}}(\mathbb{Q}, \mathbb{P}) \varepsilon(d_{\mathcal{F}}(\mathbb{Q}, \mathbb{P})).$$

It is generally possible to choose the class of functions according to the functional of interest, see for instance Arcones and Giné (1992) for general classes of M-estimators. Notice that in that case, by applying Fréchet differentiability twice, we have

$$\begin{aligned} \sqrt{N} \left(\mathsf{T}(\mathbb{P}_{R_N}^{\pi(R_N)}) - \mathsf{T}(\mathbb{P}_N) \right) &= \sqrt{N} \int \mathsf{T}^{(1)}(\mathbf{x}, \mathbb{P})(\mathbb{P}_{R_N}^{\pi(R_N)} - \mathbb{P}_N)(d\mathbf{x}) + r_N \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\varepsilon_i}{\pi_i} \mathsf{T}^{(1)}(\mathbf{X}_i, \mathbb{P}) + r_N, \end{aligned}$$

with a remainder

$$r_N = \sqrt{N} d_{\mathcal{F}}(\mathbb{P}_{R_N}^{\pi(R_N)}, \mathbb{P}) \varepsilon(d_{\mathcal{F}}(\mathbb{P}_{R_N}^{\pi(R_N)}, \mathbb{P})) + \sqrt{N} d_{\mathcal{F}}(\mathbb{P}_N, \mathbb{P}) \varepsilon(d_{\mathcal{F}}(\mathbb{P}_N, \mathbb{P})).$$

By virtue of the results in Hájek (1964), it is then obvious that the linear term in this approximation is asymptotically Gaussian with known variance. Controlling the remainder essentially amounts to controlling the behavior of $\sqrt{N} d_{\mathcal{F}}(\mathbb{P}_{R_N}^{\pi(R_N)}, \mathbb{P})$ or alternatively, by the triangular inequality, that of $\sqrt{N} d_{\mathcal{F}}(\mathbb{P}_{R_N}^{\pi(R_N)}, \mathbb{P}_N)$, which was the purpose of Section 3 and Section 4.

S1.3 Simulation-based Gaussian asymptotic confidence regions

A straightforward application consists in the building of Gaussian confidence regions for the (univariate) empirical cumulative distribution function in the entire population, denoted by $F_N(x)$, $x \in \mathbb{R}$, when the survey scheme is of the rejective type. Indeed, consider the class of functions $\mathcal{F} := \{f_x(\cdot) := \mathbb{I}\{\cdot \leq x\}, x \in \mathbb{R}\}$. Provided Assumption 2.1 is fulfilled, it respects the required conditions for Corollary 4.1 to hold (see Van der Vaart, 2000, Example 19.16 for the uniform entropy condition and take $H(x) = 1$ and $\delta = 1$ when checking condition

i^*) in 4.2), which implies in particular that $\|\mathbb{G}_{R_N}^{\pi(R_N)}\|_{\mathcal{F}}$ converges in distribution to $\|\mathbb{G}\|_{\mathcal{F}}$ as $N \rightarrow +\infty$ (Van der Vaart, 2000, Corrolary 19.21). This yields the following asymptotic uniform confidence band of level $\alpha \in (0, 1)$ for the population cdf F_N :

$$CB_\alpha := \left[F_{R_N}^{\pi(R_N)} - \frac{q_\alpha}{\sqrt{N}}, F_{R_N}^{\pi(R_N)} + \frac{q_\alpha}{\sqrt{N}} \right],$$

where $F_{R_N}^{\pi(R_N)}$ is the Horvitz-Thompson estimator of the cdf based on the rejec-tive sample and q_α the α -quantile of random variable $\|\mathbb{G}\|_{\mathcal{F}}$. Since in practice q_α is unknown, it needs to be estimated. It can be achieved by means of Monte-Carlo simulations, using a simple technique based on the Cholesky decomposition of the covariance matrix (Kroese et al., 2011, Algorithm 5.1). In the next subsections, a set of numerical experiments is performed to provide illustrative examples of this technique.

S1.3.1 Experiment setting

Simulations were based on the following model, chosen for its simplicity in terms of both computation and interpretation:

$$\begin{aligned} X &= \beta W + U, \text{ with } \beta \in \{0, 1\}, W \rightsquigarrow \mathcal{TN}(\mu, \sigma_W^2, w_*, w^*), U \rightsquigarrow \mathcal{N}(0, \sigma_U^2), \\ \mathbb{P}(W \leq w, U \leq u) &= \mathbb{P}(W \leq w) \mathbb{P}(U \leq u), \end{aligned}$$

where X is the variable of interest, W the auxiliary information, U a white noise independent from W , and $\mathcal{TN}(0, \sigma_W^2, w_*, w^*)$ refers to the truncated Normal distribution over $[w_*, w^*]$, with expectation μ and variance σ_W^2 . Such a representation enables a simple control of the dependence between X and W , since their correlation is then

$$\text{corr}(X, W) = \beta \frac{\sigma_W}{\sqrt{\sigma_W^2 + \sigma_U^2}}.$$

For a given population \mathcal{U}_N of size N , where it is assumed that $\{W_i, i \in \mathcal{U}_N\}$ (resp. $\{U_i, i \in \mathcal{U}_N\}$) are independent (hence exchangeable) realizations of W

(resp. \mathbf{U}), inclusion probabilities of the Poisson sampling scheme are defined as

$$p_i = p(W_i) = n \frac{W_i}{\sum_{j=1}^N W_j}, \quad (\text{S1})$$

with n the desired expected sample size (Hájek, 1964, Section 6, p.1512). When the inclusion probabilities are proportionate to the auxiliary variable like in Eq. (S1), the stronger the correlation between X and W , the smaller the variance of the estimator of the population mean $\frac{1}{N} \sum_{i=1}^N X_i$ (or, equivalently, of the total $\sum_{i=1}^N X_i$). Recall that under Assumption 2.1, we have $n/N \rightarrow c \in (0, 1)$ as both n and N tend to infinity. Hence, p_i can be viewed as the empirical version in the population of

$$p(W) := W \frac{c}{\mathbb{E}(W)}.$$

Observe that thus defined, $p(W) \in [p_*, p^*]$, where $p_* = c w_*/\mu$ and $p^* = c w^*/\mu$, which offers an easy way of ensuring Assumption 2.1 is fulfilled.

Numerical experiments were conducted on a set of populations with increasing sizes $N = 10^2, 5 \times 10^2, 10^3, 5 \times 10^3$ and 10^4 . Though the latter may seem quite small to study asymptotic properties, they are in fact representative of many practical situations, where populations under the microscope have moderate sizes in comparison to nationwide surveys. Several scenarios were investigated depending on both the variance parameter σ_U^2 and the coefficient β , so as to cover situations where $\text{corr}(X, W)$ is high, low or null. They are summarized in Table S1. For each scenario, two sample sizes were considered: one small with $n = 0.1 \times N$ and one relatively large with $n = 0.5 \times N$. Parameters of the distribution of W were chosen to ensure that for all $i \in \mathcal{U}_N$, $p_i \in [0.01, 1]$. Specifically, we set $\mu = 1$, $\sigma_W^2 = 0.09$, $w_* = 0.1$ and $w^* = 2$, thereby implying that $(p_*, p^*) = (0.01, 0.02)$ when $n = 0.1 \times N$ and $(p_*, p^*) = (0.05, 1)$ when $n = 0.5 \times N$.

Table S1 – *List of scenarios depending on β and $\sigma_{\mathbb{U}}^2$, and corresponding model characteristics*

Scenario	β	$\sigma_{\mathbb{U}}^2$	$\text{corr}(X, W)$
S ₁	1	0.01	0.95
S ₂	1	35.91	0.05
S ₃	0	35.91	0

For each scenario, we drew 1000 samples according to a rejective sampling scheme, following Algorithm 5.9 in Tillé (2006). The true inclusion probabilities, denoted by π_i , $1 \leq i \leq N$, were deduced from their Poisson equivalents defined in Eq. (S1) by repeating 10^5 times the basic algorithm stated in Example 2.4.

We constructed asymptotic uniform 95% confidence bands of the population cdf F_N using 10^3 Monte-Carlo approximations as in Kroese et al. (2011, Algorithm 5.1).

S1.3.2 Experiment results

The average and maximal width of the confidence bands over the 1000 simulated samples for each scenario are given in Table S2. Coverage probabilities were also estimated, the results of which are displayed in Table S3. Finally, some graphical illustrations are provided in Figure S1.

As expected, the larger N and c , the smaller the confidence bands. Regarding coverage probabilities, they appear to be close to 95%, the desired level, for any N and c . The most remarkable variability is that observed between scenarios: confidence bands get significantly tighter as the correlation between X and W decreases. As a consequence, estimated coverage probabilities are systematically smaller in scenarios S₂ and S₃ than in scenario S₁, especially when $N = 10^2$ and $c = 0.1$.

Table S2 – Average (Av) and maximal (Mx) width of confidence bands (in %)

		N									
		10^2		5×10^2		10^3		5×10^3		10^4	
	c	Av	Mx	Av	Mx	Av	Mx	Av	Mx	Av	Mx
S_1	0.5	31.24	37.12	15.68	20.70	11.43	14.61	5.10	5.76	3.58	3.94
	0.1	87.68	116.62	42.67	68.80	32.67	45.64	14.46	18.15	10.19	12.33
S_2	0.5	27.74	33.55	13.77	16.52	9.95	12.88	4.48	5.05	3.17	3.56
	0.1	76.43	103.01	37.63	53.10	28.40	40.36	12.70	15.95	8.99	10.81
S_3	0.5	27.53	32.79	13.67	16.12	9.85	13.36	4.43	5.13	3.14	3.41
	0.1	75.65	105.18	37.20	53.93	27.93	38.51	12.57	15.92	8.90	10.61

Table S3 – Estimated coverage probabilities (in %)

Scenario	c	N				
		10^2	5×10^2	10^3	5×10^3	10^4
S_1	0.5	96.30	96.80	96.80	97.80	97.00
	0.1	94.74	97.20	97.12	98.30	98.50
S_2	0.5	92.50	97.30	95.60	96.40	96.40
	0.1	89.12	93.80	93.55	95.60	96.90
S_3	0.5	91.80	96.50	95.50	96.10	95.60
	0.1	87.82	93.50	93.48	94.80	96.70

This phenomenon is due to the formula used to construct inclusion probabilities, in Eq. (S1). Let us dwell for a moment on this expression. It ensures that the Horvitz-Thompson estimator (based on the Poisson inclusion probabilities) of the expectation of W coincides with the classical empirical mean in the entire population:

$$\frac{1}{N} \sum_{i=1}^N \frac{\epsilon_i}{p_i} W_i = \frac{1}{N} \sum_{i=1}^N \frac{\epsilon_i}{n \frac{W_i}{\sum_{j=1}^N W_j}} W_i = \frac{1}{N} \sum_{i=1}^N \frac{\epsilon_i}{n} \sum_{j=1}^N W_j = \frac{1}{N} \sum_{j=1}^N W_j,$$

since $\sum_{i=1}^N \epsilon_i = n$ by definition. It is no surprise then that the stronger the correlation between X and W , the closer (in terms of variance) the weighted mean $\frac{1}{N} \sum_{i=1}^N \frac{\epsilon_i}{p_i} X_i$ is to its population counterpart. However, when considering empirical distribution functions, the standard and sample estimators for W are no longer equal. Hence, not only does the model in Eq. (S1) fail to improve the variance of the HT-cdf of X , but the deviations of $F_{R_N}^P$ are expected to grow as the link between X and W tightens. To counterbalance this drawback, we could for instance choose the inclusion probabilities p_i , $1 \leq i \leq N$, that minimize the uniform difference between the HT and the empirical cdf of W (see for instance Rueda et al., 2007). Such refinements are left for further research. Although not optimal, the confidence bands constructed on our numerical experiments are still satisfactory and advocate the utility of our asymptotic results whatever the available inclusion probabilities.

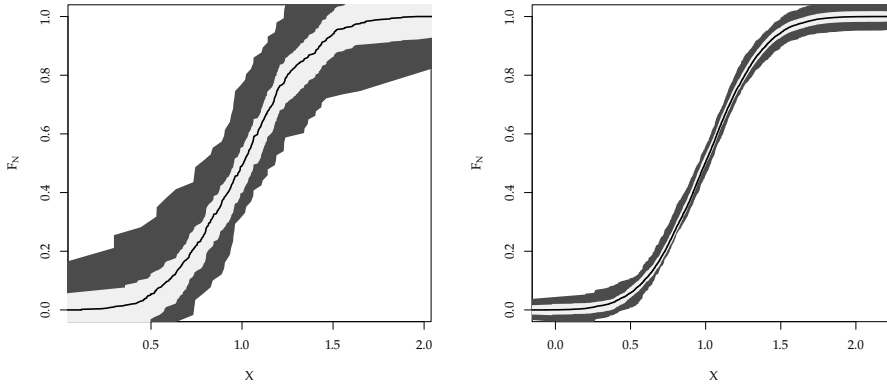


Figure S1 – Example of the 95% confidence bands of the empirical distribution function in the population F_N (black line) constructed on one of the 1000 simulated samples under scenario S_1 with $c = 0.1$ (dark area) and $c = 0.5$ (light area) for $N = 5 \times 10^2$ (left hand plot) and $N = 10^4$ (right hand plot)

S2 Appendix - Technical proofs

S2.1 Proof of Lemma 4.1

Notice first that under Assumption 4.1, we have: $\forall (f, g) \in \mathcal{F}^2$,

$$\begin{aligned} \text{cov}_{N, \mathbf{p}}(f, g) &= \frac{1}{N} \sum_{i=1}^N \frac{f(\mathbf{X}_i)}{\mathbf{p}(\mathbf{W}_i)} \frac{g(\mathbf{X}_i)}{\mathbf{p}(\mathbf{W}_i)} \mathbf{p}(\mathbf{W}_i) (1 - \mathbf{p}(\mathbf{W}_i)) \\ &\quad - \theta_{N, \mathbf{p}}(f) \theta_{N, \mathbf{p}}(g) \frac{1}{N} \sum_{i=1}^N \mathbf{p}(\mathbf{W}_i) (1 - \mathbf{p}(\mathbf{W}_i)), \end{aligned}$$

with

$$\theta_{N, \mathbf{p}}(f) = \frac{\sum_{i=1}^N (1 - \mathbf{p}(\mathbf{W}_i)) f(\mathbf{X}_i)}{\sum_{i=1}^N (1 - \mathbf{p}(\mathbf{W}_i)) \mathbf{p}(\mathbf{W}_i)}.$$

Now it is sufficient to apply the Strong Law of Large Numbers for exchangeable vectors to obtain that

$$\begin{aligned} \frac{1}{N} d_N &= \frac{1}{N} \sum_{i=1}^N (1 - \mathbf{p}(\mathbf{W}_i)) \mathbf{p}(\mathbf{W}_i) \\ &\xrightarrow{N \rightarrow \infty} \int_{\mathcal{W}} (1 - \mathbf{p}(\mathbf{w})) \mathbf{p}(\mathbf{w}) \mathbb{P}_{\mathbf{W}}(d\mathbf{w}) \text{ almost-surely.} \end{aligned}$$

The limit above is finite, positive under Assumption 2.1 (that implies there exists $\mathbf{p}_* > 0$ such that $\mathbf{p}(\mathbf{w}) > \mathbf{p}_*$). Additionally, we have with probability one

$$\begin{aligned} \frac{1}{N} \sum_{i=1}^N \frac{f(\mathbf{X}_i)}{\mathbf{p}(\mathbf{W}_i)} \frac{g(\mathbf{X}_i)}{\mathbf{p}(\mathbf{W}_i)} \mathbf{p}(\mathbf{W}_i) (1 - \mathbf{p}(\mathbf{W}_i)) &\xrightarrow{N \rightarrow \infty} \\ &\int_{\mathcal{X} \times \mathcal{W}} f(\mathbf{x}) g(\mathbf{x}) \left(\frac{1}{\mathbf{p}(\mathbf{w})} - 1 \right) \mathbb{P}_{\mathbf{X}, \mathbf{W}}(d\mathbf{x}, d\mathbf{w}). \end{aligned}$$

By virtue of Assumption 2.2, the latter integral is finite. Finally, observe that we almost-surely have

$$\theta_{N, \mathbf{p}}(f) \xrightarrow{N \rightarrow \infty} \theta_{\mathbf{p}}(f),$$

and the desired result follows. In particular notice that the limiting variance $V^2(f)$ is given by

$$\begin{aligned} V^2(f) &:= \int_{\mathcal{X} \times \mathcal{W}} f(\mathbf{x})^2 \left(\frac{1}{\mathbf{p}(\mathbf{w})} - 1 \right) \mathbb{P}_{\mathbf{X}, \mathbf{W}}(d\mathbf{x}, d\mathbf{w}) \\ &\quad - \theta_{\mathbf{p}}(f)^2 \int_{\mathcal{W}} (1 - \mathbf{p}(\mathbf{w})) \mathbf{p}(\mathbf{w}) \mathbb{P}_{\mathbf{W}}(d\mathbf{w}), \end{aligned}$$

which is strictly positive except in the degenerate case where $f(\mathbf{x}) = \mathbf{p}(\mathbf{w})$. Typically, this occurs when the inclusion probabilities are based directly on the variable of interest (or $\mathbf{W} = \mathbf{c} \times \mathbf{X}$ for some $\mathbf{c} \in \mathbb{R}$). Positivity of the operator results from Cauchy-Schwarz inequality.

S2.2 Proof of Theorem 4.2

We essentially have to check hypotheses i) – iv) of Theorem 4.1.

Concerning hypothesis i), the Lindeberg-Feller condition can be written as

$$\frac{1}{N} \sum_{i=1}^N \mathbb{E} \left(Z_{N,i}^2 \mathbb{I} \left\{ Z_{N,i} > \eta \sqrt{N} \right\} \right) \xrightarrow{N \rightarrow \infty} 0 \quad \text{for every } \eta > 0,$$

which reduces to $\mathbb{E} \left(Z_{N,i}^2 \mathbb{I} \left\{ Z_{N,i} > \eta \sqrt{N} \right\} \right) \xrightarrow{N \rightarrow \infty} 0$ by exchangeability of the components. This corresponds to condition i) in Theorem 4.2 above.

Recall that Assumption 2.2 stipulates the envelope of class \mathcal{F} is square-integrable function H and that under Assumption 2.1, there is some $p_* > 0$ such that for all $i \in \mathcal{U}_N$, $p(\mathbf{W}_i) \geq p_*$. Hence, we have

$$\sum_{i=1}^N p(\mathbf{W}_i) (1 - p(\mathbf{W}_i)) \geq p_* (N - \mathbb{E}(\mathbf{n})) = p_* N \left(1 - \frac{\mathbb{E}(\mathbf{n})}{N} \right).$$

as well as

$$|\theta_{N,\mathbf{p}}(f)| \leq \theta_{N,\mathbf{p}}(H) \leq \frac{1}{p_*} \frac{1}{N - \mathbb{E}(\mathbf{n})} \sum_{i=1}^N H(\mathbf{X}_i) < \infty.$$

We thus obtain:

$$\begin{aligned} \sup_{f \in \mathcal{F}} \left| \frac{f(\mathbf{X}_i)}{p(\mathbf{W}_i)} - \theta_{N,\mathbf{p}}(f) \right|^2 &\leq 2 \left(\sup_{f \in \mathcal{F}} \left| \frac{f(\mathbf{X}_i)}{p(\mathbf{W}_i)} \right|^2 + \sup_{f \in \mathcal{F}} |\theta_{N,\mathbf{p}}(f)|^2 \right) \\ &\leq 2 \left(\left| \frac{H(\mathbf{X}_i)}{p(\mathbf{W}_i)} \right|^2 + |\theta_{N,\mathbf{p}}(H)|^2 \right). \end{aligned}$$

Set

$$\begin{aligned} \mathcal{G}_{1,i} &= |\epsilon_i - p(\mathbf{W}_i)| \left(\left| \frac{H(\mathbf{X}_i)}{p(\mathbf{W}_i)} \right| + |\theta_{N,\mathbf{p}}(H)| \right), \\ \mathcal{G}_{2,i} &= (\epsilon_i - p(\mathbf{W}_i))^2 \left(\left| \frac{H(\mathbf{X}_i)}{p(\mathbf{W}_i)} \right|^2 + |\theta_{N,\mathbf{p}}(H)|^2 \right). \end{aligned}$$

Observe that it is thus sufficient to check that for every $\eta > 0$,

$$\mathbb{E}_{\mathbb{P}_{\mathbf{x}, \mathbf{w}}} \left(\mathbb{E}_{T_N} \left(\mathcal{G}_{2,i} \mathbb{I} \left\{ \mathcal{G}_{1,i} > \eta \sqrt{N} \right\} \mid (\mathbf{X}_i, \mathbf{W}_i)_{1 \leq i \leq N} \right) \right) \xrightarrow{N \rightarrow \infty} 0.$$

Condition ii) can be checked immediately by noticing that, in the case of the Poisson process, the equicontinuity condition becomes

$$\begin{aligned} \sup_{\rho_{\mathbb{P}}(f, g) < \delta} \sum_{i=1}^N \mathbb{E} \left((Z_{N,i}(f) - Z_{N,i}(g))^2 \right) &= \\ \sup_{\rho_{\mathbb{P}}(f, g) < \delta} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{\mathbb{P}_{\mathbf{x}, \mathbf{w}}} \left(\left(\frac{f(\mathbf{X}_i) - g(\mathbf{X}_i)}{p(\mathbf{W}_i)} - \theta_{N,p}(f) + \theta_{N,p}(g) \right)^2 \right. \\ &\quad \left. \times p(\mathbf{W}_i)(1 - p(\mathbf{W}_i)) \right) \\ &\leq \sup_{\rho_{\mathbb{P}}(f, g) < \delta} \frac{C_{p_*}}{4p_*^2} \mathbb{E}_{\mathbb{P}_{\mathbf{x}, \mathbf{w}}} \left((f(\mathbf{X}_1) - g(\mathbf{X}_1))^2 \right) \rightarrow 0, \text{ as } \delta \rightarrow 0, \end{aligned}$$

where C_{p_*} is a positive constant depending on p_* . In practice, condition iii) is checked in an easier manner by using the uniform entropy condition given here, see also Lemma 2.11.6 in van der Vaart and Wellner (1996).

Finally, condition iv) is a direct consequence of Lemma 4.1.

S2.3 Proof of Theorem 4.3

As explained in Section 3.2, with a fixed sample size n in the canonical case where $\sum_{i=1}^N p_i = n$, the recentered process $\tilde{\mathbb{G}}_{R_N}^{\mathbf{p}}$ coincides with the initial process $\mathbb{G}_{R_N}^{\mathbf{p}}$ and the distribution of $\tilde{\mathbb{G}}_{R_N}^{\mathbf{p}}$ is the same as that of $\tilde{\mathbb{G}}_{T_N}^{\mathbf{p}}$ conditional on the sample size being fixed, equal to n . Therefore, it suffices to study the joint distribution of $\left((\tilde{\mathbb{G}}_{T_N}^{\mathbf{p}} f)_{f \in \mathcal{F}}, \frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_i - p_i \right)$, which by virtue of Theorem 4.2 is marginally asymptotically Gaussian. There, checking the condition of the multivariate Lindeberg-Feller Theorem reduces to checking condition i) of Theorem 4.2. Therefore, under our set of hypotheses, the limiting process is jointly Gaussian. In addition, we have $\text{cov}_{T_N} \left(\tilde{\mathbb{G}}_{T_N}^{\mathbf{p}} f, \frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_i - p_i \right) = 0$ for all $f \in \mathcal{F}$, thus they are asymptotically independent and the limiting distribution of $\mathbb{G}_{R_N}^{\mathbf{p}}$ reduces to that of $\tilde{\mathbb{G}}_{T_N}^{\mathbf{p}}$.

S2.4 Proof of Corollary 4.1

Following in the footsteps of Hájek (1964), in the rejective sampling situation where $p(\mathbf{W}_i) = p_i$ for $i \in \{1, \dots, N\}$, we have

$$\max_{1 \leq i \leq N} \left| \frac{p_i}{\pi_i^R} - 1 \right| \xrightarrow{N \rightarrow \infty} 0.$$

We thus have

$$\max_{1 \leq i \leq N} \left| \frac{\pi_i^R}{p_i} - 1 \right| \xrightarrow{N \rightarrow \infty} 0$$

under the hypothesis that Assumption 2.1 is fulfilled by the p_i 's. Then, we can write

$$\begin{aligned} \mathbb{G}_{R_N}^{p(\mathbf{W})} f - \mathbb{G}_{R_N}^{\pi(R_N)} f &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{\epsilon_i}{p(\mathbf{W}_i)} - \frac{\epsilon_i}{\pi_i^R} \right) f(\mathbf{X}_i) \\ &= \frac{1}{\sqrt{N}} \sum_{i=1}^N \frac{\epsilon_i}{p(\mathbf{W}_i)} \left(1 - \frac{p_i}{\pi_i^R} \right) f(\mathbf{X}_i), \end{aligned}$$

and

$$\begin{aligned} &\sup_{\mathbf{b} \in \text{BL}_1(\ell^\infty(\mathcal{F}))} \mathbb{E} \left(\left| \mathbf{b} \left(\sup_{f \in \mathcal{F}} \mathbb{G}_{R_N}^{p(\mathbf{W})} f \right) - \mathbf{b} \left(\sup_{f \in \mathcal{F}} \mathbb{G}_{R_N}^{\pi(R_N)} f \right) \right| \right) \\ &\leq \mathbb{E} \left(\left| \sup_{f \in \mathcal{F}} \mathbb{G}_{R_N}^{p(\mathbf{W})} f - \sup_{f \in \mathcal{F}} \mathbb{G}_{R_N}^{\pi(R_N)} f \right| \right) \\ &\leq \mathbb{E} \left(\sup_{f \in \mathcal{F}} \left| \frac{1}{\sqrt{N}} \sum_{i=1}^N \left(\frac{\epsilon_i}{p(\mathbf{W}_i)} - \frac{\epsilon_i}{\pi_i^R} \right) f(\mathbf{X}_i) \right| \right) \\ &\leq \frac{1}{\sqrt{N}} \sum_{i=1}^N \left| \frac{\pi_i^R}{p(\mathbf{W}_i)} - 1 \right| H(\mathbf{X}_i), \end{aligned}$$

which quantity vanishes asymptotically under Assumption 2.1, according to Theorem 5.1, Equations (5.7) and (5.26) in Hájek (1964, p. 1508-1510).

The desired convergence is finally established by combining this result with Theorem 4.3 and the functional version of Slutsky's theorem (see Theorem 3.4 in Resnick, 2007 for instance).

S2.5 Proof of Lemma 4.2

Let $\mathbf{b} \in \text{BL}_1(\ell^\infty(\mathcal{F}))$. We have

$$\begin{aligned} \mathbb{E}_{\mathbf{R}_N} \left(\mathbf{b} \left(\mathbb{G}_{\tilde{\mathbf{R}}_N}^{\pi(\mathbf{R}_N)} \right) \right) - \mathbb{E}_{\tilde{\mathbf{R}}_N} \left(\mathbf{b} \left(\mathbb{G}_{\mathbf{R}_N}^{\pi(\mathbf{R}_N)} \right) \right) &= \sum_{s \in \mathcal{P}(\mathcal{U}_N)} \tilde{\mathbf{R}}_N(s) \mathbf{b} \left(\mathbb{G}_{\mathbf{R}_N(s)}^{\pi(\mathbf{R}_N)} \right) \\ &\quad - \sum_{s \in \mathcal{P}(\mathcal{U}_N)} \mathbf{R}_N(s) \mathbf{b} \left(\mathbb{G}_{\tilde{\mathbf{R}}_N(s)}^{\pi(\mathbf{R}_N)} \right) \\ &\leq \sum_{s \in \mathcal{P}(\mathcal{U}_N)} \left| \mathbf{R}_N(s) - \tilde{\mathbf{R}}_N(s) \right| \end{aligned}$$

because \mathbf{b} is bounded by 1 and for a fixed $s \in \mathcal{P}(\mathcal{U}_N)$,

$$\mathbf{b} \left(\mathbb{G}_{\mathbf{R}_N(s)}^{\pi(\mathbf{R}_N)} \right) = \mathbf{b} \left(\mathbb{G}_{\tilde{\mathbf{R}}_N(s)}^{\pi(\mathbf{R}_N)} \right),$$

their expressions depending on the first order inclusion probabilities $\pi(\mathbf{R}_N)$ solely. The last inequality follows from the usual inequality between the total variation metric and the entropy (Berger, 1998, Lemma 2 p.219).

S2.6 Proof of Theorem S1.1

The idea is essentially to apply the Hadamard differentiability property to the sequence $\mathbf{h}_N = \sqrt{N}(\mathbb{P}_{\mathbf{R}_N}^{\pi(\mathbf{R}_N)} - \mathbb{P}_N) =: \mathbb{G}_{\mathbf{R}_N}^{\pi(\mathbf{R}_N)}$, which converges to $\mathbf{h} = \mathbb{G}$ in $\ell^\infty(\mathcal{F})$ and $\mathbf{t}_N = \frac{1}{\sqrt{N}} \rightarrow 0$. We thus have, as $N \rightarrow +\infty$:

$$\sqrt{N} \left(\mathbb{T}(\mathbb{P}_{\mathbf{R}_N}^{\pi(\mathbf{R}_N)}) - \mathbb{T}(\mathbb{P}_N) \right) = \sqrt{N} \left(\mathbb{T}(\mathbb{P}_N + \frac{1}{\sqrt{N}} \mathbf{h}_N) - \mathbb{T}(\mathbb{P}_N) \right) \longrightarrow \text{d}\mathbb{T}_{\mathbb{P}}.\mathbb{G}.$$

References

- M.A. Arcones and E. Giné. On the bootstrap of M-estimators and other statistical functionals. *Exploring the Limits of Bootstrap*, ed. by R. LePage and L. Billard, Wiley, pages 13–47, 1992.
- P. Barbe and P. Bertail. *The weighted bootstrap*, volume 98. Springer Verlag, 1995.
- Y.G. Berger. Rate of convergence to normal distribution for the Horvitz-Thompson estimator. *J. Stat. Plan. Inf.*, 67(2):209–226, 1998.

- R.M. Dudley. Nonlinear functionals of empirical measures and the bootstrap. In *Probability in Banach Spaces 7*, pages 63–82. Springer, 1990.
- R.D. Gill. Non- and semiparametric maximum likelihood estimators and the von Mises method. *Scand. J. Statistics*, 22:205–214, 1989.
- J. Hájek. Asymptotic theory of rejective sampling with varying probabilities from a finite population. *The Annals of Mathematical Statistics*, 35(4):1491–1523, 1964.
- F.R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel. *Robust statistics: the approach based on influence functions*. 1986.
- D.P. Kroese, T. Taimre, and Z.I. Botev. *Handbook of Monte Carlo methods*. Wiley, 2011.
- O. Pons and E. de Turkheim. Von mises method, bootstrap and Hadamard differentiability for nonparametric general models. *Statistics= A Journal of Theoretical and Applied Statistics*, 22(2):205–214, 1991.
- S.I. Resnick. *Heavy-tail phenomena: probabilistic and statistical modeling*. Springer Verlag, 2007. ISBN 0387242724.
- M. Rueda, S. Martínez, H. Martínez, and A. Arcos. Estimation of the distribution function with calibration methods. *Journal of statistical planning and inference*, 137(2):435–448, 2007.
- G. Shorack and J.A. Wellner. *Empirical processes with applications to statistics*. Wiley, 1986.
- Y. Tillé. *Sampling algorithms*. Springer Series in Statistics, 2006.
- A.W. Van der Vaart. *Asymptotic statistics*, volume 3. Cambridge university press, 2000.
- A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes*. Springer, 1996.