Design-based predictive inference

Li-Chun Zhang[1,2]

[1]*Statistisk sentralbyrå*
[2]*University of Southampton (L.Zhang@soton.ac.uk)*

joint work with
Luis Sanguiao-Sande & Danhyang Lee
to appear in Journal of Official Statistics

Design-based prediction by an example

Fixed population U and associated values $y_U = \{y_i : i \in U\}$

Element	i_1	i_2	i_3	i_4
Value y_i	1	2	3	6

SRS: simple random sampling without replacement

Sample s	(i_1, i_2)	(i_1, i_3)	(i_1, i_4)	(i_2, i_3)	(i_2, i_4)	(i_3, i_4)
Sample mean $\bar{y}_s$	1.5	2	3.5	2.5	4	4.5
Out-of-sample mean $\bar{y}_R$	4.5	4	2.5	3.5	2	1.5
Unknown y_k for $k \notin s$	$y_3 = 3$ $y_4 = 6$	$y_2 = 2$ $y_4 = 6$	$y_2 = 2$ $y_3 = 3$	$y_1 = 1$ $y_4 = 6$	$y_1 = 1$ $y_3 = 3$	$y_1 = 1$ $y_2 = 2$
$(y_k - \hat{y}_k)^2 = (y_k - \bar{y}_s)^2$	2.25 20.25	0 16	2.25 0.25	2.25 12.25	9 1	12.25 6.25
$\bar{D}_R = \frac{1}{2}\sum_{k \notin s}(y_k - \bar{y}_s)^2$	11.25	8	1.25	7.25	5	9.25

Random $R = U \backslash s$, $\bar{y}_R$ or $\{y_k : k \notin s\}$ as sample s varies

$E_p(\bar{y}_s - \bar{y}_R) = 0$, i.e. unbiased **prediction** w.r.t. $p(s) \equiv \frac{1}{6}$

$E_p(\bar{D}_R) = 7 =$ MSE of **unit-level** prediction by $\bar{y}_s$

Design-based prediction w.r.t. $p(s)$ is well-defined, but differs completely to model-based prediction.

Descriptive inference for finite populations

Throughout the 20th century, design-based inference by probability sampling from finite-populations has been the standard approach to Official Statistics, insofar as the target parameters are *descriptive, observable summaries* of a given finite population, such as the population total, mean or quantiles of some specific values associated with the given population units.

Such kind of inference is called *descriptive* (Smith, 1983) or *predictive* (Geisser, 1993), which can be contrasted to *analytic inference of theoretical, unobservable targets* such as the life expectancy (of a hypothetical cohort) or a model that can help to understand the given population.

Fundamental epistemological distinction between predictive and analytic inference

Design-based model-assisted inference

Design-based inference can be made more efficient using known *auxiliary* population totals.

- Calibration estimation (Deville and Särndal, 1992)
- Empirical likelihood methods (Hartley and Rao, 1968; Rao and Wu, 2010; Berger and De La Riva Torres, 2016)

In particular, by the *model-assisted* approach, a model is explicitly formulated but inference remains design-based, whether or not the adopted estimator is optimal under the assumed assisting model.

- Generalised regression estimator (GREG) using linear regression models (Särndal et al, 1992)
- Model-calibrated estimators using generalised linear or non-linear models (e.g. Wu and Sitter, 2001)
- A unified "construction recipe" (Breidt and Opsomer, 2017), i.e. model prediction corrected by observed sample residuals

Such model-assisted inference can be *design consistent*, asymptotically as $N, n \to \infty$...

Finite-population Neyman-Fisher consistency

As Smith (1994) points out, the "asymptotic notion of consistency" is not immediately applicable to the given population as "a real entity", such that finite-population Fisher consistent estimators may be desirable.

> For a given population and a fixed sample size, if $t(1), ..., t(k)$ are unbiased estimators of the population totals $T(1), ..., T(k)$, then $g(t(1), ..., t(k))$ is Fisher consistent for $g(T(1), ..., T(k))$, in the sense that replacing $t(j)$ by $T(j)$ would yield the true target population parameter (Fisher 1956).

Neyman (1934) calls an interval estimator "consistent" if it achieves the nominal level of coverage for the given finite population and the prescribed method of sampling.

Finite-population Neyman-Fisher consistency requires design-unbiasedness for any given population target.

Model-assisted design-unbiased estimation

Design-unbiased ratio or linear regression estimators have been proposed by Hartley and Ross (1954) and Mickey (1959), which achieve the “component-wise unbiasedness” required for finite-population Neyman-Fisher consistency.

Sanguiao-Sande and Zhang (2021) develop a technique for design-unbiased model-assisted estimation, called

subsampling Rao-Blackwellisation (SRB),

which allows for *any* assisting Machine Learning (ML) models or algorithms that are increasingly common.
The SRB approach combines three classic ideas:

- model-assisted estimation in survey sampling,
- cross-validation for error estimation in ML,
- Rao-Blackwell Theorem (Rao, 1945; Blackwell, 1947) for efficiency improvement.

Now, use SRB for design-based predictive inference...

Finite-population prediction estimator

Fixed features $x_U = \{x_i : i \in U\}$ in addition to (U, y_U)
Given any $s \subset U$, let $\mu(x, s)$ be a predictor for any out-of-sample unit that is associated with feature vector x.
The *prediction estimator* of $Y = \sum_{i \in U} y_i$ is given as

$$\hat{Y} = \sum_{i \in s} y_i + \sum_{j \in U \setminus s} \mu(x_j, s) \quad (1)$$

where $\mu(x, s)$ is the individual predictor of y given x.

We treat y_U as constants but retain "predictor"; $\mu(x, s)$ can be given by *any* ML model or algorithm.

Eq. (1) includes the Horovitz-Thompson (HT) estimator, given $x_i = \pi_i N/n$, $\pi_i = \Pr(i \in s)$ and $n = |s|$, where

$$\mu_{HT}(x_j, s) = x_j \beta_s + \frac{1}{N - n} \sum_{i \in s} (x_i \beta_s - y_i)$$

and $\beta_s = \frac{1}{n} \sum_{i \in s} y_i / x_i$, using (π_i, N) as auxiliary information.

Sampling-subsampling pq-design

Training-test sample split: $s_1 \cup s_2 = s$ and $s_1 \cap s_2 = \emptyset$ by *subsampling design*:

$$s_1 \sim q(s_1 \mid s), \quad s_2 = s \backslash s_1$$

Typically, s_1 by SRS from s with or without replacement, or T-fold cross-validation where $n_2/n = 1/T$, $s_1 = s \backslash s_2$
E.g. let $s = \{1, 2, ..., 10\}$, by SRSWOR of s_1 with size $n_1 = 6$,

$$s_1 = \{1, 3, 4, 6, 8, 10\} \cup s_2 = \{2, 5, 7, 9\}$$
$$s_1 = \{2, 5, 6, 8, 9, 10\} \cup s_2 = \{1, 3, 4, 7\} \quad ...$$

Sanguiao-Sande and Zhang (2021) refer to the sampling-subsampling design of (s, s_1) as the pq-*design*, denote by

$$f(s_1, s) = q(s_1 \mid s)p(s) = f(s \mid s_1)f(s_1) \tag{2}$$

Given any $s_1 \sim f(s_1)$, s_2 can be regarded as probability sample from $U \backslash s_1$, where $s_1 \cup s_2 = s \sim f(s \mid s_1)$, and

$$\pi_{2i} = \Pr(i \in s_2 \mid s_1) = \sum_{s \ni i, i \notin s_1} f(s \mid s_1) \tag{3}$$

Subsample-trained prediction estimator

Let $\mu(x, s_1)$ be the predictor trained on $\{(y_i, x_i) : i \in s_1\}$, in the same way as $\mu(x, s)$ is trained on $\{(y_i, x_i) : i \in s\}$.
The *subsample-trained* prediction estimator of Y is

$$\hat{Y}_1^* = \sum_{i \in s} y_i + \sum_{j \in U \setminus s} \mu(x_j, s_1) = \sum_{i \in s} y_i + \Big(\sum_{j \in U \setminus s_1} \mu(x_j, s_1) - \sum_{j \in s_2} \mu(x_j, s_1) \Big)$$

such that

$$Y = \sum_{i \in s} y_i + \sum_{j \notin s} y_j = \sum_{i \in s} y_i + \Big(\sum_{j \in U \setminus s_1} y_j - \sum_{j \in s_2} y_j \Big)$$

$$B = \hat{Y}_1^* - Y = \sum_{j \in U \setminus s_1} \{\mu(x_j, s_1) - y_j\} - \sum_{j \in s_2} \{\mu(x_j, s_1) - y_j\} = B_1 - B(s_2)$$

Conditional on s_1, both B and $B(s_2)$ vary with $s_2 \subset U \setminus s_1$ according to $s \sim f(s \mid s_1)$, but B_1 is fixed, *as well as*

$$e_j = \mu(x_j, s_1) - y_j \quad \text{for any } j \in U \setminus s_1 = s_2 \cup (U \setminus s).$$

Unbiased $E_s\{\hat{B}_1 - B(s_2) - B \mid s_1\} = 0$ with $\hat{B}_1 = \sum_{i \in s_2} \frac{e_i}{\pi_{2i}}$

Subsampling Rao-Blackwellisation (SRB)

Applying RB to $\hat{Y}_1^*$ yields the *SRB prediction estimator*

$$\hat{Y}^{RB} = \sum_{i \in s} y_i + \sum_{j \in U \setminus s} \bar{\mu}(x_j, s) \quad (4)$$

where

$$\bar{\mu}(x_j, s) = E_q\{\mu(x_j, s_1) | \, s\} \quad (5)$$

NB. distinguish $\bar{\mu}(x, s)$ from $\mu(x, s)$ trained *once* on s
NB. unordered s as minimal sufficient statistic for $p(s)$

$$\text{Bias}(\hat{Y}^{RB}) = E_p(\hat{Y}^{RB}) - Y = E_{pq}(\hat{Y}_1^*) - Y = E_{pq}(B)$$
$$\stackrel{\text{RB}}{\Longrightarrow} \quad \hat{B}^{RB} = E_q(\hat{B} \mid s)$$

is p-unbiased for $\text{Bias}(\hat{Y}^{RB})$, i.e.

$$E_p(\hat{B}^{RB}) = E_p\{E_q(\hat{B} \mid s)\} = E_{s_1}\{E_s(\hat{B} \mid s_1)\} = E_{s_1}\{E_s(B \mid s_1)\} = E_{pq}(B)$$

Design-unbiased estimator $\hat{B}^{RB}$ of $\text{Bias}(\hat{Y}^{RB})$

Mean squared error of total prediction

Theorem 1. *For any given $\mu(\cdot)$, an unbiased estimator of the MSE of the SRB prediction estimator $\hat{Y}^{RB}$, over $s \sim p(s)$, is given by*

$$\textit{mse}^{RB} = E_q\{\hat{B}^2 - \hat{V}_s(\hat{B} \mid s_1) + \hat{V}_s\{B(s_2) \mid s_1\} \mid s\} - V_q(\hat{Y}_1^* \mid s)$$

where $\hat{B} = \sum_{j \in s_2}(\pi_{2j}^{-1} - 1)\{\mu(x_j, s_1) - y_j\}$, *and* $\hat{V}_s(\hat{B} \mid s_1)$ *is unbiased for*

$$V_s(\hat{B} \mid s_1) = \sum_{i \notin s_1}\sum_{j \notin s_1}(\pi_{2ij} - \pi_{2i}\pi_{2j})(\frac{1}{\pi_{2i}} - 1)(\frac{1}{\pi_{2j}} - 1)e_{1i}e_{1j}$$

where $\pi_{2ij} = \Pr(i, j \in s_2 \mid s_1)$, *and* $\hat{V}_s\{B(s_2) \mid s_1\}$ *is unbiased for*

$$V_s\{B(s_2) \mid s_1\} = \sum_{i \notin s_1}\sum_{j \notin s_1}(\pi_{2ij} - \pi_{2i}\pi_{2j})e_{1i}e_{1j} \ .$$

Design-unbiased estimator mse^{RB} of $\text{MSE}(\hat{Y}^{RB})$
NB. $\text{MSE}(\hat{Y}^{RB}) \approx \text{MSE}(\hat{Y})$ using $\bar{\mu}(x, s)$ or $\mu(x, s)$

An illustration

Generate and fix a population of size $N = 1000$ by

$$y_i = \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i$$

$$x_{1i} \overset{\text{IID}}{\sim} \text{LogN}(1,1), \quad x_{2i} \overset{\text{IID}}{\sim} \text{Poisson}(5), \quad \epsilon_i \overset{\text{IID}}{\sim} N(0, S^2_{x_1}/4)$$

Obtain s by SRSWOR, where $n = 100$. Let $\mu(x_1, s) = a + x_1 b$ be misspecified, on x_1 only, where (a, b) are sample OLS.

MSE estimation from 250 samples, $\mu(x, s)$ for $\hat{Y}$ and Monte Carlo $\bar{\mu}(x, s)$ for $\hat{Y}^{RB}$ with $T = 10^3$, (training, test) set of size (n_1, n_2), RE against variance of HT-estimator.

(n_1, n_2)	MSE($\hat{Y}$)	RE($\hat{Y}$)	MSE($\hat{Y}^{RB}$)	RE($\hat{Y}^{RB}$)	CV($\widetilde{\text{mse}}^{RB}$)
(98, 2)	386532.7	0.44	386632.4	0.44	3.48
(80, 20)	363613.9	0.41	363441.5	0.41	0.31
(70, 30)	362673.0	0.41	357146.9	0.41	0.21

NB. Bias of $\hat{Y}$ and $\hat{Y}^{RB}$ negligible, details omitted here

NB. $n_2 = 20$ or 30 suffices for $\widetilde{\text{mse}}^{RB}$ in practice; exact RB with $n_2 = 2$ has CV(mse^{RB}) = 0.14 and CV($\hat{V}_{HT}$) = 0.32

Summary so far

We obtain design-unbiased bias and MSE estimation of any ML-based SRB-prediction estimator (1) using $\bar{\mu}(x, s)$.

This allows one to use *any* ML models or algorithms that may compare favourably to the traditional estimators by calibration, model-assisted GREG, etc.

Our *inference* of the design-based bias or MSE is finite-population Neyman-Fisher consistent, without the need to resort to (possibly difficult) asymptotic justifications.

It is simple to use the SRB predictor $\bar{\mu}(x, s)$ instead of the once-trained predictor $\mu(x, s)$, as long as tuning or error estimation by cross-validation is needed.

Next, design-based inference at unit/individual level...

Risk of individual prediction

Regardless how $\mu(x, s)$ is obtained from $\{(y_i, x_i) : i \in s\}$, by whichever model or algorithm, its *total squared error (TSE)* over $R = U \backslash s$ is given by

$$D(s; \mu) = \sum_{i \in R} \{\mu(x_i, s) - y_i\}^2$$

For design-based individual-level predictive inference, we define the *risk* of $\mu(x, s)$ to be the expectation of $D(s; \mu)$ over repeated sampling of $s \sim p(s)$, denoted by

$$\tau(\mu) = E_p \{D(s; \mu)\} \quad (6)$$

NB. only s is random in (6), while y_U and x_U are fixed
NB. MSE($\hat{Y}$) for total Y is E_p of *squared total error (STE)*

SRB provides a unified approach to both STE and TSE, by appropriate averaging of subsample-trained $\mu(x, s_1)$.

Subsample-trained $\mu(x, s_1)$

Under the pq-design, $s_1 \cup s_2 = s$, let

$$D_R(s_1; \mu) = \sum_{i \in R} e_i(\mu, s_1)^2 \quad \text{and} \quad e_i(\mu, s_1) = \mu(x_i, s_1) - y_i$$

be the TSE of $\mu(x, s_1)$ over $R = U \backslash s$. Let

$$A(s_2) = \sum_{i \in s_2} e_i(\mu, s_1)^2 = \sum_{i \in U \backslash s_1} e_i(\mu, s_1)^2 - D_R(s_1; \mu)$$

Given s_1, both A_2 and $D_R(s_1; \mu)$ vary with s_2, but their sum $A_1 = \sum_{i \in U \backslash s_1} e_i(\mu, s_1)^2$ is fixed. The predictor

$$\hat{D}_R(s_1; \mu) = \hat{A}_1 - A(s_2) = \sum_{i \in s_2} \pi_{2i}^{-1} e_i(\mu, s_1)^2 - A(s_2)$$

is unbiased for $D_R(s_1; \mu)$ conditional on s_1,

$$E_s\{\hat{D}_R(s_1; \mu) - D_R(s_1; \mu) \mid s_1\} = 0.$$

NB. similarly to $E_s\{\hat{B}_1 - B(s_2) - B \mid s_1\} = 0$ before

Risk of SRB-predictor

Theorem 2. *For any given* $\mu(\cdot)$, $a_i(\mu, s_1) = \mu(x, s_1) - \bar{\mu}(x, s)$, *an unbiased estimator of the risk* $\tau(\bar{\mu})$ *of the corresponding SRB-predictor* $\bar{\mu}(x, s)$, *over* $s \sim p(s)$, *is given by*

$$\hat{D}(s; \bar{\mu}) = E_q\Big(\sum_{i \in s_2} (\pi_{2i}^{-1} - 1) \left\{ e_i(\mu, s_1)^2 - a_i(\mu, s_1)^2 \right\} \mid s \Big) \, .$$

When exact SRB is infeasible, one can use the Monte Caro SRB predictor based on T subsamples

$$\begin{cases} \tilde{\mu}(x_i, s) = T^{-1} \, \sum_{t=1}^{T} \mu(x_i, s_1^{(t)}) & \textbf{if} \;\; i \in R \\ \mathring{\mu}(x_i, s) = T_i^{-1} \sum_{t=1}^{T} \mathbb{I}\big(i \notin s_1^{(t)}\big) \mu(x_i, s_1^{(t)}) & \textbf{if} \;\; i \in s \end{cases}$$

where $T_i = \sum_{t=1}^{T} \mathbb{I}\big(i \notin s_1^{(t)}\big)$, and the risk estimator

$$\tilde{D}(s; \bar{\mu}) = \frac{1}{T} \sum_{t=1}^{T} \sum_{i \in s_2^{(t)}} \big(\pi_{2i}^{-1} - 1\big) \{ e_i(\mu, s_1^{(t)})^2 - a_i(\mu, s_1^{(t)})^2 \}$$

$$a_i(\mu, s_1^{(t)}) = \mu(x_i, s_1^{(t)}) - \mathring{\mu}(x_i, s) \quad \Lsh$$

Illustration

Generate *ad hoc* 200 sets of y_U, each of size $N = 2000$, where half of them by M1 and half by M2,

$$\text{(M1)} \quad y = x_1 + 0.5x_2 + \epsilon, \quad \epsilon \sim \begin{cases} N(0,1) & \text{if } z = 1 \Leftrightarrow x_2 < 3 \\ N(-2,1) & \text{if } z = 2 \Leftrightarrow 3 \leqslant x_2 < 7 \\ N(2,1) & \text{if } z = 3 \Leftrightarrow x_2 \geqslant 7 \end{cases}$$

$$\text{(M2)} \quad y = 0.5 + 1.5x_1 + x_2 + \epsilon, \quad \epsilon \sim \chi_1^2 + N(0, 0.25)$$

$$x_1 \overset{\text{IID}}{\sim} N(0,1) \qquad x_2 \overset{\text{IID}}{\sim} \text{Poisson}(5)$$

From each population, draw sample s by

- SRS of size $n = 200$
- Poisson Sampling, with $\pi_i^{-1} \propto 1 + 1/\exp(\alpha + 0.5y_i)$ and $\sum_{i \in U} \pi_i = n$, where $\alpha \in \{1, -0.1, -1\}$ leads to coefficient of variation of π_i over U, denoted by cv_π, to be about 15%, 30% and 45%, respectively.

Models

- linear regression (LR)
- random forest (RF)
- support vector machine (SVM)

Illustration

MSE and estimates given LR, RF or SVM over 200 simulations.
PS, Poisson Sampling; CrV-based, cross-validation-based.

	SRSWOR			PS (cv$_\pi$=15%)		
MSE $D/\|R\|$	LR	RF	SVM	LR	RF	SVM
Average, true	8.399	9.013	9.272	8.566	9.225	9.671
Design, proposed	8.409	9.073	9.326	8.416	9.182	9.615
Model, CrV	8.457	9.481	9.862	8.014	9.214	9.405
Model, residual	8.162	5.105	7.706	7.766	4.945	7.578
	PS (cv$_\pi$=30%)			PS (cv$_\pi$=45%)		
MSE $D/\|R\|$	LR	RF	SVM	LR	RF	SVM
Average, true	8.957	9.726	10.451	9.866	10.884	11.573
Design, proposed	8.711	9.559	10.196	9.288	10.364	10.974
Model, CrV	7.624	8.880	8.799	6.992	8.262	7.933
Model, residual	7.369	4.731	7.330	6.776	4.367	6.758

CrV-based mse $= \frac{1}{T}\sum_{t=1}^{T}\frac{1}{n_2}\sum_{i\in s_2^{(t)}}\{\mu(x_i, s_1^{(t)}) - y_i\}^2$ and
residual-based mse $= \frac{1}{n}\sum_{i\in s}(\tilde{\mu}(x_i, s) - y_i)^2$ under IID error model

NB. IID-model CrV-based MSE is biased under Poisson Sampling

Spanish Structural Business Survey 2020

Stratified SRS, take-some stratum sample size > 2
No. strata $= 9\ 681$, $|U| = 2\ 018\ 561$, $|s| = 80\ 280$
Reduced sample: $|s^*| = 40\ 514$, stratum sample size > 2
Models:

- LR, $x =$ admin turnover, operating income (by model selection)
- RF, additional features 1st-digit NACE, no. employees

q-design: SRS, 80-20 split for LR, 50-50 for RF

Results $(\times 10^9)$, $T = 10^5$, reduced sample size if unspecified

Estimator, model	$\hat{Y}$	Bias	MSE	RErr	MC error
HT-estimator (full sample size)	258	0	94	0.04	-
HT-estimator	252	0	151	0.05	-
SRB-estimator, LR	229	0	122	0.05	1
SRB-estimator, RF	234	0	107	0.04	2
SRB-prediction estimator, LR	227	-2	50	0.03	3
SRB-prediction estimator, RF	238	4	27	0.02	5

SRB-estimator, design-unbiased (Sanguiao-Sande & Zhang, 2021)

Final remarks

Design-based predictive inference from finite-population probability sampling is developed for the first time.

In addition to population-level estimation, it provides a theoretical basis for creating census-like population data or statistical registers for descriptive statistics.

Finite-population design-unbiased estimation of the bias and MSE of prediction are obtained, without the need of asymptotic justifications, given arbitrary ML model or algorithm (either existing or yet to be invented).

Finally, some obvious, non-exhaustive topics in future:

- Lee et al. (2022) apply ensemble-SRB to missing data imputation. A unified quasi-randomisation approach?
- Other individual prediction losses, coverage of interval estimator for population total...
- Better balance between total and individual prediction?

References

[1] Bates, S., Hastie, T. and Tibshirani, R. (2023) Cross-Validation: What Does It Estimate and How Well Does It Do It? *Journal of the American Statistical Association*, https://doi.org/10.1080/01621459.2023.2197686

[2] Beaumont, J.-F. and Haziza, D. (2022). Statistical inference from finite population samples: A critical review of frequentist and Bayesian approaches. *The Canadian Journal of Statistics*, 50:1186-1212.

[3] Berger, Y.G. and De La Riva Torres, O. (2016). Empirical Likelihood Confidence Intervals for Complex Sampling Designs. *Journal of the Royal Statistical Society Series B*, 78:319-341.

[4] Blackwell, D. (1947). Conditional expectation and unbiased sequential estimation. *Annals of Mathematical Statististics*, 18: 105-110.

[5] Breidt, F. J. and Opsomer, J. D. (2017). Model-assisted survey estimation with modern prediction techniques. *Statistical Science*, 32:190-205.

[6] Deville, J.-C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.

[7] Deville, J.-C. and Tillé, Y. (2004). Efficient balanced sampling: The cube method. *Biometrika*, 91:893-912.

[8] Dietterich, T.G. (2000). Ensemble methods in machine learning. In *International workshop on multiple classifier systems*. Springer.

[9] Dong, X., Yu, Z., Cao, W., Shi, Y. and Ma, Q. (2020). A survey on ensemble learning. *Frontiers of Computer Science*, 14:241-258.

[10] Fisher, R.A. (1956). *Statistical Methods and Scientific Inference*. Oliver and Boyd, Edinburgh and London.

[11] Geisser, S. (1993). *Predictive Inference*. Chapman & Hall.

[12] Hansen, M. (1987). Some History and Reminiscences on Survey Sampling. *Statistical Science*, 2:180-190.

[13] Hansen, M. H., Madow, W. G., and Tepping, B. J. (1983). An evaluation of dependent and probability sampling inferences in sample surveys. *Journal of the American Statistical Association*, 78:776-793.

[14] Hartley, H. O. and Rao, J. N. K. (1968). A new estimation theory for sample surveys. *Biometrika*, 55, 547-557.

[15] Hartley, H. O. and Ross, A. (1954). Unbiased Ratio Estimators. *Nature*, August 7, pp. 270-271.

[16] Horvitz, D. G. and Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663-685.

[17] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.

[18] Lee, D., Zhang, L.-C. and Chen, S. (2022). Robust quasi-randomization-based estimation with ensemble learning for missing data. *Scandinavian Journal of Statistics*. DOI:10.1111/sjos.12626

[19] Mickey, M.R. (1959). Some finite population unbiased ratio and regression estimators. *Journal of the American Statistical Association*, 54:594-612.

[20] Neyman, J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. *Journal of the Royal Statistical Society, pp. 558-625.*

[21] Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of Calcutta Mathematical Society*, 37:81-91.

[22] Rao, J. N. K. (2005). Interplay between sample survey theory and practice: An appraisal. *Survey Methodology*, 31:117-138.

[23] Rao, J. N. K. (2011). Impact of frequentist and Bayesian methods on survey sampling practice: A selective appraisal. *Statistical Science*, 26:240-256.

[24] Rao, J. N. K. and Wu, C. (2010). Bayesian pseudo-empirical-likelihood intervals for complex surveys. *Journal of the Royal Statistical Society B*, 72:533-544.

[25] Sagi, O. and Rokach, L. (2018). Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, e1249.

[26] Sanguiao-Sande, L. and Zhang, L.-C. (2021). Design-Unbiased Statistical Learning in Survey Sampling. *Sankhya A*, Centenary Issue in Honour of C. R. Rao, 83:714-744.

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

[28] Smith, T.M.F. (1983). On the validity of inferences from non-random sample. *Journal of the Royal Statistical Society, Series A*, 146:394-403.

[29] Smith, T. M. F. (1994). Sample surveys 1975–1990; an age of reconciliation? (with discussion). *International Statistical Review*, 62:5-34.

[30] Valliant, R., Dorfman, R. M., and Royall, R. M. (2000). *Finite Population Sampling and Inference: A Prediction Approach*. Wiley, New York.

[31] Wu, C. and Sitter, R. R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *Journal of the American Statistical Association*, 96:185-193.

[32] Zhou, Z.H. (2012). *Ensemble Methods: Foundations and Algorithms*. CRC press.