1 Sample space of two-stage sampling & its probability function

Let $U^{(1)}$ be a finite set of size $N^{(1)}$. Let $U_1^{(2)}, U_2^{(2)}, \ldots, U_{N^{(1)}}^{(2)}$ be finite sets of sizes $N_1^{(2)}, N_2^{(2)}, \ldots, N_{N^{(1)}}^{(2)}$, respectively. For each $i = 1, 2, \ldots, N^{(1)}$, we enumerate the elements of $U_i^{(2)}$ as follows:

$$U_i^{(2)} \;=\; \left\{\, u_{i1}, u_{i2}, \ldots, u_{iN_i^{(2)}} \,\right\} \;=\; \left\{\, u_{ik} \;\middle\vert\; k = 1, 2, \ldots, N_i^{(2)} \,\right\}$$

Let

$$U \;:=\; \bigsqcup_{i \in U^{(1)}} U_i^{(2)} \;=\; \left\{\, u_{ik} \;\middle\vert\; i = 1, 2, \ldots, N^{(1)},\, k = 1, 2, \ldots, N_i^{(2)} \,\right\}$$

Let $p^{(1)} : \mathcal{S}^{(1)} \longrightarrow (0,1]$ be our chosen first-stage sampling design, where $\mathcal{S}^{(1)} \subseteq \mathcal{P}\big(U^{(1)}\big)$ is the set of all possible first-stage samples in the design, and $\mathcal{P}\big(U^{(1)}\big)$ is the power set of $U^{(1)}$.

For each $i \in U^{(1)}$, let $p_i^{(2)} : \mathcal{S}_i^{(2)} \longrightarrow (0,1]$ be our chosen second-stage sampling design, where $\mathcal{S}_i^{(2)} \subseteq \mathcal{P}\Big(U_i^{(2)}\Big)$ is the set of all possible second-stage samples in the design, and $\mathcal{P}\Big(U_i^{(2)}\Big)$ is the power set of $U_i^{(2)}$.

The sample space $\mathcal{S}$ of the two-stage sampling design is:

$$\mathcal{S} \;:=\; \left\{\, \left(s^{(1)}, \left\{\, s_i^{(2)} \,\right\}_{i \in U^{(1)}}\right) \in \mathcal{S}^{(1)} \times \prod_{i \in U^{(1)}} \mathcal{S}_i^{(2)} \;\middle\vert\; \begin{array}{ll} s_i^{(2)} \in \mathcal{S}_i^{(2)}, & \text{if } i \in s^{(1)} \\ s_i^{(2)} = \varnothing, & \text{if } i \notin s^{(1)} \end{array} \,\right\}$$

We will use the following abbreviation for an element in $\mathcal{S}$:

$$s \;=\; \left(\, s^{(1)} \,, \left\{\, s_i^{(2)} \,\right\}_{i \in s^{(i)}} \right)$$

We now define the probability function $p : \mathcal{S} \longrightarrow (0,1]$ as follows: For each $s \in \mathcal{S}$,

$$p(s) \;:=\; p\left(\left(\, s^{(1)} \,, \left\{\, s_i^{(2)} \,\right\}_{i \in s^{(1)}} \right)\right) \;=\; p^{(1)}\Big(s^{(1)}\Big) \cdot \prod_{i \in s^{(1)}} p_i^{(2)}\Big(s_i^{(2)}\Big)$$

We will prove in Proposition 1.2 below that $p : \mathcal{S} \longrightarrow (0,1]$ defined above is indeed a probability function on $\mathcal{S}$.

Lemma 1.1 *For each first-stage sample $s^{(1)} \in \mathcal{S}^{(1)}$, let $\Omega\big(s^{(1)}\big) := \left\{\, s_i^{(2)} \in \mathcal{S}_i^{(2)} \;\middle\vert\; i \in s^{(1)} \,\right\}$, i.e. $\Omega\big(s^{(1)}\big)$ is the collection of all second-stage samples compatible with the first-stage sample $s^{(1)} \in \mathcal{S}^{(1)}$. Then, we have:*

$$\sum_{\xi \in \Omega(s^{(1)})} p\Big((\, s^{(1)}, \xi \,)\Big) \;=\; p^{(1)}\Big(s^{(1)}\Big).$$

PROOF Let n be the number of elements in $s^{(1)}$, we write $s^{(1)} = \{\, i_1, i_2, \ldots, i_n \,\}$. Then,

$$p\left(\left(\, s^{(1)} \,, \left\{\, s_{i_1}^{(2)}, s_{i_2}^{(2)}, \ldots, s_{i_n}^{(2)} \,\right\}\right)\right) \;=\; p^{(1)}\Big(s^{(1)}\Big) \cdot p_{i_1}^{(2)}\Big(s_{i_1}^{(2)}\Big) \cdot p_{i_2}^{(2)}\Big(s_{i_2}^{(2)}\Big) \;\cdots\; p_{i_n}^{(2)}\Big(s_{i_n}^{(2)}\Big)$$

Hence,

$$\begin{aligned}
\sum_{\xi\in\Omega(s^{(1)})} p\Big(\big(\,s^{(1)},\xi\,\big)\Big) &= \sum_{\zeta_1\in\mathcal{S}^{(2)}_{i_1}}\sum_{\zeta_2\in\mathcal{S}^{(2)}_{i_2}}\cdots\sum_{\zeta_n\in\mathcal{S}^{(2)}_{i_n}} p\Big(\Big(\,s^{(1)}\,,\{\,\zeta_1,\zeta_2,\ldots,\zeta_n\,\}\Big)\Big) \\
&= \sum_{\zeta_1\in\mathcal{S}^{(2)}_{i_1}}\sum_{\zeta_2\in\mathcal{S}^{(2)}_{i_2}}\cdots\sum_{\zeta_n\in\mathcal{S}^{(2)}_{i_n}} p^{(1)}\Big(s^{(1)}\Big)\cdot p^{(2)}_{i_1}(\zeta_1)\cdot p^{(2)}_{i_2}(\zeta_2)\,\cdots\,p^{(2)}_{i_n}(\zeta_n) \\
&= p^{(1)}\Big(s^{(1)}\Big)\cdot\left(\sum_{\zeta_1\in\mathcal{S}^{(2)}_{i_1}} p^{(2)}_{i_1}(\zeta_1)\right)\cdot\left(\sum_{\zeta_2\in\mathcal{S}^{(2)}_{i_2}} p^{(2)}_{i_2}(\zeta_2)\right)\cdots\left(\sum_{\zeta_n\in\mathcal{S}^{(2)}_{i_n}} p^{(2)}_{i_n}(\zeta_n)\right) \\
&= p^{(1)}\Big(s^{(1)}\Big)\cdot(\,1\,)\cdot(\,1\,)\,\cdots\,(\,1\,) \\
&= p^{(1)}\Big(s^{(1)}\Big)
\end{aligned}$$

□

Proposition 1.2

$$\sum_{s\in\mathcal{S}} p(s) \;=\; 1$$

Proof

$$\sum_{s\in\mathcal{S}} p(s) \;=\; \sum_{(s^{(1)},\xi)\in\mathcal{S}} p(s^{(1)},\xi) \;=\; \sum_{s^{(1)}\in\mathcal{S}^{(1)}}\;\sum_{\xi\in\Omega(s^{(1)})} p(s^{(1)},\xi) \;=\; \sum_{s^{(1)}\in\mathcal{S}^{(1)}} p^{(1)}(s^{(1)}) \;=\; 1,$$

where the second-last equality follows from the preceding Lemma. □

2 Estimation in two-stage sampling

Let $\mathbf{y} : U \longrightarrow \mathbb{R}^m$ be an $\mathbb{R}^m$-valued function defined on U (such a $\mathbf{y}$ commonly called a "population parameter"). We will use the common notation $\mathbf{y}_{kl}$ for $\mathbf{y}(u_{kl})$, for $k = 1, 2, \ldots, N^{(1)}$ and $l = 1, 2, \ldots, N^{(2)}_k$. We wish to estimate

$$\mathbf{T}_{\mathbf{y}} \;:=\; \sum_{u\in U}\mathbf{y}(u) \;=\; \sum_{k\in U^{(1)}}\;\sum_{l\in U^{(2)}_k}\mathbf{y}_{kl} \;=\; \sum_{k=1}^{N^{(1)}}\sum_{l=1}^{N^{(2)}_k}\mathbf{y}_{kl} \;\in\; \mathbb{R}^m$$

via two-stage sampling. We consider estimators for $\widehat{T}_{\mathbf{y}}$ of the following form:

$$\begin{aligned}
\widehat{\mathbf{T}}_{\mathbf{y}} :\quad \mathcal{S} \quad &\longrightarrow\quad \mathbb{R}^m \\
\Big(s^{(1)},\{s^{(2)}_k\}_{k\in s^{(1)}}\Big) \quad &\longmapsto\quad \sum_{k\in s^{(1)}} w^{(1)}_k(s^{(1)})\,\widehat{\mathbf{T}}_{\mathbf{y}|k}(s^{(2)}_k) \;=\; \sum_{k\in U^{(1)}} I_k(s^{(1)})\,w^{(1)}_k(s^{(1)})\,\widehat{\mathbf{T}}_{\mathbf{y}|k}(s^{(2)}_k),
\end{aligned}$$

where, for each $k \in U^{(1)}$, $w^{(1)}_k : \mathcal{S}^{(1)} \longrightarrow \mathbb{R}$ and $\widehat{\mathbf{T}}_{\mathbf{y}|k} : \mathcal{S}^{(2)}_k \longrightarrow \mathbb{R}^m$ are random variables.

Proposition 2.1 *Suppose:*

- *The first-stage weights $w^{(1)}_k(\,\cdot\,)$ satisfy the following:*

$$E^{(1)}\Big[\;\widehat{T}_z\;\Big] \;=\; T_z \;:=\; \sum_{k\in U^{(1)}} z_k,\quad \text{for any function } z : U^{(1)} \longrightarrow \mathbb{R},$$

where $\widehat{T}_z : \mathcal{S}^{(1)} \longrightarrow \mathbb{R}$ is a random variable defined by $\widehat{T}_z(s^{(1)}) := \sum_{k\in s^{(1)}} w^{(1)}_k(s^{(1)})\,z_k$.

- *For each $k \in U^{(1)}$, the random variable $\widehat{\mathbf{T}}_{\mathbf{y}|k} : \mathcal{S}^{(2)}_k :\longrightarrow \mathbb{R}$ is a design-unbiased estimator for $T_{\mathbf{y}|k}$, i.e.*

$$E^{(2)}_k\left[\,\widehat{\mathbf{T}}_{\mathbf{y}\,|\,k}\,\right] \;=\; \mathbf{T}_{\mathbf{y}\,|k} \;:=\; \sum_{l\in U^{(2)}_k} \mathbf{y}_{kl}$$

Definition 2.2

A random variable $\widehat{\mathbf{T}}_{\mathbf{y}} : \mathcal{S} \longrightarrow \mathbb{R}^m$ is said to be linear *in the population parameter* $\mathbf{y} : U \longrightarrow \mathbb{R}$ *if it has the following form:*

$$\begin{array}{rccl} \widehat{\mathbf{T}}_{\mathbf{y}} : & \mathcal{S} & \longrightarrow & \mathbb{R}^m \\ & s & \longmapsto & \sum_{k\in s} w_k(s)\,\mathbf{y}_k \;=\; \sum_{k\in U} I_k(s)\,w_k(s)\,\mathbf{y}_k, \end{array}$$

where, for each $k \in U$, $w_k : \mathcal{S} \longrightarrow \mathbb{R}$ is itself an $\mathbb{R}$-valued random variable, and $I_k : \mathcal{S} \longrightarrow \{0, 1\}$ is the indicator random variable defined by:

$$I_k(s) \;=\; \left\{\begin{array}{ll} 1, & \text{if } k \in s, \\ 0, & \text{otherwise} \end{array}\right.$$

We call the w_k's the weights *of $\widehat{\mathbf{T}}_{\mathbf{y}}$, and we use the notation $\widehat{\mathbf{T}}_{\mathbf{y};w}$ to indicate that the random variable depends on the weights w_k.*

Nomenclature In the context of finite-population probability sampling, under a design $p : \mathcal{S} \longrightarrow (0, 1]$, an "estimator" is precisely just a random variable defined on the space $\mathcal{S}$ of all admissible samples in the design.

Proposition 2.3

Let $\widehat{\mathbf{T}}_{\mathbf{y};w} : \mathcal{S} \longrightarrow \mathbb{R}^m$, with $\widehat{\mathbf{T}}_{\mathbf{y};w}(s) = \sum_{k\in U} I_k(s)\,w_k(s)\,\mathbf{y}_k = \sum_{k\in s} w_k(s)\,\mathbf{y}_k$, be a random variable linear in the population parameter $\mathbf{y} : U \longrightarrow \mathbb{R}$. Then,

$$E\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w}\,\right] \;=\; \mathbf{T}_{\mathbf{y}},\ \text{for arbitrary } \mathbf{y} \quad\Longleftrightarrow\quad E[\,I_k\,w_k\,] \;=\; 1,\ \text{for each } k \in U.$$

Proof Note:

$$E\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w}\,\right] \;=\; E\left[\,\sum_{k\in s} w_k\,\mathbf{y}_k\,\right] \;=\; E\left[\,\sum_{k\in U} I_k\,w_k\,\mathbf{y}_k\,\right] \;=\; \sum_{k\in U} E[\,I_k\,w_k\,]\;\mathbf{y}_k$$

Hence, since $\mathbf{y} : U \longrightarrow \mathbb{R}$ is arbitrary,

$$E\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w}\,\right] \;=\; \mathbf{T}_{\mathbf{y}} \;:=\; \sum_{k\in U}\mathbf{y}_k \quad\Longleftrightarrow\quad \sum_{k\in U}\left(E[\,I_k\,w_k\,]-1\right)\cdot\mathbf{y}_k \;=\; \mathbf{0} \quad\Longleftrightarrow\quad E[\,I_k\,w_k\,] \;=\; 1,\ \text{for each } k \in U.$$

The proof of the Proposition is now complete. □

Corollary 2.4

Let $U = \{1, 2, \ldots, N\}$ be a finite population. For any fixed but arbitrary population parameter $\mathbf{y} : U \longrightarrow \mathbb{R}^m$ and for any sampling design $p : \mathcal{S} \longrightarrow (0, 1]$ such that each of its first-order inclusion probabilities is strictly positive, the Horvitz-Thompson estimator $\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}$ is well-defined and it is the unique unbiased estimator for $\mathbf{T}_{\mathbf{y}}$, which is linear in $\mathbf{y}$ and whose weights are constant in s.

Proof Recall that the Horvitz-Thompson estimator is defined as:

$$\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}(s) \;:=\; \sum_{k\in s}\frac{1}{\pi_k}\,\mathbf{y}_k \;:=\; \sum_{k\in U} I_k(s)\,\frac{1}{\pi_k}\,\mathbf{y}_k,$$

where $\pi_k := E[\,I_k\,] = \sum_{k\in U} p(s)\,I_k(s) = \sum_{s\ni k} p(s)$ is the inclusion probability of $k \in U$ under the sampling design $p : \mathcal{S} \longrightarrow (0, 1]$. Clearly, $\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}$ is linear in $\mathbf{y}$ with weights constant in s. Next, note that:

$$E\left[\,\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}\,\right] \;=\; E\left[\,\sum_{k\in s}\frac{1}{\pi_k}\,\mathbf{y}_k\,\right] \;=\; E\left[\,\sum_{k\in U} I_k\,\frac{\mathbf{y}_k}{\pi_k}\,\right] \;=\; \sum_{k\in U} E[\,I_k\,]\;\frac{\mathbf{y}_k}{\pi_k} \;=\; \sum_{k\in U}\pi_k\,\frac{\mathbf{y}_k}{\pi_k} \;=\; \sum_{k\in U}\mathbf{y}_k \;=\; \mathbf{T}_y$$

Hence, $\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}$ is an unbiased estimator for $\mathbf{T}_{\mathbf{y}}$. Conversely, let

$$\widehat{\mathbf{T}}_{y;w}(s) \;=\; \sum_{k\in s} w_k\,\mathbf{y}_k$$

be any unbiased estimator for $\mathbf{T}_{\mathbf{y}}$ which linear in $\mathbf{y}$ with weights w_k constant in s. Thus,

$$\sum_{k\in U}\mathbf{y}_k \;=\; \mathbf{T}_{\mathbf{y}} \;=\; E\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w}\,\right] \;=\; E\left[\,\sum_{k\in s} w_k\,\mathbf{y}_k\,\right] \;=\; E\left[\,\sum_{k\in U} I_k\,w_k\,\mathbf{y}_k\,\right] \;=\; \sum_{k\in U} E[\,I_k\,]\,w_k\,\mathbf{y}_k \;=\; \sum_{k\in U}\pi_k\,w_k\,\mathbf{y}_k.$$

Since $\mathbf{y}$ is arbitrary, the above equation immediately implies that

$$\pi_k w_k - 1 = 0,$$

or equivalently, $w_k = \dfrac{1}{\pi_k}$; in other words, $\widehat{\mathbf{T}}_{\mathbf{y};w}$ is in fact equal to the Horvitz-Thompson estimator. The proof of the Corollary is now complete. □

Lemma 2.5

Let $(\Omega,\mathcal{A},p)$ be a probability space, $X, Y : \Omega \longrightarrow \mathbb{R}$ be two $\mathbb{R}$-valued random variables defined on Ω, and $\mathbf{u},\mathbf{v}\in\mathbb{R}^m$ be two fixed vectors in $\mathbb{R}^m$. Then,

$$\mathrm{Cov}(\,X\cdot\mathbf{u}\,,\,Y\cdot\mathbf{v}\,) \;=\; \mathrm{Cov}(X,Y)\cdot\mathbf{u}\cdot\mathbf{v}^T \;\in\; \mathbb{R}^{m\times m}$$

Proof Note:

$$\begin{aligned}
\mathrm{Cov}(\,X\cdot\mathbf{u}\,,\,Y\cdot\mathbf{v}\,) &:= E\left[\,(X\,\mathbf{u}-\mu_X\mathbf{u})\cdot(Y\,\mathbf{v}-\mu_Y\mathbf{v})^T\,\right] \;=\; E\left[\,(X-\mu_X)\,\mathbf{u}\cdot(Y-\mu_Y)\,\mathbf{v}^T\,\right] \\
&= E\left[\,(X-\mu_X)\,(Y-\mu_Y)\cdot\mathbf{u}\cdot\mathbf{v}^T\,\right] \;=\; E[\,(X-\mu_X)\,(Y-\mu_Y)\,]\cdot\mathbf{u}\cdot\mathbf{v}^T \\
&= \mathrm{Cov}(X,Y)\cdot\mathbf{u}\cdot\mathbf{v}^T,
\end{aligned}$$

as required. □

Proposition 2.6

Let $\widehat{\mathbf{T}}_{\mathbf{y};w} : \mathcal{S}\longrightarrow\mathbb{R}$, with $\widehat{\mathbf{T}}_{\mathbf{y};w}(s) = \sum_{k\in s} w_k(s)\,\mathbf{y}_k = \sum_{k\in U} I_k(s)\,w_k(s)\,\mathbf{y}_k$, be a random variable linear in the population parameter $\mathbf{y}: U\longrightarrow\mathbb{R}$. Then, the covariance matrix of $\widehat{\mathbf{T}}_{\mathbf{y};w}$ is given by:

$$\mathrm{Var}\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w}\,\right] \;=\; \sum_{i\in U}\sum_{k\in U}\mathrm{Cov}[\,I_i\,w_i\,,\,I_k\,w_k\,]\;\mathbf{y}_i\cdot\mathbf{y}_k^T \;\in\mathbb{R}^{m\times m}$$

Furthermore, if the first-order and second-order inclusion probabilities of the sampling design $p:\mathcal{S}\longrightarrow(0,1]$ are all strictly positive, i.e. $\pi_k = \pi_{kk} := \sum_{s\ni k} p(s) > 0$, for each $k\in U$, and $\pi_{ik} := \sum_{s\ni i,k} p(s) > 0$, for any distinct $i,k\in U$, then an unbiased estimator for $\mathrm{Var}\left[\,\widehat{\mathbf{T}}_{y;w}\,\right]$ is given by:

$$\widehat{\mathrm{Var}}\left[\,\widehat{\mathbf{T}}_{y;w}\,\right](s) \;:=\; \sum_{i,k\in s}\frac{\mathrm{Cov}(I_iw_i,I_kw_k)}{\pi_{ik}}\,\mathbf{y}_i\cdot\mathbf{y}_k^T \;=\; \sum_{k\in s}\frac{\mathrm{Var}(I_kw_k)}{\pi_k}\,\mathbf{y}_k\cdot\mathbf{y}_k^T + \sum_{\substack{i,k\in s\\ i\neq k}}\frac{\mathrm{Cov}(I_iw_i,I_kw_k)}{\pi_{ik}}\,\mathbf{y}_i\cdot\mathbf{y}_k^T,\;\text{ for each } s\in\mathcal{S}.$$

Proof First, note that Lemma 2.5 implies:

$$\mathrm{Var}\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w}\,\right] \;=\; \mathrm{Cov}\left[\,\sum_{i\in U} I_i\,w_i\,\mathbf{y}_i\,,\,\sum_{k\in U} I_k\,w_k\,\mathbf{y}_k\,\right] \;=\; \sum_{i\in U}\sum_{k\in U}\mathrm{Cov}[\,I_i\,w_i\,,\,I_k\,w_k\,]\cdot\mathbf{y}_i\cdot\mathbf{y}_k^T \;\in\; \mathbb{R}^{m\times m}$$

Next,

$$
\begin{aligned}
E\Big(\widehat{\mathrm{Var}}\Big[\, \widehat{\mathbf{T}}_{\mathbf{y};w} \,\Big] \Big) &= \sum_{s\in\mathcal{S}} p(s)\cdot \widehat{\mathrm{Var}}\Big[\, \widehat{\mathbf{T}}_{\mathbf{y};w} \,\Big](s) \;=\; \sum_{s\in\mathcal{S}} p(s)\cdot \left(\sum_{i,k\in s} \frac{\mathrm{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \right) \\
&= \sum_{s\in\mathcal{S}} p(s)\cdot \left(\sum_{i,k\in U} I_i(s) I_k(s) \cdot \frac{\mathrm{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \right) \\
&= \sum_{i,k\in U} \left(\sum_{s\in\mathcal{S}} p(s) I_i(s) I_k(s) \right) \cdot \frac{\mathrm{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \\
&= \sum_{i,k\in U} \left(\sum_{s\ni i,k} p(s) \right) \cdot \frac{\mathrm{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \\
&= \sum_{i,k\in U} \pi_{ik} \cdot \frac{\mathrm{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \;=\; \sum_{i,k\in U} \mathrm{Cov}(I_i w_i, I_k w_k) \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \\
&= \mathrm{Var}\Big[\, \widehat{\mathbf{T}}_{y;w} \,\Big]
\end{aligned}
$$

Lastly, recall that $\pi_{kk} = \pi_k$ and $\mathrm{Cov}(I_k w_k, I_k w_k) = \mathrm{Var}[\, I_k w_k \,]$, and the validity of the following identity is thus trivial:

$$
\sum_{i,k\in s} \frac{\mathrm{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T \;=\; \sum_{k\in s} \frac{\mathrm{Var}(I_k w_k)}{\pi_k} \cdot \mathbf{y}_k \cdot \mathbf{y}_k^T + \sum_{\substack{i,k\in s \\ i\neq k}} \frac{\mathrm{Cov}(I_i w_i, I_k w_k)}{\pi_{ik}} \cdot \mathbf{y}_i \cdot \mathbf{y}_k^T
$$

The proof of the Proposition is complete. $\square$

3 Calibrated linear estimators for (multivariate) population totals

Definition 3.1

Let $\widehat{\mathbf{T}}_{\mathbf{y};w} : \mathcal{S} \longrightarrow \mathbb{R}^m$ be an $\mathbb{R}^m$-valued random variable which is linear in the $\mathbb{R}^m$-valued population parameter $\mathbf{y} : U \longrightarrow \mathbb{R}^m$, i.e.

$$
\begin{array}{rccccl}
\widehat{\mathbf{T}}_{\mathbf{y};w} : & \mathcal{S} & \longrightarrow & \mathbb{R}^m & & \\
& s & \longmapsto & \sum_{k\in s} w_k(s)\cdot \mathbf{y}_k & = & \sum_{k\in U} I_k(s)\, w_k(s)\cdot \mathbf{y}_k,
\end{array}
$$

where, for each $k \in U$, $w_k : \mathcal{S} \longrightarrow \mathbb{R}$ is itself an $\mathbb{R}$-valued random variable, and $I_k : \mathcal{S} \longrightarrow \{0,1\}$ is the indicator random variable defined by:

$$
I_k(s) \;=\; \begin{cases} 1, & \text{if } k \in s, \\ 0, & \text{otherwise} \end{cases}
$$

Let $x : U \longrightarrow \mathbb{R}$ be an $\mathbb{R}$-valued population parameter and $T_x := \sum_{k\in U} x_k$.

Then, $\widehat{\mathbf{T}}_{\mathbf{y};w}$ is said to be <u>calibrated with respect to x</u> *if*

$$
\sum_{k\in s} w_k(s)\, x_k \;=\; T_x, \quad \text{for each } s \in \mathcal{S}.
$$

Example 3.2

If the sampling design has fixed sample size and each of its first-order inclusion probabilities is strictly positive, then Horvitz-Thompson estimator is calibrated with respect to the first-order inclusion probabilities.

To see this, let $U = \{1, 2, \ldots, N\}$ be a finite population, $\mathbf{y} : U \longrightarrow \mathbb{R}^m$ a population parameter, and $p : \mathcal{S} \subset \mathcal{P}(U) \longrightarrow (0,1]$ a sampling design such that $\pi_k := \sum_{s\ni k} p(s) > 0$, for each $k \in U$. The Horvitz-Thompson estimator $\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}} : \mathcal{S} \longrightarrow \mathbb{R}$ is then well-defined and is given by:

$$
\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}(s) \;:=\; \sum_{k\in s} \frac{\mathbf{y}_k}{\pi_k}
$$

Let $x : U \longrightarrow \mathbb{R}$ be defined by

$$x_k \;=\; \pi_k, \quad \text{for each } k \in U,$$

i.e. x_k is simply the inclusion probability of $k \in U$ under the sampling design $p : \mathcal{S} \longrightarrow (0,1]$.

Now, suppose that the sampling design has a fixed sample size n, and we shall show that $\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}$ is consequently calibrated with respect to $x : U \longrightarrow \mathbb{R}$. Indeed, recall that the weights of the Horvitz-Thompson estimator are simply $w_k(s) = 1/\pi_k$, for each $k \in U$ and each $s \in \mathcal{S}$. Hence,

$$\sum_{k \in s} w_k(s)\, x_k \;=\; \sum_{k \in s} \frac{1}{\pi_k}\,\pi_k \;=\; \sum_{k \in s} 1 \;=\; \left(\begin{array}{c}\text{sample}\\ \text{size of } s\end{array}\right) \;=\; n,$$

since the sampling design has fixed size n. On the other hand,

$$T_x \;=\; \sum_{k \in U} x_k \;=\; \sum_{k \in U} \pi_k \;=\; \sum_{k \in U} E[\, I_k \,] \;=\; E\left[\,\sum_{k \in U} I_k\,\right] \;=\; E\left[\begin{array}{c}\text{sample}\\ \text{size}\end{array}\right] \;=\; n,$$

again since the sample size is fixed and equals n. Therefore, we have, for any $s \in \mathcal{S}$,

$$\sum_{k \in s} w_k(s)\, x_k \;=\; n \;=\; T_x$$

Therefore, the Horvitz-Thompson estimator, under the assumption of fixed sample size, is indeed calibrated with respect to the inclusion probabilities $x : U \longrightarrow \mathbb{R}$, $x_k = \pi_k := \sum_{s \ni k} p(s)$, for each $k \in U$. $\square$

Proposition 3.3

Let $\widehat{\mathbf{T}}_{\mathbf{y};w,x} : \mathcal{S} \longrightarrow \mathbb{R}^m$ be an $\mathbb{R}^m$-valued random variable which is linear in the $\mathbb{R}^m$-valued population parameter $\mathbf{y} : U \longrightarrow \mathbb{R}^m$ and calibrated with respect to the population parameter $x : U \longrightarrow \mathbb{R}$, with $x_k \neq 0$ for each $k \in U$.
Then, the mean squared error matrix of $\widehat{\mathbf{T}}_{\mathbf{y};w,x}$ as an estimator of $\mathbf{T}_{\mathbf{y}}$ is given by:

$$\mathrm{MSE}\Big[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\Big] \;=\; -\frac{1}{2}\sum_{\substack{i,k \in U\\ i \neq k}} a_{ik}\left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right)\cdot\left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right)^T x_i\, x_k \;\in\; \mathbb{R}^{m \times m}, \quad \text{where } \; a_{ik} \;:=\; E[\,(I_i\, w_i - 1)\,(I_k\, w_k - 1)\,].$$

Proof

$$\begin{aligned}
\mathrm{MSE}\Big[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\Big] &= E\left[\left(\widehat{\mathbf{T}}_{\mathbf{y};w,x} - \mathbf{T}_{\mathbf{y}}\right)\cdot\left(\widehat{\mathbf{T}}_{\mathbf{y};w,x} - \mathbf{T}_{\mathbf{y}}\right)^T\,\right] \;=\; E\left[\left(\sum_{i \in U}(I_i w_i - 1)\,\mathbf{y}_i\right)\cdot\left(\sum_{k \in U}(I_k w_k - 1)\,\mathbf{y}_k\right)^T\,\right] \\
&= \sum_{i \in U}\sum_{k \in U} E[\,(I_i w_i - 1)\,(I_k w_k - 1)\,]\cdot\mathbf{y}_i\cdot\mathbf{y}_k^T \;=\; \sum_{k \in U} a_{kk}\cdot\mathbf{y}_k\cdot\mathbf{y}_k^T + \sum_{\substack{i,k \in U\\ i \neq k}} a_{ik}\cdot\mathbf{y}_i\cdot\mathbf{y}_k^T \\
&= \sum_{k \in U} a_{kk}\left(\frac{\mathbf{y}_k\cdot\mathbf{y}_k^T}{x_k^2}\right)x_k^2 + \sum_{\substack{i,k \in U\\ i \neq k}} a_{ik}\left(\frac{\mathbf{y}_i}{x_i}\right)\cdot\left(\frac{\mathbf{y}_k}{x_k}\right)^T x_i\, x_k
\end{aligned}$$

On the other hand,

$$\begin{aligned}
&-\frac{1}{2}\sum_{\substack{i,k \in U\\ i \neq k}} a_{ik}\left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right)\cdot\left(\frac{\mathbf{y}_i}{x_i} - \frac{\mathbf{y}_k}{x_k}\right)^T x_i\, x_k \\
=\;& -\frac{1}{2}\sum_{\substack{i,k \in U\\ i \neq k}} a_{ik}\left[\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T - \left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T - \left(\frac{\mathbf{y}_k}{x_k}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T + \left(\frac{\mathbf{y}_k}{x_k}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T\right] x_i\, x_k \\
=\;& -\frac{1}{2}\sum_{\substack{i,k \in U\\ i \neq k}} a_{ik}\left[\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T + \left(\frac{\mathbf{y}_k}{x_k}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T\right] x_i\, x_k + \sum_{\substack{i,k \in U\\ i \neq k}} a_{ik}\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T x_i\, x_k
\end{aligned}$$

Thus, the proof of the present Proposition will be complete once we show:

$$\underbrace{\sum_{k\in U} a_{kk}\left(\frac{\mathbf{y}_k}{x_k}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T x_k^2}_{\frac{1}{2}\sum\limits_{\substack{i,k\in U\\ i=k}} a_{ik}\left[\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T+\left(\frac{\mathbf{y}_k}{x_k}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T\right]x_i\,x_k} \quad = \quad -\frac{1}{2}\sum_{\substack{i,k\in U\\ i\neq k}} a_{ik}\left[\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T+\left(\frac{\mathbf{y}_k}{x_k}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T\right]x_i\,x_k,$$

which is equivalent to:

$$\sum_{i\in U}\sum_{k\in U} a_{ik}\left[\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T+\left(\frac{\mathbf{y}_k}{x_k}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T\right]x_i\,x_k \;=\; 0. \tag{3.1}$$

Observe that

$$\begin{aligned}
\mathrm{LHS}(3.1) &= \sum_{i\in U}\sum_{k\in U} a_{ik}\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T x_i\,x_k + \sum_{i\in U}\sum_{k\in U} a_{ik}\left(\frac{\mathbf{y}_k}{x_k}\right)\left(\frac{\mathbf{y}_k}{x_k}\right)^T x_i\,x_k \\
&= 2\sum_{i\in U}\sum_{k\in U} a_{ik}\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T x_i\,x_k \;=\; 2\sum_{i\in U} x_i\left(\frac{\mathbf{y}_i}{x_i}\right)\left(\frac{\mathbf{y}_i}{x_i}\right)^T\left(\sum_{k\in U} a_{ik}x_k\right).
\end{aligned}$$

Hence, (3.1) follows once we show

$$\sum_{k\in U} a_{ik}x_k = 0, \quad \text{for each } i\in U. \tag{3.2}$$

Lastly, we now claim that (3.2) follows from the hypothesis that $\widehat{T}_{y;w;x}$ is calibrated with respect to x. Indeed,

$$\begin{aligned}
\sum_{k\in U} a_{ik}x_k &= \sum_{k\in U} E[\,(I_iw_i-1)(I_kw_k-1)\,]\,x_k \;=\; \sum_{k\in U}\left[\sum_{s\in\mathcal{S}} p(s)(I_i(s)w_i(s)-1)(I_k(s)w_k(s)-1)\right]x_k \\
&= \sum_{s\in\mathcal{S}} p(s)\cdot(I_i(s)w_i(s)-1)\cdot\left[\sum_{k\in U}(I_k(s)w_k(s)-1)\cdot x_k\right] \\
&= \sum_{s\in\mathcal{S}} p(s)\cdot(I_i(s)w_i(s)-1)\cdot\left[\underbrace{\left(\sum_{k\in s} w_k(s)\,x_k\right)-T_x}_{0}\right] \\
&= 0
\end{aligned}$$

The proof of the present Proposition is now complete. □

Proposition 3.4 (The Yates-Grundy-Sen Variance Estimator for calibrated linear population total estimators)

Let $p:\mathcal{S}\longrightarrow(0,1]$ be a sampling design each of whose first-order and second-order inclusion probabilities is strictly positively. Let $\widehat{\mathbf{T}}_{\mathbf{y};w,x}:\mathcal{S}\longrightarrow\mathbb{R}^m$ be a random variable which is linear in the population parameter $\mathbf{y}:U\longrightarrow\mathbb{R}^m$ and calibrated with respect to the population parameter $x:U\longrightarrow\mathbb{R}$, with $x_k\neq 0$ for each $k\in U$. Suppose that $\widehat{\mathbf{T}}_{\mathbf{y};w,x}$ is an unbiased estimator for $\mathbf{T}_{\mathbf{y}}:=\sum\limits_{k\in U}\mathbf{y}_k$, for arbitrary $\mathbf{y}$. Then, the following is an unbiased estimator of the variance $\mathrm{Var}\Big[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\Big]$ of $\widehat{\mathbf{T}}_{\mathbf{y};w,x}$: For each $s\in\mathcal{S}$ admissible in the sampling design $p:\mathcal{S}\longrightarrow(0,1]$,

$$\widehat{\mathrm{Var}}\Big[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\Big](s) \;:=\; -\frac{1}{2}\sum_{\substack{i,k\in s\\ i\neq k}}\left(w_i(s)w_k(s)-\frac{1}{\pi_{ik}}\right)\left(\frac{\mathbf{y}_i}{x_i}-\frac{\mathbf{y}_k}{x_k}\right)\cdot\left(\frac{\mathbf{y}_i}{x_i}-\frac{\mathbf{y}_k}{x_k}\right)^T x_i\,x_k$$

Terminology: $\widehat{\mathrm{Var}}\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\right]$ is called the Yates-Grundy-Sen Variance Estimator.

PROOF Since $\widehat{\mathbf{T}}_{\mathbf{y};w,x}$ is an unbiased estimator for $\mathbf{T}_{\mathbf{y}}$ by hypothesis, we have $\mathrm{Var}\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\right] = \mathrm{MSE}\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\right]$. By Proposition 3.3, we thus have:

$$\mathrm{Var}\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\right] \;=\; -\frac{1}{2}\sum_{\substack{i,k\in U\\ i\neq k}} a_{ik}\left(\frac{\mathbf{y}_i}{x_i}-\frac{\mathbf{y}_k}{x_k}\right)^2 x_i\,x_k,\quad \text{where}\;\; a_{ik} \;:=\; E[\,(I_i\,w_i-1)\,(I_k\,w_k-1)\,]\,.$$

On the other hand,

$$E\Big(\,\widehat{\mathrm{Var}}\left[\,\widehat{\mathbf{T}}_{\mathbf{y};w,x}\,\right]\Big) \;=\; -\frac{1}{2}\sum_{\substack{i,k\in U\\ i\neq k}} E\left[\,I_iI_k\left(w_iw_k-\frac{1}{\pi_{ik}}\right)\right]\left(\frac{\mathbf{y}_i}{x_i}-\frac{\mathbf{y}_k}{x_k}\right)\cdot\left(\frac{\mathbf{y}_i}{x_i}-\frac{\mathbf{y}_k}{x_k}\right)^T x_i\,x_k$$

Thus, it remains only to show:

$$a_{ik} \;=\; E\left[\,I_iI_k\left(w_iw_k-\frac{1}{\pi_{ik}}\right)\right].$$

Now,

$$E\left[\,I_iI_k\left(w_iw_k-\frac{1}{\pi_{ik}}\right)\right] \;=\; E[\,I_iI_kw_iw_k\,]-\frac{1}{\pi_{ik}}E[\,I_iI_k\,] \;=\; E[\,I_iI_kw_iw_k\,]-\frac{1}{\pi_{ik}}\,\pi_{ik} \;=\; E[\,I_iI_kw_iw_k\,]-1,$$

and

$$\begin{aligned} a_{ik} &= E[\,(I_i\,w_i-1)\,(I_k\,w_k-1)\,] \;=\; E[\,I_i\,I_k\,w_i\,w_k\,]-E[\,I_i\,w_i\,]-E[\,I_k\,w_k\,]+1 \\ &= E[\,I_i\,I_k\,w_i\,w_k\,]-1-1+1 \;=\; E[\,I_i\,I_k\,w_i\,w_k\,]-1 \\ &= E\left[\,I_iI_k\left(w_iw_k-\frac{1}{\pi_{ik}}\right)\right], \end{aligned}$$

where third last equality follows from Proposition 2.3 and the unbiasedness hypothesis on $\widehat{\mathbf{T}}_{\mathbf{y};w,x}$ as an estimator for $\mathbf{T}_{\mathbf{y}}$. The proof of the present Proposition is now complete. □

4 Unbiased variance estimators for the Horvitz-Thompson Estimator

Let $U=\{1,2,\ldots,N\}$ be a finite population. Let $\mathbf{y}=(y_1,y_2,\ldots,y_m):U\longrightarrow\mathbb{R}^m$ be an $\mathbb{R}^m$-valued function defined on U (commonly called a "population parameter"). We will use the common notation $\mathbf{y}_k$ for $\mathbf{y}(k)$. We wish to estimate $\mathbf{T}_{\mathbf{y}} := \sum_{k\in U}\mathbf{y}_k\in\mathbb{R}^m$ via survey sampling. Let $p:\mathcal{S}\longrightarrow(0,1]$ be our chosen sampling design, where $\mathcal{S}\subseteq\mathcal{P}(U)$ is the set of all possible samples in the design, and $\mathcal{P}(U)$ is the power set of U.

Proposition 4.1

Suppose the first-order and second-order inclusion probabilities of $p:\mathcal{S}\longrightarrow(0,1]$ are all strictly positive, i.e.

$$\pi_k := \sum_{s\ni k}p(s) = \sum_{k\in U}I_k(s)p(s) > 0 \quad\text{and}\quad \pi_{ik} := \sum_{s\ni i,k}p(s) = \sum_{i,k\in U}I_i(s)I_k(s)p(s) > 0,$$

for any $i,k\in U$. Then, the Horvitz-Thompson estimator for $\mathbf{T}_{\mathbf{y}}$ is:

$$\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}(s) \;:=\; \sum_{k\in s}\frac{1}{\pi_k}\,\mathbf{y}_k,$$

and the covariance matrix of the Horvitz-Thompson estimator can be given by:

$$\mathrm{Var}\left[\,\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}\,\right] \;=\; \sum_{i,k\in U}(\pi_{ik}-\pi_i\pi_k)\cdot\left(\frac{\mathbf{y}_i}{\pi_i}\right)\cdot\left(\frac{\mathbf{y}_k}{\pi_k}\right)^T$$

An unbiased estimator for the covariance matrix of the Horvitz-Thompson estimator is given by:

$$\widehat{\mathrm{Var}}\left[\,\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}\,\right](s) \;=\; \sum_{i,k\in s}\left(\frac{\pi_{ik}-\pi_i\pi_k}{\pi_{ik}}\right)\cdot\left(\frac{\mathbf{y}_i}{\pi_i}\right)\cdot\left(\frac{\mathbf{y}_k}{\pi_k}\right)^T, \;\text{ for each } s\in\mathcal{S}.$$

Furthermore, if the sampling design has fixed sample size, then an alternative expression of the covariance matrix of the Horvitz-Thompson estimator is:

$$\mathrm{Var}\left[\,\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}\,\right] \;=\; -\frac{1}{2}\sum_{i,k\in U}\left(\pi_{ik}-\pi_i\pi_k\right)\cdot\left(\frac{\mathbf{y}_i}{\pi_i}-\frac{\mathbf{y}_k}{\pi_k}\right)\cdot\left(\frac{\mathbf{y}_i}{\pi_i}-\frac{\mathbf{y}_k}{\pi_k}\right)^T$$

and the corresponding Yates-Grundy-Sen variance estimator is:

$$\widehat{\mathrm{Var}}^{\mathrm{YGS}}\left[\,\widehat{\mathbf{T}}^{\mathrm{HT}}_{\mathbf{y}}\,\right](s) \;:=\; -\frac{1}{2}\sum_{i,k\in s}\left(\frac{\pi_{ik}-\pi_i\pi_k}{\pi_{ik}}\right)\cdot\left(\frac{\mathbf{y}_i}{\pi_i}-\frac{\mathbf{y}_k}{\pi_k}\right)\cdot\left(\frac{\mathbf{y}_i}{\pi_i}-\frac{\mathbf{y}_k}{\pi_k}\right)^T$$

Proof By Proposition 2.6, for any random variable (a.k.a. estimator) $\widehat{\mathbf{T}}_{\mathbf{y};w}$ linear in the population parameter $\mathbf{y}:\mathcal{S}\longrightarrow\mathbb{R}^m$ with weights $w_k:\mathcal{S}\longrightarrow\mathbb{R}$, $k\in U$, the following

$$\widehat{\mathrm{Var}}\left[\,\widehat{\mathbf{T}}_{y;w}\,\right](s) \;:=\; \sum_{i,k\in s}\frac{\mathrm{Cov}(I_iw_i,I_kw_k)}{\pi_{ik}}\,\mathbf{y}_i\cdot\mathbf{y}_k^T \tag{4.3}$$

always gives an unbiased estimator for the covariance matrix of $\widehat{\mathbf{T}}_{y;w}$. For the Horvitz-Thompson estimator, the weights are $w_k=1/\pi_k$, for each $k\in U$, and the weights are independent of the sample $s\in\mathcal{S}$. Thus, for the Horvitz-Thompson estimator, the right-hand side of equation (4.3) becomes:

$$\begin{aligned}\sum_{i,k\in s}\frac{\mathrm{Cov}(I_iw_i,I_kw_k)}{\pi_{ik}}\,\mathbf{y}_i\cdot\mathbf{y}_k^T &= \sum_{i,k\in s}\frac{\mathrm{Cov}(I_i,I_k)}{\pi_{ik}}\left(\frac{\mathbf{y}_i}{\pi_i}\right)\cdot\left(\frac{\mathbf{y}_k}{\pi_k}\right)^T\\ &= \sum_{i,k\in s}\frac{E(I_iI_k)-E(I_i)E(I_k)}{\pi_{ik}}\left(\frac{\mathbf{y}_i}{\pi_i}\right)\cdot\left(\frac{\mathbf{y}_k}{\pi_k}\right)^T\\ &= \sum_{i,k\in s}\frac{\pi_{ik}-\pi_i\pi_k}{\pi_{ik}}\left(\frac{\mathbf{y}_i}{\pi_i}\right)\cdot\left(\frac{\mathbf{y}_k}{\pi_k}\right)^T,\end{aligned}$$

which coincides with the right-hand side of the equation of the conclusion of the present Proposition. Thus this present Proposition is but a special case of Proposition 2.6, specialized to the Horvitz-Thompson estimator, and the proof is now complete. □

5 Estimation of Domain Totals

6 Conditional inference in finite-population sampling

In this section, we give a justification for making inference conditional on the observed sample size for sampling designs with random sample size.

Observation ("mixture" of experiments) [see [?], p.15.]

Consider a population $\mathcal{U}$ of 1000 units. We wish to estimate the total T_y of a certain population characteristic $\mathbf{y}=(y_1,y_2,\ldots,y_{1000})$. Suppose we use the following two-step sampling scheme:

- Step 1: We first flip a fair coin.
  Define the random variable X by letting $X=1$ if the coin lands heads, and $X=0$ if it lands tails.

- Step 2: If $X = 1$, we select an SRS from $\mathcal{U}$ of size 100. If $X = 0$, we take a census on all of $\mathcal{U}$.

Let $\mathcal{S} \subset \mathcal{P}(\mathcal{U})$ denote the probability space of all possible samples induced by the (two-step) sampling design above. Note that $\mathcal{S} = \mathcal{S}_0 \sqcup \mathcal{S}_1$, where $\mathcal{S}_0 = \{\,\mathcal{U}\,\}$ and $\mathcal{S}_1$ is the set of all subsets of $\mathcal{U}$ of size 100. The sampling design is determined by the following probability distribution on $\mathcal{S}$:

$$P(\mathcal{U}\,) = \frac{1}{2} \quad \text{and} \quad P(\,s\,) = \frac{1}{2\begin{pmatrix}1000\\100\end{pmatrix}}, \quad \text{for each } s \in \mathcal{S}_1.$$

Let $\widehat{T}_y : \mathcal{S} \longrightarrow \mathbb{R}$ denote our chosen estimator for T_y. Then the (unconditional) probability distribution of $\widehat{T}_y$ can be "decomposed" as follows:

$$\begin{aligned}
P\left(\,\widehat{T}_y = t \,\middle|\, \mathbf{y}\,\right) &= P\left(\,\widehat{T}_y = t,\, X = 0 \,\middle|\, \mathbf{y}\,\right) + P\left(\,\widehat{T}_y = t,\, X = 1 \,\middle|\, \mathbf{y}\,\right) \\
&= P\left(\,\widehat{T}_y = t \,\middle|\, X = 0,\, \mathbf{y}\,\right) \cdot P(\,X = 0 \,|\, \mathbf{y}\,) + P\left(\,\widehat{T}_y = t \,\middle|\, X = 1,\, \mathbf{y}\,\right) \cdot P(\,X = 1 \,|\, \mathbf{y}\,) \\
&= P\left(\,\widehat{T}_y = t \,\middle|\, X = 0,\, \mathbf{y}\,\right) \cdot P(\,X = 0\,) + P\left(\,\widehat{T}_y = t \,\middle|\, X = 1,\, \mathbf{y}\,\right) \cdot P(\,X = 1\,),
\end{aligned}$$

where the last equality follows because the distribution of X is independent of $\mathbf{y}$. Suppose the observation we make consists of $\left(\,\widehat{T}_y\,,\,X\,\right)$. The unconditional probability distribution of $\widehat{T}_y$, given by $P\left(\,\widehat{T}_y = t \,\middle|\, \mathbf{y}\,\right)$ above, describes of course the randomness of the estimator $\widehat{T}_y$ as induced by both the randomness of the sample $s \in \mathcal{S} = \mathcal{S}_0 \sqcup \mathcal{S}_1$ as well as that of X (the outcome of the coin flip in Step 1). Now, suppose we have indeed carried out the sampling procedure and have obtained an observation of $\left(\,\widehat{T}_y\,,\,X\,\right)$. Suppose it happened that $X = 1$. Hence, we know that the estimate $\widehat{T}_y(s)$ we actually obtained was generated from an SRS of size 100 (rather than a census). Note also that the probability distribution of X is independent of $\mathbf{y}$ and the observation of X gives no information about $\mathbf{y}$. One school of thought therefore argues that downstream inferences about $\mathbf{y}$ should be carried out using the conditional probability $P\left(\,\widehat{T}_y = t \,\middle|\, X = 1\,,\,\mathbf{y}\,\right)$, rather than the unconditional probability $P\left(\,\widehat{T}_y = t \,\middle|\, \mathbf{y}\,\right)$. In other words, in the present example, as far as making inferences about $\mathbf{y}$ is concerned, only the randomness in Step 2 is relevant, and the randomness in Step 1 (i.e. the randomness of X, the outcome of the coin flip) is irrelevant to any inference about $\mathbf{y}$. Consequently randomness of X "should" be removed in any inference procedure for $\mathbf{y}$, and this is achieved by conditioning on the observed value of X. $\square$

Conditioning on obtained sample size for sample designs with random sample size

Suppose $\mathcal{U}$ is a finite population. We wish to estimate the total $T_y = \sum_{i\in\mathcal{U}} y_i$ of a population characteristic $\mathbf{y} : \mathcal{U} \longrightarrow \mathbb{R}$, using a sample design $p : \mathcal{S} \longrightarrow [\,0\,,1\,]$ and a estimator $\widehat{T} : \mathcal{S} \longrightarrow \mathbb{R}$. We make the assumption that the sampling design p is independent of $\mathbf{y}$. Let $N : \mathcal{S} \longrightarrow \mathbb{N} \cup \{\,0\,\}$ be the random variable of sample size, i.e. $N(s) =$ number of elements in s, for each possible sample $s \in \mathcal{S}$. Then,

$$\begin{aligned}
P\left(\,\widehat{T} = t \,\middle|\, \mathbf{y}\,\right) &= \sum_n P\left(\,\widehat{T} = t,\, N = n \,\middle|\, \mathbf{y}\,\right) \\
&= \sum_n P\left(\,\widehat{T} = t \,\middle|\, N = n,\, \mathbf{y}\,\right) \cdot P(\,N = n \,|\, \mathbf{y}\,) \\
&= \sum_n P\left(\,\widehat{T} = t \,\middle|\, N = n,\, \mathbf{y}\,\right) \cdot P(\,N = n\,),
\end{aligned}$$

where the last equality follows from the assumed independence of the probability distribution $p : \mathcal{S} \longrightarrow [\,0\,,\,1\,]$ (hence that of N) from $\mathbf{y}$. The key observation to make now is that: Although the actual sampling procedure operationally may or may not have been a two-step procedure, the independence of p from $\mathbf{y}$ makes it probabilistically equivalent to a two-step procedure, as shown by the above decomposition of $P\left(\,\widehat{T} = t \,\middle|\, \mathbf{y}\,\right)$ — Step (1): randomly select a sample size $N = n$ according to the distribution $P(N = n)$, and then Step (2): randomly select a sample s of size n chosen in Step (1) according to the distribution $P(\,s \,|\, N = n\,)$. By the statistical reasoning explained in the preceding observation, it follows that post-sampling inference about $\mathbf{y}$ should be made based on the conditional distribution $P\left(\,\widehat{T}=t\,\middle|\,N=n\,,\,\mathbf{y}\right)$, rather than the unconditional distribution $P\left(\,\widehat{T}=t\,\middle|\,\mathbf{y}\right)$. This is because the sampling scheme is probabilistically equivalent to a two-step procedure, with the probability distribution of the first step (choosing a sample size) independent of the parameters of interest (T_y), and thus only the probability distribution of the second step (choosing a sample of the size chosen in first step) should be used to make inference about T_y. □

Caution

In more formal parlance, the random variable $N : \mathcal{S} \longrightarrow \mathbb{N} \cup \{\,0\,\}$ is ancillary to the parameter $\mathbf{y}$. Thus, conditioning on sample size, for finite-population sampling schemes with random sample size, *partially* conforms to the **Conditionality Principle**, which states that statistical inference about a parameter should be made conditioned on observed values of statistics ancillary to that parameter. The conformance is only partial due to the (obvious) fact that it is the sample s itself which is ancillary to the parameter of interest $\mathbf{y}$, not just its sample size $N(s)$. Thus, full conformance to the Conditionality Principle would require inference about $\mathbf{y}$ be made conditioned on the observed sample s itself (rather than its size $N(s)$). However, if we did condition on the obtained sample s itself, the domain of the estimator $\widehat{T}$ would be restricted to the singleton $\{\,s\,\}$, and $\widehat{T}$ could then attain only one value under conditioning on s, and no randomization-based (i.e. design-based) inference — apart from the observed value of $\widehat{T}(s)$ — could be made any longer.