

1 The Hájek Central Limit Theorem for Simple Random Sampling without Replacement

Suppose we are given a sequence of finite populations, on each of which is defined an \mathbb{R} -valued population characteristic. Suppose on each of the finite populations, we use SRSWOR (of fixed sample size) to select a sample, observe the values of the corresponding population characteristics on the selected elements, and use the Horvitz-Thompson estimator for the population mean. We seek to determine a necessary and sufficient condition for the (associated sequence of) “standardized deviations from the mean” of the Horvitz-Thompson estimator for population mean to converge in distribution to the standard Gaussian distribution.

Theorem 1.1

Suppose we have the following:

- Let $\{U_\nu\}_{\nu \in \mathbb{N}}$ be a sequence of finite populations, and $N_\nu = |U_\nu|$ be the population size of U_ν . Let the elements of U_ν be indexed by $1, 2, 3, \dots, N_\nu$.
- For each $\nu \in \mathbb{N}$, let $y^{(\nu)} : U_\nu \rightarrow \mathbb{R}$ be an \mathbb{R} -valued population characteristic. For each $i \in U_\nu$, let $y_i^{(\nu)}$ denote $y^{(\nu)}(i)$, the value of $y^{(\nu)}$ evaluated at the i^{th} element of U_ν .
- For each $\nu \in \mathbb{N}$, let $n_\nu \in \{1, 2, 3, \dots, N_\nu\}$ be given, and let \mathcal{S}_ν be the set of all n_ν -element subsets of U_ν . Let \mathcal{S}_ν be endowed with the uniform probability function, namely

$$P(s) = \frac{1}{\binom{N_\nu}{n_\nu}}, \text{ for each } s \in \mathcal{S}_\nu.$$

- For each $\nu \in \mathbb{N}$, let $\widehat{Y}_\nu : \mathcal{S}_\nu \rightarrow \mathbb{R}$ be the random variable defined as follows:

$$\widehat{Y}_\nu(s) := \frac{1}{n_\nu} \sum_{i \in s} y_i^{(\nu)}, \text{ for each } s \in \mathcal{S}_\nu$$

Let

$$\mu_\nu := E[\widehat{Y}_\nu] = \frac{1}{N_\nu} \sum_{i \in U_\nu} y_i^{(\nu)} \quad \text{and} \quad \sigma_\nu^2 := \text{Var}[\widehat{Y}_\nu] = \left(1 - \frac{n_\nu}{N_\nu}\right) \frac{S_\nu^2}{n_\nu},$$

where

$$S_\nu^2 := \frac{1}{N_\nu - 1} \sum_{i \in U_\nu} \left(y_i^{(\nu)} - \mu_\nu\right)^2$$

- For each $\nu \in \mathbb{N}$ and each $\delta > 0$ define:

$$U_\nu(\delta) := \left\{ i \in U_\nu \mid |y_i^{(\nu)} - \mu_\nu| > \delta \sqrt{\sigma_\nu^2} \right\} \subset U_\nu.$$

Suppose $n_\nu \rightarrow \infty$ and $N_\nu - n_\nu \rightarrow \infty$. Then,

$$\lim_{\nu \rightarrow \infty} P\left\{ s \in \mathcal{S}_\nu \mid \frac{\widehat{Y}_\nu(s) - \mu_\nu}{\sqrt{\sigma_\nu^2}} < x \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt$$

if and only if

$$\lim_{\nu \rightarrow \infty} \frac{\sum_{i \in U_\nu(\delta)} \left(y_i^{(\nu)} - \mu_\nu\right)^2}{\sum_{i \in U_\nu} \left(y_i^{(\nu)} - \mu_\nu\right)^2} = 0, \text{ for every } \delta > 0.$$

References

- [1] HÁJEK, J. Limiting distributions in simple random sampling from a finite population. *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* 5 (1960), 361–374.
- [2] LOHR, S. L. *Sampling: Design and Analysis*, first ed. Duxbury Press, 1999.
- [3] VALLIANT, R., DORFMAN, A. H., AND ROYALL, R. M. *Finite Population Sampling and Inference*, first ed. John-Wiley & Sons, 2000.