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* $1,2,3,\ldots,N_\nu$.
- *For each* $\nu\in\mathbb{N}$, *let* $y^{(\nu)} : U_\nu \longrightarrow \mathbb{R}$ *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* $n_\nu \in \{1,2,3,\ldots,N_\nu\}$ *be given, and let* $\mathcal{S}_\nu$ *be the set of all* n_ν*-element subsets of* U_ν. *Let* $\mathcal{S}_\nu$ *be endowed with the uniform probability function, namely*
$$P(s) \;=\; \frac{1}{\left(\begin{array}{c} N_\nu \\ n_\nu \end{array}\right)}, \;\; \text{for each } s\in\mathcal{S}_\nu .$$
- *For each* $\nu\in\mathbb{N}$, *let* $\widehat{\overline{Y}}_\nu : \mathcal{S}_\nu \longrightarrow \mathbb{R}$ *be the random variable defined as follows:*
$$\widehat{\overline{Y}}_\nu(s) \;:=\; \frac{1}{n_\nu}\sum_{i\in s} y_i^{(\nu)}, \;\; \text{for each } s\in\mathcal{S}_\nu$$
Let
$$\mu_\nu \;:=\; E\!\left[\,\widehat{\overline{Y}}_\nu\,\right] \;=\; \frac{1}{N_\nu}\sum_{i\in U_\nu} y_i^{(\nu)} \quad\text{and}\quad \sigma_\nu^2 \;:=\; \mathrm{Var}\!\left[\,\widehat{\overline{Y}}_\nu\,\right] \;=\; \left(1-\frac{n_\nu}{N_\nu}\right)\frac{S_\nu^2}{n_\nu},$$
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:*
$$U_\nu(\delta) \;:=\; \left\{\, i\in U_\nu \;\middle|\; |y_i^{(\nu)} - \mu_\nu| > \delta\,\sqrt{\sigma_\nu^2}\,\right\} \;\subset\; U_\nu .$$

Suppose $n_\nu \longrightarrow \infty$ *and* $N_\nu - n_\nu \longrightarrow \infty$. *Then*
$$\lim_{\nu\to\infty} P\left\{\, s\in\mathcal{S}_\nu \;\middle|\; \frac{\widehat{\overline{Y}}_\nu(s) - \mu_\nu}{\sqrt{\sigma_\nu^2}} < x \,\right\} \;=\; \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t$$
if and only if
$$\lim_{\nu\to\infty}\frac{\sum_{i\in U_\nu(\delta)}\left(y_i^{(\nu)}-\mu_\nu\right)^2}{\sum_{i\in U_\nu}\left(y_i^{(\nu)}-\mu_\nu\right)^2} \;=\; 0, \;\; \text{for every } \delta>0.$$

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.