# The Bahadur Representation of Sample Quantiles in General Unequal Probability Sampling Designs

**Hitoshi MOTOYAMA**[1]

[1] Delft University of Technology, Netherlands;
Aoyama Gakuin University, Japan;
The Institute of Statistical Mathematics, Japan
e-mail: H.M.Motoyama@tudelft.nl; hitoshi@aoyamagakuin.jp.

## Abstract

This presentation establishes the Bahadur representation of sample quantiles in general unequal probability sampling designs.

The Bahadur representation, proposed by and named after Bahadur(1966), is a useful linear representation of quantiles that has been extended to $M$-estimators and time-series data. Moreover, in a sample survey framework, the Bahadur representation of quantiles plays an important role. Francisco and Francisco and Fuller(1991) and Shao(1994) developed the Bahadur representation for stratified cluster sampling and stratified multistage sampling, respectively. However, it has been considered difficult to develop Bahadur representations in the general sampling framework. In fact, Wu and Thompson(2020), citing Chen and Wu(2002), state that "With complex survey data, however, Bahadur representations are difficult to establish even under very restrictive regularity conditions."

In this presentation, we establish the Bahadur representation of sample quantiles in general unequal probability sampling designs under the fairly general and mild conditions similar to Boistard *et al.*(2017).

Consider a sequence of finite population $\mathcal{U}^N$, of size $N = 1, 2, \ldots$. With each population, we associate a set of indices $U_N = \{1, 2, \ldots, N\}$. Furthermore, for each index $i \in U_N$, we have a number $x_i \in \mathbb{R}$ which is the values of the variable of interest. For all $N = 1, 2, \ldots$, let a sequence of sample $\mathcal{S}_N = \{s : s \subset U_N\}$ be the collection of subsets of $U_N$. We define the sample size $n = n(N)$ for the sample $\mathcal{S}_N$ as the cardinal number of $\mathcal{S}_N$. Moreover, we define the first-order inclusion probability $\pi_i$ of the unit $i$ as $\pi_i = P(i \in s)$.

Define $F_N$ and $F_n$ be the population distribution function and the Hájek estimator for $F_N$, respectively,

$$F_N(x) = \frac{1}{N}\sum_{i=1}^{N} \mathrm{I}(x_i \le x) \quad \text{and} \quad F_n(x) = \frac{\sum_{i\in s} I(x_i \le x)/\pi_i}{\sum_{i\in s} 1/\pi_i}, \tag{1}$$

where $\mathrm{I}(A)$ is the indicator function defined by 1 if $A$ is true and 0 otherwise.

Under some regularity conditions, we establish the Bahadur representation of sample quantiles:

$$\hat{\theta}_n = \theta_N + \frac{p - F_n(\theta_N)}{f_N(\theta_N)} + o_p(n^{-1/2}), \tag{2}$$

where $\hat{\theta}_n = \inf\{x : F_n(x) \ge p\}$ are the $p$-th sample quantiles and $\theta_N = \inf\{x : F_N(x) \ge p\}$ are the $p$-th population quantiles.

**Keywords:** Bahadur representation; Quantiles; Finite population; Survey Sampling.

## References


Boistard, Hélène and Lopuhaä, Hendrik P. and Ruiz-Gazen, Anne. (2017). Functional central limit theorems for single-stage sampling designs. *The Annals of Statistics*, **45**, 1728–1758.

Chen, Jiahua and Wu, Changbao. (2002). Estimation of distribution function and quantiles using the model-calibrated pseudo empirical likelihood method. *Statistica Sinica*, **12**, 1223–1239.

Francisco, Carol A. and Fuller, Wayne A. (1991). Quantile estimation with a complex survey design. *The Annals of Statistics*, **19**, 454–469.

Shao, Jun. (1994). *L*-statistics in complex survey problems. *The Annals of Statistics*, **22**, 946–967.

Wu, Changbao and Thompson, Mary E. (2020). *Sampling theory and practice*. Springer.