# Note on a simple derivation of the asymptotic normality of a sample quantile from a finite population

**元山 斉** リスク解析戦略研究センター 融合プロジェクト特任研究員

## Notations

| | |
|---|---|
| $N$ | total number of units in finite population |
| $x_1, \ldots, x_N$ | characteristic of pupulation |
| $n$ | total number of units in sample |
| $X_1, \ldots, X_n$ | characteristic of sample |

$$F_{N_k}(x) = \frac{1}{N_k}\sum_{i=1}^{N_k} I_{(-\infty,x]}(x_i) \quad \text{Distribution Funstion(Population)}$$

$$F_{n_k}(x) = \frac{1}{n_k}\sum_{i=1}^{n_k} I_{(-\infty,x]}(X_i) \quad \text{Empirical Distribution Funstion(Sample)}$$

$$\theta_k = F_{N_k}^{-1}(p) = \inf\{x : F_{N_k}(x) \geqq p\} \quad \text{Population } p\text{-th quantile}$$

$$\hat{\theta}_k = F_{n_k}^{-1}(p) = \inf\{x : F_{n_k}(x) \geqq p\} \quad \text{Sample } p\text{-th quantile}$$

## Sampling Design

Simple random samples $X_1, \ldots, X_n$ of size $n$ are chosen without replacement from the populations $x_1, \ldots, x_N$.

## Asymptotics in Finite Population

$\{\mathcal{P}_k\}_1^\infty$: Population Sequence
$N_k \to \infty$ as $k \to \infty$

$$\begin{array}{llllll} x_{11} & \cdot & x_{1N_1} & & & \text{Population } \mathcal{P}_1 \\ x_{21} & \cdot & \cdot & x_{2N_2} & & \\ \cdot & \cdot & \cdot & & & \\ x_{k1} & x_{k2} & & & x_{kN_k} & \text{Population } \mathcal{P}_k \\ \cdot & \cdot & \cdot & \cdot & \cdot\ \cdot\ \cdot & \end{array}$$

$\{\mathcal{P}_k\}_1^\infty$: Sample from Population $\mathcal{P}_k$
$n_k \to \infty$ as $k \to \infty$

$$\begin{array}{llllll} X_{11}, & X_{12}, & \cdots, & X_{1n_k} & & \text{Sample from } \mathcal{P}_1 \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{k1} & X_{k2} & & & X_{kn_k} & \text{Sample from } \mathcal{P}_k \\ \vdots & \vdots & & \vdots & \vdots & \end{array}$$

## Asymptotic Normality

(A1) The sequence $\{\theta_k\}$ is bounded.
(A2) There is a sequence of functions $\{f_k\}$ such that

$$\lim_{k\to\infty}\left[\frac{F_{N_k}(\theta_k+\delta_k)-p}{\delta_k} - f_k(\theta_k)\right] = 0,$$

for any sequence $\{\delta_k\}$ of order $\sim O(n_k^{-1/2})$ and

$$0 < \inf_k f_k(\theta_k) \leqq \sup_k f_k(\theta_k) < \infty.$$

(A3) $0 \leqq \frac{n_k}{N_k} < \epsilon < 1$

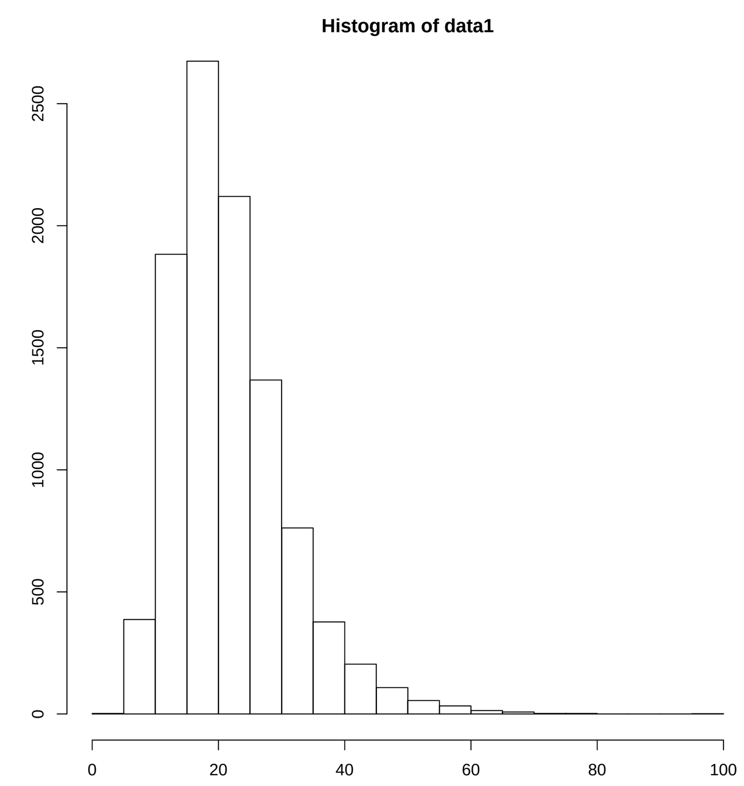
**Theorem 1.** *Under the assumptions* (A1), (A2) *and* (A3), *we have*

$$P((\hat{\theta}_k - \theta_k)/a_k \leqq x) \xrightarrow{d} N(0,\ 1) \quad n_k,\ N_k - n_k \to \infty$$

*where* $a_k = ((N_k - n_k)p(1-p)/f_k^2(\theta_k)n_k(N_k-1))^{1/2}$.

## Simulation

Pseudo population: lognormal random number $(3,\ 0.4)$




Confidence Intervals of Median(Sampling Fraction10%)

| | 90% | 95% | 99% |
|---|---|---|---|
| Sample Size500 Population Size5,000 | 0.90636 | 0.95292 | 0.99251 |
| Sample Size600 Population Size6,000 | 0.89872 | 0.94755 | 0.99129 |
| Sample Size700 Population Size7,000 | 0.90283 | 0.95165 | 0.99248 |
| Sample Size800 Population Size8,000 | 0.89735 | 0.95276 | 0.99129 |
| Sample Size900 Population Size9,000 | 0.89257 | 0.9477 | 0.98978 |
| Sample Size1,000 Population Size10,000 | 0.89818 | 0.94831 | 0.99091 |

Confidence Intervals of Median(Sampling Fraction30%)

| | 90% | 95% | 99% |
|---|---|---|---|
| Sample Size1,500 Population Size5,000 | 0.88282 | 0.93482 | 0.99091 |
| Sample Size1,800 Population Size6,000 | 0.87778 | 0.93947 | 0.98971 |
| Sample Size2,100 Population Size7,000 | 0.90417 | 0.93855 | 0.98976 |
| Sample Size2,400 Population Size8,000 | 0.92029 | 0.95009 | 0.98853 |
| Sample Size2,700 Population Size9,000 | 0.91278 | 0.94908 | 0.98678 |
| Sample Size3,000 Population Size10,000 | 0.90292 | 0.94974 | 0.98661 |

### References

Francisco and Fuller (1991). *The Annals of Statistics* **19**, 454-469.

Motoyama and Takahashi (2008). *Journal of The Japan Statistical Society*, **38**, 475-504.

Shao (1994). *The Annals of Statistics*, **22**, 946-67.

Wretman (1978). *Scandinavian Journal of Statistics*, **5**, 123-124.