A New Family of Quantiles Estimators with P-Auxiliary Information in Successive Sampling

Nursel Koyuncu[1]
Hacettepe University, Ankara, Turkey, nkoyuncu@hacettepe.edu.tr
Hulya Cingi
Hacettepe University, Ankara, Turkey, hcingi@hacettepe.edu.tr

In this paper, we have suggested a new family of quantiles estimators using p-auxiliary information in successive sampling. We have found the bias and mean square error (MSE) expressions to compare the efficiency over other quantiles estimators in the sampling literature. To illustrate the efficiency of suggested estimator with the existing estimators we use a real data. This data set uses 40 stick-figure cartoons to investigate the structure of cartoons covered a range of social situations. Most of the cartoons portrayed negative acts representing physical and verbal aggression and social exclusion. We use bullying as study variable and injury, threat, annoyance as auxiliary variables. It has been shown that suggested family of estimator is more efficient than existing estimators.

Keywords: auxiliary information, mean square error, ratio estimator, successive sampling

1. Introduction

Successive sampling has been used when the value of study character of a finite population is subject to change over time. Using information of single occasion will provide the characteristics for the given occasion only and can't give any information of the characteristic over different occasions. Many authors such as Rueda et. Al. (2008), Singh et.al. (2007), Martinez-Miranda et. al. (2005) etc. suggested quantile estimation under successive sampling. In this study we have suggested a new family of quantiles estimators using p-auxiliary information in successive sampling

Consider a finite population U with size N, which is assumed to retain its composition over two (or more) time periods. Assume a sample, s', with the size n' which is drawn on the previous occasion. On the current occasion a sub-sample s_m, with size m, is taken from the previously selected n' units, and u=n−m units are replaced by the new units selected independently of the matched portion. This last sample is called the unmatched sample s_u. Let $\chi = m/n$ be the matched fraction and $s = s_m \cup s_u$.

Let y denote the survey variable, with values $y_1, \ldots, y_N$ for the N population elements.

In the previous occasion exists p-auxiliary variables, $x_1, \ldots, x_p$ which are used to build the proposed multivariate ratio estimator. The finite population distribution

---

[1] Corresponding Author

function of y is given by $F_Y(t)=\frac{\sum_{i\in U}\Delta(t-y_i)}{N}$ with $\Delta(\alpha)=1$ if $\alpha\geq 0$ and $\Delta(\alpha)=0$ otherwise. The corresponding finite population β-quantile, $Q_Y(\beta)$, of y is

$$Q_Y(\beta)=F_Y^{-1}(\beta)=\inf\{t:F_Y(t)\geq\beta\},$$

where $F_Y^{-1}(t)$ is the inverse function and $0<\beta<1$. Estimation of quantiles is given by

$$\hat{Q}_Y(\beta)=\hat{F}_Y^{-1}(\beta)=\inf\{t:\hat{F}_Y(t)\geq\beta\}$$

To estimate the β-quantile based upon the sample s of the current occasion is

$$\hat{Q}_{Yn}(\beta)=\hat{F}_{Yn}^{-1}(\beta)=\inf\{t:\hat{F}_{Yn}(t)\geq\beta\}$$

where $\hat{F}_{Yn}(t)=\frac{\sum_{i\in U}\Delta(t-y_i)}{n}$ is the sample estimator for the finite population distribution function, which coincides with the Horvitz–Thompson type estimator under simple random sampling. Similarly, let $\hat{Q}_{Xi}(\beta)$ $(i=1,\ldots,p)$ be the sample quantiles of order β based on s'. Denote by $\hat{Q}_{Xim}(\beta)$ and $\hat{Q}_{Ym}(\beta)$ the sample quantiles of the matched sample on the auxiliary and principle variables, and $\hat{Q}_{Yu}(\beta)$ the sample quantile based upon the unmatched sample of the current occasion. (See Rueda et al. (2008))

Rueda et al. (2008) proposed a class of estimator is given by

$$\hat{Q}_{Ym}^{MR}(\beta)=\sum_{1\leq i\leq p} w_i\frac{\hat{Q}_{Ym}(\beta)}{\hat{Q}_{Xim}(\beta)}\hat{Q}_{Xi}(\beta)=\sum_{1\leq i\leq p} w_i\hat{Q}_{Yrim}(\beta) \tag{1.1}$$

The mean square error of $\hat{Q}_{Ym}^{MR}(\beta)$ and optimum value of w are given by respectively

$$V\left(\hat{Q}_{Ym}^{MR}(\beta)\right)=\sum_{1\leq i\leq p} w_i^2 V\left(\hat{Q}_{Yrim}(\beta)\right)+2\sum_{i<j} w_i w_j \mathrm{Cov}\left(\hat{Q}_{Yrim}(\beta),\hat{Q}_{Yrjm}(\beta)\right) \tag{1.2}$$

$$w_{opt} = \frac{B^{-1}e}{e'B^{-1}e} \tag{1.3}$$

$$V_{\min}\left(\hat{Q}_{Ym}^{MR}(\beta)\right) = \frac{1}{e'B^{-1}e} \tag{1.4}$$

Then they defined a composite estimator as

$$\hat{Q}_Y(\beta) = W\hat{Q}_{Ymopt}^{MR}(\beta) + (1-W)\hat{Q}_{Yu}(\beta) \tag{1.5}$$

$$W_{opt} = \frac{V\left(\hat{Q}_{Yu}(\beta)\right)}{V\left(\hat{Q}_{Yu}(\beta)\right) + V\left(\hat{Q}_{Ymopt}^{MR}(\beta)\right)} \tag{1.6}$$

$$V\left(\hat{Q}_{Yopt}(\beta)\right) = W_{opt}^2 V\left(\hat{Q}_{Ymopt}^{MR}(\beta)\right) + (1-W_{opt})^2 V\left(\hat{Q}_{Yu}(\beta)\right) \tag{1.7}$$

2. Suggested Estimator

We consider following estimator for population quantiles

$$\hat{Q}_P(\beta) = W\hat{Q}_{Pm}^{(l)} + (1-W)\hat{Q}_{Pu} \tag{2.1}$$

where

$$\hat{Q}_{Pm}^{(l)} = \hat{Q}_{y(m)}(\beta) + b\left\{\hat{Q}_x(\beta) - \hat{Q}_{x(m)}(\beta)\right\} \tag{2.2}$$

regression estimator with

$$b = \frac{\hat{f}_x\left(\hat{Q}_{x(m)}(\beta)\right)}{\hat{f}_y\left(\hat{Q}_{y(m)}(\beta)\right)}\left\{\frac{\hat{P}_{xy(m)}}{\beta(1-\beta)} - 1\right\} \tag{2.3}$$

and

$$\hat{Q}_{Pu} = \hat{Q}_{Yu}\left(\frac{\phi Q_Z + \lambda}{\phi\hat{Q}_{Zu} + \lambda}\right)^{\eta} \tag{2.4}$$

ratio type estimator where ϕ and λ any suitable auxiliary variable of Z. The mean square error of $\hat{Q}_{Pm}^{(l)}$ and $\hat{Q}_{Pu}$ are given, respectively,

$$MSE\left(\hat{Q}_{Pu}\right) = Q_Y(\beta)^2\left(\eta^2\omega^2\frac{N-u}{Nu}\frac{\beta(1-\beta)}{\left(Q_Z(\beta)\,f_Z\left(Q_Z(\beta)\right)\right)^2} + \frac{N-u}{Nu}\frac{\beta(1-\beta)}{\left(Q_y(\beta)\,f_y\left(Q_y(\beta)\right)\right)^2}\right. \tag{2.5}$$

$$\left. -2\eta\omega\frac{N-u}{Nu}\frac{\left(P_{11}(y,z) - \beta(1-\beta)\right)}{Q_Z(\beta)Q_Y(\beta)\,f_Z\left(Q_Z(\beta)\right)f_Y\left(Q_Y(\beta)\right)}\right)$$

where $\omega = \dfrac{\phi Q_Z(\beta)}{\phi Q_Z + \lambda}$ and

$$\eta_{opt} = \frac{\left(P_{11}(y,z) - \beta(1-\beta)\right)}{\omega\beta(1-\beta)} \frac{Q_Z(\beta)\, f_Z\left(Q_Z(\beta)\right)}{Q_Y(\beta)\, f_Y\left(Q_Y(\beta)\right)} \tag{2.6}$$

Substituting (2.6) in (2.5) we have

$$MSE\left(\hat{Q}_{Pu}\right) = \frac{N-u}{Nu} \frac{\beta(1-\beta)}{\left\{f_Y\left(Q_Y(\beta)\right)\right\}^2} \left(1 - \left\{\frac{P_{11}(y,z)}{\beta(1-\beta)} - 1\right\}^2\right) \tag{2.7}$$

$$MSE\left(\hat{Q}_{Pm}^{(l)}\right) = \sum_{1\le i\le p} w_i^2 MSE\left(\hat{Q}_{Yim}^{(l)}(\beta)\right) + 2\sum_{i<j} w_i w_j Cov\left(\hat{Q}_{Yim}^{(l)}(\beta), \hat{Q}_{Yjm}^{(l)}(\beta)\right) \tag{2.8}$$

We can rewrite (2.8)

$$MSE\left(\hat{Q}_{Pm}^{(l)}\right) = w'Bw \tag{2.9}$$

where $\mathrm{w} = \left(\mathrm{w}_1, \ldots, \mathrm{w}_\mathrm{p}\right)'$, $\mathrm{B} = \left(\mathrm{b}_{\mathrm{ij}}\right)$, $b_{ij} = Cov\left(\hat{Q}_{Yim}^{(l)}(\beta), \hat{Q}_{Yjm}^{(l)}(\beta)\right)$ $(\mathrm{i,j} = 1, \ldots, \mathrm{p})$.

$$\mathrm{w}_{\mathrm{opt}} = \frac{\mathrm{B}^{-1}\mathrm{e}}{\mathrm{e}'\mathrm{B}^{-1}\mathrm{e}} \tag{2.10}$$

Substituting (2.10) in (2.8) we have

$$MSE_{\min}\left(\hat{Q}_{Pm}^{(l)}\right) = \frac{1}{e'B^{-1}e} \tag{2.11}$$

Using (2.7) and (2.11), the optimum mean square error of (2.1) is given by

$$V\left(\hat{Q}_P(\beta)\right) = W^2 V\left(\hat{Q}_{Pm}^{(l)}\right) + (1-W)^2 V\left(\hat{Q}_{Pu}\right)$$

(2.12)

$$W_{opt} = \frac{V\left(\hat{Q}_{Pu}\right)}{V\left(\hat{Q}_{Pm}^{(l)}\right) + V\left(\hat{Q}_{Pu}\right)} \tag{2.13}$$

$$V\left(\hat{Q}_P(\beta)\right) = \frac{V\left(\hat{Q}_{Pu}\right) V\left(\hat{Q}_{Pm}^{(l)}\right)}{V\left(\hat{Q}_{Pm}^{(l)}\right) + V\left(\hat{Q}_{Pu}\right)}$$

(2.14)

3. Numerical Example

To illustrate the efficiency of suggested estimator with the existing estimators we use a real data. This data set uses 40 stick-figure cartoons to investigate the structure of cartoons covered a range of social situations. Most of the cartoons portrayed negative acts representing physical and verbal aggression and social exclusion. We use bullying as study variable and injury, threat, annoyance as auxiliary variables. The mean square error value of suggested estimator and existing estimators are calculated and given in Table1. From Table1 we have observed that suggested estimator is more efficient than existing estimator for 0.10, 0.25 and 0.50 quantiles.

References

Martinez-Miranda, M.D., Rueda-Garcia, M., Cebrian, A.A., Roman Montoya, Y., Gonzalez-Aguilera, S. (2005)"Quantile estimation under successive sampling," *Computational Statistics*, 20, 385-399.

Rueda, M., Munoz, J., Arcos A. (2008)"Successive sampling to estimate quantiles with p-auxiliary variables," *Quality & Quantity*, 42,427–443.

Singh, H.P., Tailor R., Singh S., Kim J.M. (2007)"Quantile estimation in successive sampling," *Journal of the Korean Statistical Society*, 36,4, 543-556.

Table 1: Mean Square Error

$\hat{Q}_Y(0.10)$	
Estimator	**MSE**
$\hat{Q}_{Yu}$	225.564
$\hat{Q}_{Ym}^{MR}(\beta)$	176.364
$\hat{Q}_Y(\beta)$	98.976
$\hat{Q}_P(\beta)$	83.897
$\hat{Q}_Y(0.25)$	
Estimator	**MSE**
$\hat{Q}_{Yu}$	33.785
$\hat{Q}_{Ym}^{MR}(\beta)$	61.343
$\hat{Q}_Y(\beta)$	21.786
$\hat{Q}_P(\beta)$	12.673
$\hat{Q}_Y(0.50)$	
Estimator	**MSE**
$\hat{Q}_{Yu}$	33.444
$\hat{Q}_{Ym}^{MR}(\beta)$	17.838
$\hat{Q}_Y(\beta)$	11.633
$\hat{Q}_P(\beta)$	2.945