

A note on Woodruff confidence intervals for quantiles

Randy R. Sitter^{a,*}, Changbao Wu^b

^aDepartment of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada V5A 1S6

^bDepartment of Statistics and Actuarial Science, University of Waterloo, Waterloo, ONT, Canada N2L 3G1

Received May 1999; received in revised form November 1999

Abstract

Woodruff (1952) proposed a simple confidence interval for quantiles in complex surveys based upon inverting the usual confidence intervals for the distribution function. In the moderate to extreme tail regions of the distribution function the usual confidence interval performs poorly for moderate sample size. In this paper we demonstrate that despite this fact, the Woodruff intervals based upon inverting these badly behaved intervals perform very well. We go on to explain this rather surprising fact. © 2001 Elsevier Science B.V. All rights reserved

Keywords: Complex survey; Distribution function; Percentiles; Stratified

1. Introduction

In this article we will consider inference for quantiles in surveys. If we let $F(t) = N^{-1} \sum_{i=1}^N I_{[y_i \leq t]}$ be the finite population distribution function evaluated at t , then the p -quantile can be written as $\xi_p = F^{-1}(p) = \inf\{t: F(t) \geq p\}$, where N is the number of units in the finite population. This is then estimated by $\hat{\xi}_p = \hat{F}^{-1}(p) = \inf\{t: \hat{F}(t) \geq p\}$, where $\hat{F}(t) = \sum_{i \in s} w_i I_{[y_i \leq t]}$, s is a sample of units from the finite population under some complex sampling design and $\{w_i: i \in s\}$ are the survey weights.

The conventional $1 - \alpha$ normal confidence interval for a distribution function $F(t)$ at fixed t is based on $\hat{F}(t)$ and its estimated variance $\hat{v} = v[\hat{F}(t)]$ obtained from the general sampling design, and is constructed as follows:

$$(\hat{F}(t) - z_{\alpha/2} \hat{v}^{1/2}, \hat{F}(t) + z_{\alpha/2} \hat{v}^{1/2}), \quad (1.1)$$

where $z_{\alpha/2}$ is the $1 - \alpha/2$ quantile from $N(0, 1)$. The interval in (1.1) relies on the asymptotic normality of $\hat{F}(t)$,

$$\frac{\hat{F}(t) - F(t)}{\hat{v}^{1/2}} \xrightarrow{d} N(0, 1) \quad (1.2)$$

* Corresponding author. Tel.: +1-604-291-3334; fax: +1-604-291-4947.

E-mail address: sitter@stat.sfu.ca (R.R. Sitter).

which can be justified in many situations (see, for example, Francisco and Fuller, 1991, Theorem 2). Woodruff (1952) proposed a large sample confidence interval for ξ_p for general sampling designs, $(\hat{\xi}_p^L, \hat{\xi}_p^U)$, where

$$\hat{\xi}_p^L = \inf\{t: \hat{F}(t) \geq p - z_{\alpha/2}v^{1/2}[\hat{F}(\hat{\xi}_p)]\},$$

$$\hat{\xi}_p^U = \inf\{t: \hat{F}(t) \geq p + z_{\alpha/2}v^{1/2}[\hat{F}(\hat{\xi}_p)]\}$$

and

$$P(\hat{\xi}_p^L \leq \xi_p \leq \hat{\xi}_p^U) \doteq 1 - \alpha. \quad (1.3)$$

Approximation (1.3) is closely related to (1.1) and relies on the asymptotic normality of $\hat{F}(t)$. Francisco and Fuller (1991) have justified (1.3) under certain assumptions.

In the cases of large or small values of p and moderate sample size, the sampling distribution of $\hat{F}(\xi_p)$ is usually not symmetric. The smaller (larger) the value of p , the farther the departure from symmetry. The consequence is that the normal confidence interval (1.1) performs poorly: the right and left tail-errors are very unbalanced and the coverage probability differs from the nominal value. One might believe that because of this the Woodruff confidence interval for large or small quantiles should not work and should not be recommended in practice in such situations. However, a Monte Carlo study reported in Section 2 shows that Woodruff intervals for large or small quantiles work far better than intuition might suggest. The reason for this interesting phenomenon is discussed in Section 3, with suggestions and concluding remarks given in Section 4.

2. A simulation

The finite population we use here was modified from the one used by Francisco and Fuller (1991) in a Monte Carlo study to investigate the performance of some statistics including the Woodruff interval. Their study did not include large or small quantiles and they did not vary the sample size.

We generate the finite population from a superpopulation which is a mixture of 10 lognormal distributions: $F(t) = \sum_{i=1}^{10} w_i F_i(t)$, with weights 0.08, 0.08, 0.10, 0.10, 0.12, 0.12, 0.14, 0.10, 0.10, 0.06 and means and standard deviations of the distributions of the log of the random variables given by (1.5, 0.3), (2.0, 0.4), (2.1, 0.4), (3.0, 0.6), (2.5, 0.5), (1.8, 0.4), (1.6, 0.3), (1.8, 0.4), (3.0, 0.6), (3.8, 0.8), respectively. The finite population has ten strata with stratum sizes proportional to the above weights. The observations in each of the strata were generated as simple random samples from the corresponding lognormal distributions.

The simulation was first conducted with finite population size $N = 500$. For each simulation, a finite population was first generated and then a stratified simple random sample of size $n = 50$ was selected from the finite population, with 5 elements selected from each of the 10 strata. Normal confidence intervals for $F(t)$ at $t = \xi_{0.05}, \xi_{0.10}, \xi_{0.25}, \xi_{0.50}, \xi_{0.75}, \xi_{0.90}, \xi_{0.95}$ were constructed using (1.1), and Woodruff confidence intervals for these same quantiles were obtained using (1.3). This process was repeated 10,000 times. The simulation was then re-conducted with finite populations of sizes $N = 1000, 2000$ and sample sizes of $n = 100, 200, 10$ and 20 elements from each of the 10 strata, respectively. The Monte Carlo error in this simulation study is negligible.

Table 1 reports the estimated coverage probabilities (CP) of 90% confidence intervals and estimated left and right tail errors (L, U: nominal value is 5% for both) for $F(t)$ at various quantiles and for the quantiles

Table 1
 Estimated coverage probabilities and tail errors and standardized length of 90% confidence intervals from stratified simple random sampling

		$p = 0.05$	$p = 0.10$	$p = 0.25$	$p = 0.50$	$p = 0.75$	$p = 0.90$	$p = 0.95$
Normal Confidence Interval for Distribution Function $F(t)$ at $t = \xi_p$								
$N = 500$	CP	82.5	85.7	88.4	88.8	86.9	86.5	80.7
$n = 50$	L	1.7	2.8	5.0	5.9	7.5	9.5	16.1
	U	15.8	11.5	6.6	5.3	5.6	4.0	3.2
	SL	0.93	0.97	0.99	1.00	1.00	0.96	0.94
$N = 1000$	CP	86.0	87.6	89.1	88.8	89.3	88.3	85.9
$n = 100$	L	1.9	3.2	5.3	6.1	5.9	8.4	11.3
	U	12.1	9.2	5.6	5.1	4.8	3.3	2.8
	SL	0.98	1.00	1.00	1.00	1.00	0.98	0.97
$N = 2000$	CP	88.1	88.7	89.6	89.8	89.3	89.1	88.1
$n = 200$	L	2.5	3.6	4.8	5.1	5.2	6.8	8.7
	U	9.4	7.7	5.6	5.1	5.5	4.1	3.2
	SL	0.99	1.01	1.01	1.00	1.01	1.00	1.00
Woodruff Confidence Interval for Quantile ξ_p								
		$\xi_{0.05}$	$\xi_{0.10}$	$\xi_{0.25}$	$\xi_{0.50}$	$\xi_{0.75}$	$\xi_{0.90}$	$\xi_{0.95}$
$N = 500$	CP	89.5	92.3	90.5	89.6	90.4	88.8	85.2
$n = 50$	L	7.0	3.5	5.1	6.0	5.3	6.5	8.4
	U	3.5	4.2	4.4	4.4	4.3	4.7	6.4
	SL	0.98	1.12	1.03	1.00	1.01	1.09	1.56
$N = 1000$	CP	92.7	91.2	90.2	89.6	90.4	89.3	85.9
$n = 100$	L	3.0	4.1	4.7	5.5	4.9	5.7	8.3
	U	4.3	4.7	5.1	4.9	4.7	4.9	5.8
	SL	1.17	1.05	1.00	0.98	1.01	1.04	0.99
$N = 2000$	CP	91.8	90.4	90.1	90.2	89.9	90.1	88.8
$n = 200$	L	3.6	4.8	5.0	5.5	5.7	5.5	6.5
	U	4.6	4.8	4.9	4.3	4.4	4.4	4.7
	SL	1.05	1.03	1.02	1.01	1.01	1.04	1.02

as well (all entries are presented as percentages). The standardized length (SL) of confidence intervals were estimated by dividing the average length of all 10,000 simulated intervals by $2z_{0.05}(\text{MSE})^{1/2}$, where MSE was estimated by another 10,000 simulations.

For fixed sample size, normal confidence intervals for $F(\xi_p)$ performed well at $p = 0.50$ and reasonably well for $p = 0.25, 0.75$, that is, coverage probabilities were very close to 90% nominal and one-sided tail errors were well balanced. However, for larger or smaller p , normal confidence intervals for $F(t)$ performed poorly: coverage probabilities became lower and one-sided tail errors were no longer balanced. The smaller (larger) the value of p , the more severe the unbalance of the two tails.

As for Woodruff intervals, they performed surprisingly well for all quantiles. The unbalanced tail-error behavior for large and small quantiles observed in normal confidence intervals for $F(t)$ mysteriously disappear or are much improved and coverage probabilities are close to the nominal value and more conservative. Meanwhile, the good performance of normal confidence intervals for middle-range quantiles were well passed to the Woodruff intervals.

When we increase the sample size from 50 to 100 and 200, the quality of both normal confidence intervals for $F(t)$ and Woodruff confidence intervals for ξ_p improve, as they should due to asymptotic properties. However, the above phenomenon is still evident and the improvement for Woodruff intervals seems much quicker than that of normal intervals.

Table 2
Estimated coverage probabilities and tail errors of 90% idealized Woodruff intervals

		$\xi_{0.05}$	$\xi_{0.10}$	$\xi_{0.25}$	$\xi_{0.50}$	$\xi_{0.75}$	$\xi_{0.90}$	$\xi_{0.95}$
$N = 1000$	CP	86.3	87.8	89.3	89.1	89.6	89.0	86.8
$n = 100$	L	12.1	9.2	5.6	5.1	4.8	3.3	2.8
	U	1.6	3.0	5.1	5.8	5.6	7.7	10.4

The above phenomenon was not case-specific. We conducted more simulation studies, including simple random sampling from a finite population generated from a single lognormal distribution and another “large-scale” stratified simple random sampling ($N = 14,000$) using the well-known Hansen, Madow & Tepping synthetic population (Hansen et al., 1983). We also included 95% confidence intervals in each case. The results were similar and are not presented.

3. Investigation of the phenomenon

To illustrate what is causing this phenomenon, first let us re-do the simulations using $v[\hat{F}(\xi_p)]$ in the Woodruff interval instead of $v[\hat{F}(\hat{\xi}_p)]$. We will refer to this as the idealized Woodruff interval. We can do this in the simulation since we know ξ_p . In practice, however, we could not since ξ_p is unknown. The results appear in Table 2. To save space we present only the case $N = 1000$ and $n = 100$. The other cases are qualitatively similar (Wu, 1999). We see that the coverage probabilities and tail errors are almost identical to that of confidence intervals for the distribution function (Table 1), except that the left and right tail errors are switched. So, for some reason replacing ξ_p with $\hat{\xi}_p$ improves things.

Given these and the simulation results of the previous section, one is compelled to further investigate the formulation of Woodruff intervals. We consider simple random sampling where $\hat{F}(t) = 1/n \sum_{i \in s} I_{[y_i \leq t]}$,

$$v[\hat{F}(t)] = \frac{1-f}{n-1} \hat{F}(t)[1 - \hat{F}(t)] \tag{3.1}$$

and f is the finite population correction factor. Consider that

$$\hat{F}(\hat{\xi}_p^L) \doteq p - z_{\alpha/2} v^{1/2} [\hat{F}(\hat{\xi}_p)] \quad \text{and} \quad \hat{F}(\hat{\xi}_p^U) \doteq p + z_{\alpha/2} v^{1/2} [\hat{F}(\hat{\xi}_p)],$$

noting that the equalities are approximate only as a result of the discreteness of the empirical distribution function. The variance, $v[\hat{F}(t)]$, when viewed as a function of $\hat{F}(t)$, is monotone increasing for $\hat{F}(t) \in [0.0, 0.5]$ and monotone decreasing for $\hat{F}(t) \in (0.5, 1.0]$, with maximum absolute derivative at $\hat{F}(t) = 0.0$ and 1.0 and minimum absolute derivative at $\hat{F}(t) = 0.5$.

Let us first consider why using $\hat{\xi}_p$ instead of ξ_p decreases the unsatisfactorily high left tail error when p is small ($\ll 0.5$). Assuming the sample size fixed, this follows from

$$\begin{aligned} P\{F(\xi_p) \geq \hat{F}(\xi_p) + z_{\alpha/2} v^{1/2} [\hat{F}(\xi_p)]\} &\geq P\{\hat{F}(\xi_p) \leq p - z_{\alpha/2} v^{1/2} [\hat{F}(\xi_p)]\} \\ &\geq P\{\hat{F}(\xi_p) \leq p - z_{\alpha/2} v^{1/2} [\hat{F}(\hat{\xi}_p)]\} \\ &\doteq P\{\hat{F}(\xi_p) \leq \hat{F}(\hat{\xi}_p^L)\} \\ &\geq P(\xi_p \leq \hat{\xi}_p^L). \end{aligned} \tag{3.2}$$

The key step in the above is the second inequality, which follows from $v[\hat{F}(\xi_p)] < v[\hat{F}(\hat{\xi}_p)]$. This can be justified as follows. It is clear that if $\hat{F}(\xi_p) \leq p - z_{\alpha/2} v^{1/2} [\hat{F}(\xi_p)]$, then $p > \hat{F}(\xi_p)$. Since, by definition, $p \leq \hat{F}(\hat{\xi}_p)$, we get $\hat{F}(\xi_p) < p \leq \hat{F}(\hat{\xi}_p)$, which in turn, by (3.1), implies $v[\hat{F}(\xi_p)] < v[\hat{F}(\hat{\xi}_p)]$.

One can similarly argue that

$$P\{F(\xi_p) < \hat{F}(\xi_p) - z_{\alpha/2}v^{1/2}[\hat{F}(\xi_p)]\} \leq P(\xi_p > \hat{\xi}_p^U). \tag{3.3}$$

This gives some explanation for why the tail error rates are more balanced in the Woodruff interval than one might expect. This alone would be a desirable property even if the interval coverage rate remained unchanged. However, the coverage rate of the Woodruff intervals appear to be slightly higher (i.e. more conservative) and closer to the nominal, as well. This does not follow directly from (3.2) and (3.3). To understand this consider that $\hat{F}(\xi_p) = n^{-1} \sum_{i \in S} I_{[y_i \leq \xi_p]}$ can be viewed as the standard estimate of the finite population proportion $F(\xi_p) = N^{-1} \sum_{i=1}^N I_{[y_i \leq \xi_p]}$. If $p \ll 0.5$, the distribution of $\hat{F}(\xi_p)$ will be highly skewed and thus $\hat{F}(\xi_p)$ will more often be less than its mean, $F(\xi_p)$, than not. But, if $\hat{F}(\xi_p) < F(\xi_p)$, then $\xi_p < \hat{\xi}_p$ which implies $v[\hat{F}(\xi_p)] < v[\hat{F}(\hat{\xi}_p)]$, and consequently the interval $(\hat{F}(\xi_p) \pm z_{\alpha/2}v^{1/2}[\hat{F}(\hat{\xi}_p)])$ will be longer than the interval $(\hat{F}(\xi_p) \pm z_{\alpha/2}v^{1/2}[\hat{F}(\xi_p)])$. Noting that

$$\begin{aligned} &P\{\hat{F}(\xi_p) - z_{\alpha/2}v^{1/2}[\hat{F}(\hat{\xi}_p)] < F(\xi_p) < \hat{F}(\xi_p) + z_{\alpha/2}v^{1/2}[\hat{F}(\hat{\xi}_p)]\} \\ &\quad \doteq P\{p - z_{\alpha/2}v^{1/2}[\hat{F}(\hat{\xi}_p)] < \hat{F}(\xi_p) < p + z_{\alpha/2}v^{1/2}[\hat{F}(\hat{\xi}_p)]\} \\ &\quad \doteq P\{\hat{F}(\hat{\xi}_p^L) < \hat{F}(\xi_p) < \hat{F}(\hat{\xi}_p^U)\} \\ &\quad \leq P(\hat{\xi}_p^L < \xi_p < \hat{\xi}_p^U), \end{aligned}$$

this gives an explanation for the observed behaviour.

To summarize, the large right tail error of the normal confidence interval for $F(\xi_p)$ is switched to a smaller left tail error for the Woodruff confidence interval on ξ_p , while the small left tail error on $F(\xi_p)$ is “inverted” to a larger right tail error on ξ_p . In addition, the coverage probability on $F(\xi_p)$ is inflated a bit on ξ_p . The result is a better confidence interval for ξ_p . Similar arguments hold for $p \gg 0.5$. For general sampling designs, as long as the changing pattern of $v[\hat{F}(t)]$ over $\hat{F}(t)$ is similar to that of (3.1), the justification will be identical.

4. Concluding remarks

Woodruff confidence intervals for quantiles are basically an inversion of normal confidence intervals for the distribution function, with only one modification: we replace the unknown quantile ξ_p in $v[\hat{F}(\xi_p)]$ by its estimate, $\hat{\xi}_p$. Although this is a common practice in statistics, sometimes we lose precision or efficiency with this replacement, especially when the sample size is not very large. For Woodruff intervals, this replacement not only works but improves things: it keeps the good performance of the normal confidence interval of distribution functions for the middle-range quantiles and adjusts its ill behavior for large and small quantiles in the desirable direction.

Given the known poor performance of the interval for $F(t)$ given in (1.1) when t is large or small, one might be tempted to use a better interval before applying Woodruff’s method. Wu (1999) used a logit transformation method which drastically improved the performance of the intervals for $F(t)$. However, when one then uses this interval in Woodruff’s method, similar arguments as those given in Section 3 suggest the Woodruff interval will perform badly. This was verified in a simulation (Wu, 1999).

The high coverage probability of Woodruff intervals for small and large quantiles is partially due to the inflated length of the interval. In fact, according to a weak version of the Bahadur representation of the sample

quantile (Francisco and Fuller, 1991; Rao et al., 1990),

$$\hat{\xi}_p = \xi_p - [f(\xi_p)]^{-1}[\hat{F}(\xi_p) - F(\xi_p)] + o_p(n^{-1/2})$$

where $f(t)$ is the derivative of the limiting function of $F(t)$ as $N \rightarrow \infty$. We see that the sampling distribution of $\hat{\xi}_p$ is closely related to that of $\hat{F}(\xi_p)$. As we pointed out earlier, when sample size is small or moderate and p is small or large, the sampling distribution of $\hat{F}(\xi_p)$ is not symmetric. This is also true for $\hat{\xi}_p$. Therefore, the confidence interval for ξ_p with balanced two tail errors will be longer than the symmetric normal interval, which implies that the SL of better-behaved Woodruff intervals will be larger than 1. As sample size increases, the sampling distributions of both $\hat{F}(\xi_p)$ and $\hat{\xi}_p$ tend to be more symmetric and consequently, the SL of these intervals will approach 1.

There exist more elaborate methods for constructing confidence intervals for quantiles, for example, the test inversion procedure (Francisco and Fuller, 1991) or possibly a bootstrap interval (Sitter, 1992). However, Woodruff intervals are simple and easy to implement and thus are popular with practitioners. The good performance for large and small quantiles demonstrated here makes them even more appealing.

Acknowledgements

This research was supported by grants from the Natural Sciences and Engineering Council of Canada. We thank the referee for helpful comments and suggestions.

References

- Francisco, C.A., Fuller, W.A., 1991. Quantiles estimation with a complex survey design. *Ann. Statist.* 19, 454–469.
- Hansen, M.H., Madow, W.G., Tepping, B.J., 1983. An evaluation of model-dependent and probability-sampling inferences in sample surveys. *J. Amer. Statist. Assoc.* 78, 776–793.
- Rao, J.N.K., Kovar, J.G., Mantel, H.J., 1990. On estimating distribution functions and quantiles from survey data using auxiliary information. *Biometrika* 77, 365–375.
- Sitter, R.R., 1992. Comparing three bootstrap methods for survey data. *Can. J. Statist.* 20, 135–154.
- Woodruff, R.S., 1952. Confidence intervals for medians and other position measures. *J. Amer. Statist. Assoc.* 47, 635–646.
- Wu, C., 1999. The Effective Use of Complete Auxiliary Information From Survey Data. Ph.D. Dissertation, Simon Fraser University, Burnaby, BC, Canada.