

SAMPLING-SKEWED BIOLOGICAL POPULATIONS: BEHAVIOR OF CONFIDENCE INTERVALS FOR THE POPULATION TOTAL

TIMOTHY G. GREGOIRE^{1,3} AND OLIVER SCHABENBERGER²

¹Department of Forestry, Virginia Polytechnic Institute and State University,
Blacksburg, Virginia 24061 USA

²Department of Crop and Soil Sciences, Department of Statistics and Probability,
Michigan State University, East Lansing, Michigan 48824-1325 USA

Abstract. Four populations were repeatedly sampled with the aim of examining interval estimation of the cumulative amount, T , of some characteristic that has a positively skewed distribution, a feature of many biological populations. With samples of various sizes, the empirical sampling distribution of the simple expansion estimator was remarkably symmetric and approximately normal, as was that of the customary ratio-of-means estimator. While the nominal coverage rate of $(1 - \alpha)100\%$ intervals based on the Student's t distribution was nearly achieved in all cases, there was a substantially greater tendency for such intervals to fail from below rather than above. This behavior is attributed to the strong positive correlation between the estimator of the population total and the corresponding estimator of its variance when sampling from a finite, and perforce nonnormal, population.

Key words: asymmetry; failure rates; normality; simple random sampling.

INTRODUCTION AND BACKGROUND

Owing to their size and the expense of complete enumeration, biological populations often are sampled with an aim towards estimating the cumulative amount, say T , of some characteristic. Moreover, many biological populations exhibit positive skewness, especially when the characteristic of interest is some expression of size of an organism. The biological relevance of this distributional measure has provoked a modicum of discussion in the ecological literature (cf. Weiner and Solbrig 1984, Weiner and Thomas 1986, Knox et al. 1989). Notwithstanding its biological interpretation, or lack thereof, skewness affects estimation of population parameters, as documented in a rich literature dating at least back to Neyman and Pearson (1928). The particular concern of the present paper is the effect of skewness on the coverage of design-based confidence intervals following simple random sampling without replacement from a finite population. (In accordance with traditional sampling theory, as elucidated by Cochran (1977) and other prominent texts on sampling, we adopt a design basis for inference. Thus, the population is regarded as fixed, not random, and the properties of an estimator are deduced from its distribution under the probability sampling design. This sampling distribution likewise serves as the reference distribution for interval estimation. In contrast, model-based inference relies on

a reference distribution implied by an assumed model of the population.)

Cochran (1977) touched briefly on the distortion of interval statements caused by population skewness, but failed to elaborate on its magnitude or offer any explanation as to the cause of this phenomenon. He cautioned that a two-sided 95% confidence interval for the population mean value will fall below the mean with a relative frequency $>2.5\%$, and conversely it will fall above the mean with a relative frequency $<2.5\%$. The actual coverage will be $<95\%$. He suggested minimum sample sizes of $25G_1^2$ to more nearly achieve the nominal coverage of 95% confidence intervals, where G_1 is the Fisher skewness coefficient. Dalén (1986) suggested a slightly different sample size formula.

Johnson (1978) tried an alternative tactic to correct "bias and skewness effects due to the skewness of the nonnormal population distribution." The correction to the usual confidence interval based on Student's t distribution is to add a term involving the third sample moment.

More recent strategies that have been proposed to counter population skewness include that of Fuller (1991), who took a model-based, or superpopulation, approach to dealing with the ill effects caused by a few "very large observations." By fitting a presumed Weibull distribution to the right tail of the sample distribution, he derived an estimator of the population mean that has minimum mean square error. While Fuller's estimator may be attractive to researchers who find superpopulation inference acceptable, no guidance was provided on confidence interval construction or coverage. Winsorization is another approach that has been suggested recently by Rivest (1994) and Rivest and

Manuscript received 27 January 1998; revised 26 April 1998; accepted 1 May 1998; final version received 26 May 1998.

³ Present address: School of Forestry and Environmental Studies, Yale University, 360 Prospect Street, New Haven, Connecticut 06511 USA. E-mail: Timothy.Gregoire@Yale.edu

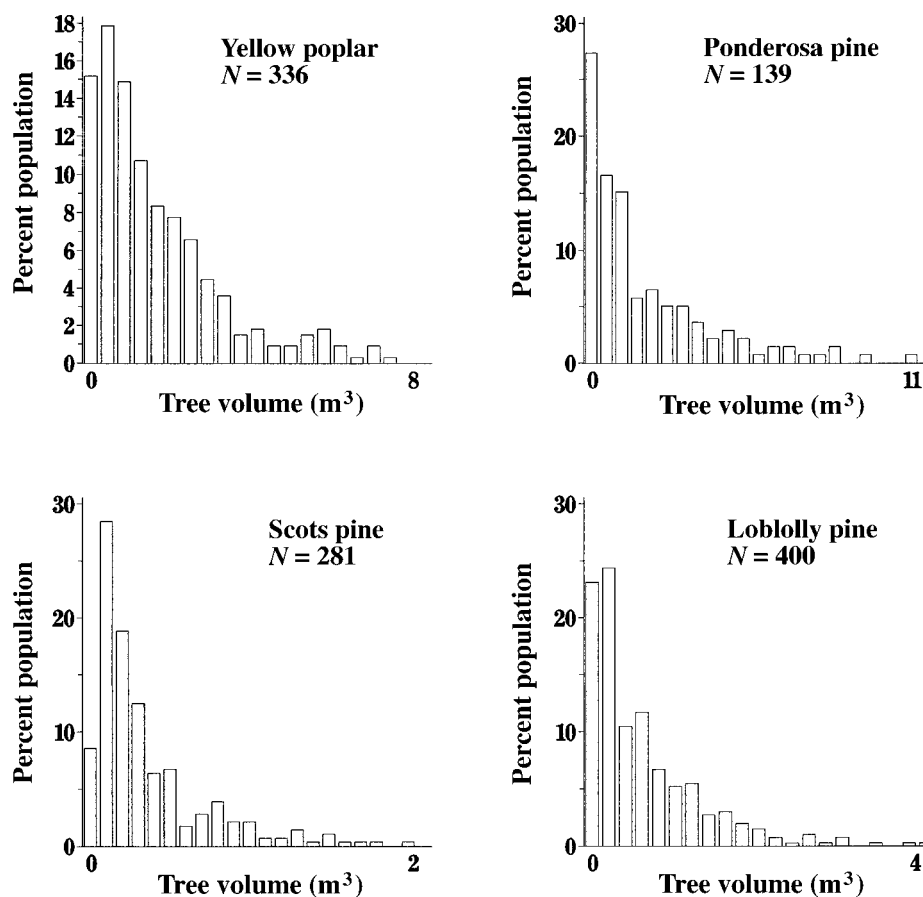


FIG. 1. Relative frequency distributions of tree volumes in four populations that were sampled by simple random sampling without replacement.

Hurtubise (1995) to ameliorate the influence of “outliers that are much larger than most sample values.” Winsorizing is a technique whereby data values are replaced by cutoff values. We speculate that in consideration of the expense and effort to collect data, some ecological researchers would be reluctant to replace a portion of the observed sample data with subjectively selected cutoff values.

We are concerned with conventional interval estimation following sampling from moderately skewed populations, as exhibited in Fig. 1. Skewness in these populations is not due to the presence of isolated extreme values, as per the above authors, but rather results from an approximately exponential decrease in relative frequency with increasing size. While these four populations are of bole volumes of trees, we believe that similarly shaped distributions are encountered commonly in ecological research. We examine confidence-interval coverage for the populations shown in Fig. 1 with samples of size $n = 20, 40,$ and $25G^2$.

We document a largely unheralded aspect of interval estimation with the design-based framework of inference. The positive skewness of the population imparts

a positive correlation to the simple expansion estimator of T , say \hat{T} , and the customary estimator of its variance. It is this, we posit, that explains the behavior noted by Cochran (1977). This artifactual correlation, due jointly to the skewness and finiteness of the parent population, is not widely recognized or frequently mentioned. No sampling texts refer to this correlation. Texts in mathematical statistics teach that the sample mean and variance are independent random variables when sampling from a Gaussian (normally distributed) population. Independence does not hold when sampling from non-normal, finite populations, and a finite-population version of the central limit theorem holds only asymptotically as sample size, n , and parent population size, N , jointly become infinitely large.

There has been isolated mention of this phenomenon. Campbell and Meyer (1978) make passing reference to the correlation between \hat{T} and estimators of variance in their simulation study of the unconditional coverage of t -based confidence intervals for two-stage survey data from a stratified population, but fail to pursue the matter. Royall and Cumberland (1985) studied the conditional coverage properties of finite population con-

TABLE 1. Populations of tree volumes used in Monte Carlo simulations.

Population	N^\dagger	\bar{D}_q^\ddagger (cm)	T_v (m ³)	σ_v (m ³)	cv § (%)	$G_{i }$	ρ^\P
Yellow poplar	336	24.5	516.9	1.54	100	1.5	0.99
Ponderosa pine	139	41.7	249.2	2.13	119	1.8	0.96
Scots pine	281	15.8	96.9	0.36	104	1.9	0.97
Loblolly pine	400	18.2	239.7	0.45	111	1.9	0.95

† Population size.
 ‡ Quadratic mean diameter.
 § Coefficient of variation among tree volumes.
 $||$ Skewness coefficient of tree volumes.
 ¶ Linear correlation coefficient of tree volume and basal area.

confidence intervals and concluded that the nonnormality due to skewness has a strong effect on the distribution of the standardized error of both the traditional ratio (-of-means) and regression estimators.

Our results extend and directly reinforce those of Royall and Cumberland, and we place them in a context relevant to ecological and biological researchers. We show that even when the conventional confidence interval based on the t distribution nearly achieves its nominal coverage, the frequency with which it fails to cover T is exceedingly asymmetric. Intervals are much more likely to fail from below than from above.

Inasmuch as the expansion estimator commonly associated with simple random sampling is an inverse-probability weighted estimator, it is a special case of the Horvitz–Thompson (HT) estimator that is commonly associated with unequal probability sampling. In other unrelated work reported in the ecological statistics literature, Overton and Stehman’s (1994: Table 3) results with the Horvitz–Thompson estimator reveal an asymmetry of coverage following unequal probability sampling, too. We did not duplicate their effort. However, because of the prevalence of ratio estimation, we did examine the performance of confidence intervals based on the ratio estimator and several alternative estimators of its sampling variance that have been proposed. The same pattern characterized coverage by intervals based on ratio estimators.

STUDY POPULATIONS

The four populations of tree bole volumes shown in Fig. 1 were used in the Monte Carlo sampling. The pertinent characteristics of these populations are summarized in Table 1. Although the sizes of these populations are unrealistically small, their distributional characteristics are typical of those encountered in forestry and ecological applications; hence their relevance to this investigation.

Let v_i denote the bole volume of the i th tree bole in a population, and let b_i denote its corresponding basal area, which is the customary term to describe the cross-sectional area of the bole at a height 1.37 m above-ground. Bole diameter at 1.37 m is easily measured on standing trees, thus basal area is a readily computed measure of tree size. Moreover, because basal area is

directly related to foliar surface area and is very highly correlated with bole biomass, it is a measure of enormous importance in forest ecology and silviculture. Of pertinence to this study, basal area is the auxiliary covariate when estimating aggregate bole volume of a population with the ratio estimator.

STUDY DESIGN AND ESTIMATORS

We repeatedly sampled the study populations, described above, to generate empirical sampling distributions of the HT and ratio estimators of T , as well as various estimators of standard error. These parent populations varied in skewness, coefficient of variation, and size, N . For each combination of population and estimator of T , 25 000 samples were drawn of size $n = 20, 40, \text{ and } 25G_1^2$. The approximate normality of both estimators was judged qualitatively by means of normal probability plots, the coverage of nominal standard normal (z -based) confidence intervals and t -based confidence intervals was evaluated, and the failure rates and patterns of both types of intervals were examined.

In the sequel we follow fairly standard notation, $i \in U$ and $i \in S$, to indicate the range of summation over all elements in the population and sample, respectively. The Bebbington (1975) algorithm for drawing a simple random sample without replacement was coded in the SAS programming language using its RANUNI function initiated with a different eight-digit integer seed for each Monte Carlo simulation. With data from each sample the expansion estimator (\hat{T}_v) and the ratio-of-means estimator (\hat{T}_R) of $T_v = \sum_{i \in U} v_i = N\mu_v$ were evaluated as follows:

$$\hat{T}_v = \frac{N}{n} \sum_{i \in S} v_i \tag{1}$$

$$\hat{T}_R = \hat{R}T_B \tag{2}$$

where $\hat{R} = \hat{T}_v / \hat{T}_B$,

$$\hat{T}_B = \frac{N}{n} \sum_{i \in S} b_i \tag{3}$$

and $T_B = \sum_{i \in U} b_i = N\mu_B$. (The mean values were μ_v and μ_B , respectively.) These commonly used estimators are fully explained in Cochran (1977).

The design-based variance of \hat{T}_v is

$$V(\hat{T}_v) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \sigma_v^2 \tag{4}$$

where $\sigma_v^2 = (N - 1)^{-1} \sum_{i \in U} (v_i - \mu_v)^2$. For the purpose of this simulation study σ_v^2 was computable; hence, $V(\hat{T}_v)$ could be evaluated and used in z -based confidence intervals of the form $\hat{T}_v \pm z_{(1-\alpha/2)} \sqrt{V(\hat{T}_v)}$. In practice, such intervals are not feasible, as $V(\hat{T}_v)$ requires information about the entire population. Nonetheless, their coverage constitutes a benchmark for comparison in this study.

A design-unbiased estimator of $V(\hat{T}_V)$ is

$$\hat{V}(\hat{T}_V) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \hat{\sigma}_V^2 \tag{5}$$

where $\hat{\sigma}_V^2 = (n - 1)^{-1} \sum_{i \in S} (v_i - \bar{v})^2$, and $\bar{v} = \hat{T}_V/N$.

To a first-order Taylor series approximation, the variance of \hat{T}_R is

$$V(\hat{T}_R) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \sigma_R^2 \tag{6}$$

where $\sigma_R^2 = (N - 1)^{-1} \sum_{i \in U} (v_i - Rb_i)^2$ and $R = T_V/T_B$. A consistent estimator of $V(\hat{T}_R)$ is

$$\hat{V}(\hat{T}_R) = N^2 \left(\frac{1}{n} - \frac{1}{N} \right) \hat{\sigma}_R^2 \tag{7}$$

where $\hat{\sigma}_R^2 = (n - 1)^{-1} \sum_{i \in S} (v_i - \hat{R}b_i)^2$. An alternative consistent estimator of $V(\hat{T}_R)$ is

$$\hat{V}_a(\hat{T}_R) = \left(\frac{T_B}{\hat{T}_B} \right)^2 \hat{V}(\hat{T}_R) \tag{8}$$

which apparently was first suggested by Rao and Rao (1971). A number of studies since Rao and Rao (1971) have indicated that $\hat{V}_a(\hat{T}_R)$ is a more accurate estimator of $V(\hat{T}_R)$ than $\hat{V}(\hat{T}_R)$, yet the latter is still more commonly used.

Let $\hat{T}_{V(-i)}$ and $\hat{T}_{B(-i)}$ denote the HT estimators of T_V and T_B omitting the i th sample datum, and let $\hat{R}_{(-i)} = \hat{T}_{V(-i)}/\hat{T}_{B(-i)}$. The delete-one jackknife version of the ratio estimator of T_V is

$$\hat{T}_J = \hat{R}_J T_B, \tag{9}$$

where $\hat{R}_J = n^{-1} \sum_{i \in S} \hat{R}_i$ and \hat{R}_i is the i th pseudo-value $\hat{R}_i = n\hat{R} - (n - 1)\hat{R}_{(-i)}$. The corresponding jackknife estimator of the variance of \hat{T}_J is

$$\hat{V}(\hat{T}_J) = T_B^2 (n(n - 1))^{-1} \sum_{i=1}^n (\hat{R}_i - \hat{R}_J)^2 \tag{10}$$

which has also been suggested as an estimator of a bias-corrected version of T_R (cf. Efron 1982).

With an estimator of T , say \hat{T} , and its estimated variance, say $\hat{V}(\hat{T})$, many texts on sampling techniques (e.g., Cochran 1977, Särndal et al. 1992, and Schreuder et al. 1993) suggest or endorse the construction of an approximate $(1 - \alpha)100\%$ confidence interval for T of the form

$$\hat{T} \pm t_{1-\alpha/2}^{n-1} \sqrt{\hat{V}(\hat{T})} \tag{11}$$

where $t_{\alpha/2}^{n-1}$ is the upper $1 - \alpha/2$ quantile of the (central) t distribution with $n - 1$ degrees of freedom. For the situations and estimators considered here, t -based intervals generally are considered to be robust to deviations from normality of the parent population. For any particular population, i.e., for fixed N , T , and σ^2 , it is usually thought that the success of such t -based intervals in achieving their nominal coverage is often a function chiefly of sample size, n , especially when n is not too small relative to N .

We investigated various combinations of Eqs. 1–10 in constructing intervals of the form of Expression 11 in order to examine any systematic deviation when t -based intervals failed to cover the targeted parameter. Given the implicit invocation of a finite-population version of the central limit theorem to justify Expression 11, we also examined Gaussian probability plots of \hat{T} and the failure characteristics of z -based intervals of the form

$$\hat{T} \pm z_{(1-\alpha/2)} \sqrt{V(\hat{T})} \tag{12}$$

where $z_{\alpha/2}$ is the upper $1 - \alpha/2$ quantile of the standard Gaussian distribution.

We also investigated the ‘‘corrected’’ interval proposed by Johnson (1978), which is

$$\hat{T} + N^2 \hat{\mu}_3 / (6ns^2) \pm t_{1-\alpha/2}^{n-1} \sqrt{\hat{V}(\hat{T})} \tag{13}$$

where $\hat{\mu}_3$ and s^2 are the conventional sample-based estimates of the population’s third and second central moments, respectively.

To ensure that the 25 000-sample simulation provided empirical sampling distributions that were not contaminated with appreciable sampling error arising from the simulation itself, various checks were instituted to compare simulation averages to their theoretical values. The results, reported in Table A1 of the Appendix, verify that the simulation provided results that were accurate to well within 1% with each of the three accuracy checks used.

RESULTS

For the $n = 40$ samples of the four populations, the Monte Carlo sampling distributions of \hat{T}_V are shown in the left column of Fig. 2, and the corresponding normal probability plots of \hat{T}_V are displayed in the adjacent column. The symmetry of these sampling distributions is noteworthy when compared to the asymmetry of the parent populations: the skewness coefficients of these 25 000-member empirical sampling distributions were 0.20, 0.14, 0.24, and 0.46, respectively. Their approximate normality is indicated in the probability plots in the right column of the Fig. 2. Corresponding to Fig. 2 are the results in the left column of numerical results displayed in Table 2, which show the actual coverage rates of the z -based intervals at the 80%, 90%, and 95% levels, and the proportion of times the interval failed from below (upper endpoint of the interval is $<T$) and from above (lower endpoint of the interval is $>T$). The actual coverage rate of these intervals matches the nominal rate rather well, but there is an apparent tendency to fail from above when sampling from these positively skew distributions. This tendency disappeared when the samples were selected from symmetrized versions of these populations, as shown in the last four lines of the left column of numerical results (see Appendix B for details on the algorithm to symmetrize the populations).

The coverage and failure rates of t -based intervals,

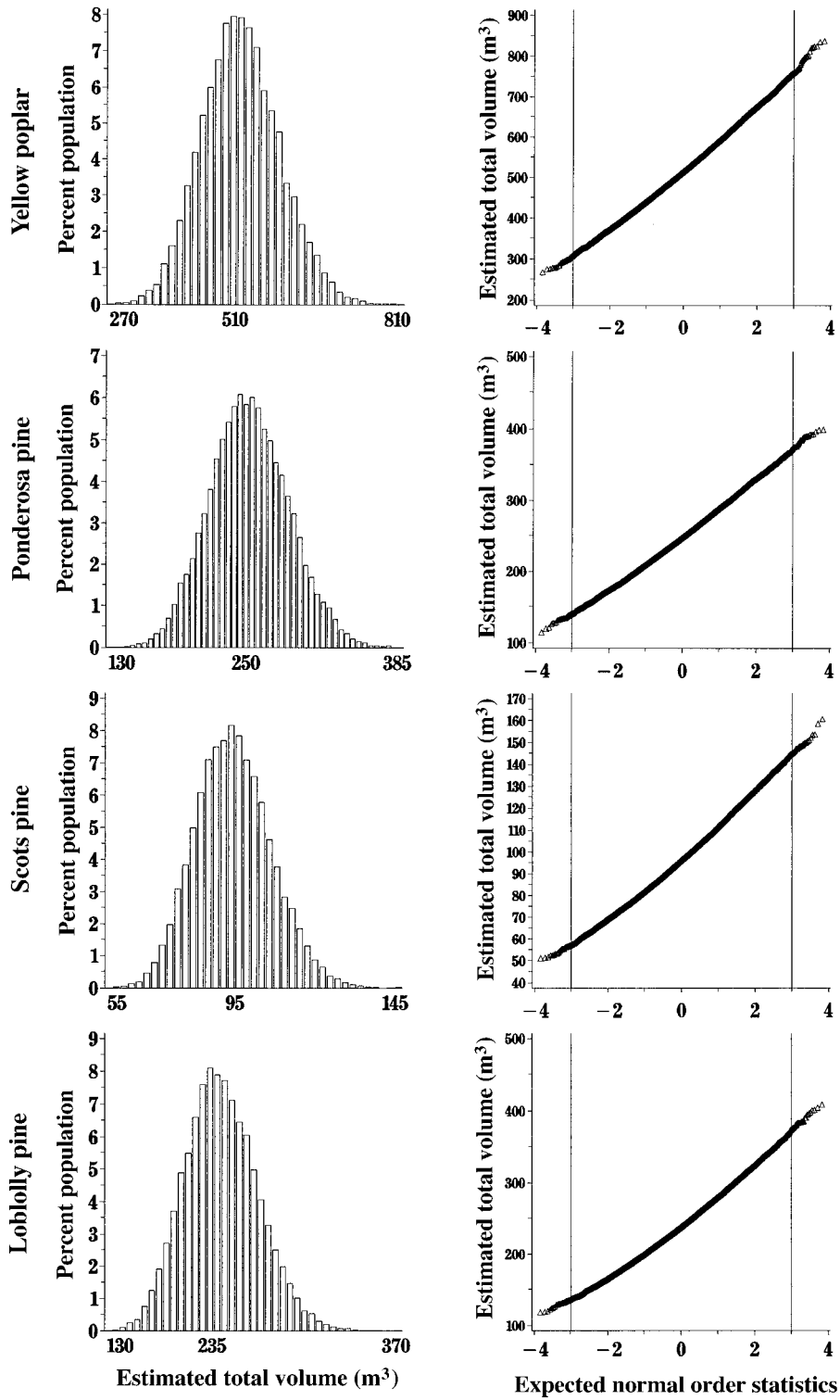


FIG. 2. Empirical sampling distributions of \hat{T}_V based on 25 000 Monte Carlo samples are shown on the left, and the corresponding normal probability plots are shown on the right.

TABLE 2. Coverage rates and failure rates of confidence intervals (given as percentages) for T_V constructed from samples of $n = 40$ trees and the Horvitz–Thompson estimator of T_V .

Population	$\hat{T}_V \pm z_{(1-\alpha/2)}\sqrt{V(\hat{T}_V)}$ (z-based interval)		$\hat{T}_V \pm t_{1-\alpha/2}^{n-1}\sqrt{\hat{V}(\hat{T}_V)}$ (t-based interval)	
	Below (Cover)	Above	Below (Cover)	Above
Nominal CI = 80%				
Yellow poplar	9.5 (79.8)	10.7	13.2 (79.0)	7.8
Ponderosa pine	10.0 (79.9)	10.2	13.6 (79.3)	7.1
Scots pine	9.6 (79.8)	10.6	14.3 (78.8)	7.0
Loblolly pine	9.5 (80.0)	10.4	14.3 (78.9)	6.9
Nominal CI = 90%				
Yellow poplar	4.3 (89.8)	5.9	8.1 (88.9)	3.0
Ponderosa pine	4.7 (89.8)	5.5	8.8 (88.5)	2.7
Scots pine	4.9 (90.2)	5.0	9.0 (88.4)	2.5
Loblolly pine	4.1 (90.1)	5.8	9.2 (88.5)	2.4
Nominal CI = 95%				
Yellow poplar	1.9 (95.0)	3.1	5.2 (93.7)	1.1
Ponderosa pine	2.0 (95.0)	3.0	5.8 (93.3)	0.9
Scots pine	1.7 (95.2)	3.1	5.9 (93.2)	0.8
Loblolly pine	1.7 (95.3)	3.0	6.1 (93.2)	0.8
Symmetric yellow poplar	2.7 (94.8)	2.6	2.5 (95.1)	2.4
Symmetric Ponderosa pine	2.5 (94.9)	2.6	3.4 (94.5)	2.1
Symmetric Scots pine	2.4 (95.0)	2.5	2.4 (95.2)	2.4
Symmetric loblolly	2.4 (95.1)	2.5	2.4 (95.2)	2.4

Note: All rates are based on the coverage of intervals from 25 000 samples selected in the simulation.

viz. $\hat{T}_V \pm t_{1-\alpha/2}^{n-1}\sqrt{\hat{V}(\hat{T}_V)}$, are displayed in the right column of Table 2, again for samples of size $n = 40$. While the coverage rate of these intervals is slightly less than the nominal rate in all cases, the difference is arguably small, ranging at most to a 3.4% deviation from the nominal rate. The striking feature, however, is the overwhelming tendency for these intervals to fail, not from above as with the z-based intervals, but from below. For example, an 80% confidence interval for yellow poplar aggregate volume ($T_V = 516.9 \text{ m}^3$) failed to cover T_V nearly twice as often from below than from above, and, at the 95% level, the estimated interval failed nearly five times as often from below. Even at $n = 56$, which for this population corresponds to Cochran’s recommended sample size of $25G_1^2$, the overall coverage rate improved to 94% at the nominal 95% level, yet the failure rate from below was still four times greater than that from above. Johnson’s corrected interval helped to balance the above/below failure rates by a negligible amount.

We concluded that the asymmetric failure rate is attributable to the strong positive correlation between \hat{T}_V and $\sqrt{\hat{V}(\hat{T}_V)}$ which is evident in the four frames of Fig. 3. For the four Monte Carlo distributions shown in Fig. 2, the empirical correlation coefficient between \hat{T}_V and $\sqrt{\hat{V}(\hat{T}_V)}$ was 0.75, 0.79, 0.79, and 0.76, respectively. As a consequence of this correlation, when T is underestimated, $\hat{V}(\hat{T}_V)$ will tend to be small in value, too, thereby generating an interval that tends to be too narrow to cover T . Conversely, when T is overestimated, the generated interval tends to be so wide as to nearly always include T . (Exactly opposite behavior was ob-

served when the populations were reflected in a way that preserved the mean volume but induced a negatively skewed distribution of the same magnitude as the positive skew in the original population.)

The behavior of \hat{T}_R is similar to that of \hat{T}_V but with some intriguing differences. The Monte Carlo sampling distributions of \hat{T}_R for the four populations are exhibited on the left in Fig. 4, and the corresponding normal probability plots of the 25 000 estimates are shown on the right. As expected, each distribution of \hat{T}_R is much less dispersed than the corresponding distribution of \hat{T}_V in Fig. 2. Unexpectedly, the distribution of \hat{T}_R appears to be more closely Gaussian than that of \hat{T}_V .

The left column of numerical results in Table 3 shows the coverage and failure rates of z-based intervals, $\hat{T}_R \pm z_{(1-\alpha/2)}\sqrt{V(\hat{T}_R)}$. While the observed difference between the actual and nominal coverage of these intervals is slightly greater here than for \hat{T}_V , it is still within a percentage point in all but one case. As with \hat{T}_V , here too there is a slight asymmetry in the failure rate of z-based intervals, but in this case the pattern is reversed from that evident in the left column of Table 2: a slight tendency to fail more often from below is evident. The reason for this reversal is unknown.

The remainder of Table 3 shows the coverage and failure rates of various t-based intervals constructed from the ratio estimator of T_V . For the intervals whose behavior is summarized in the middle two columns of numerical results, both are similar in that they are centered on \hat{T}_R ; however, they differ in that the estimate of variance is $\hat{V}(\hat{T}_R)$ and $\hat{V}_a(\hat{T}_R)$ across the two columns, respectively. The intervals summarized in the final col-

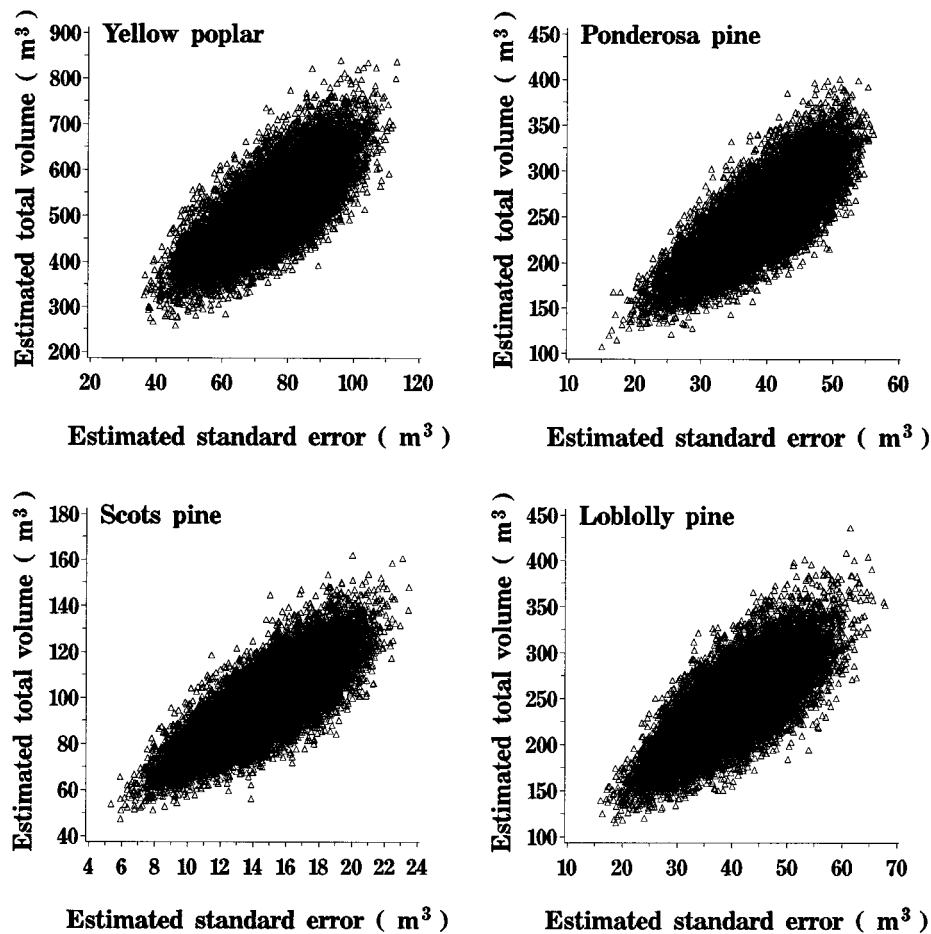


FIG. 3. Scatterplots of \hat{T}_V vs. $\hat{V}(\hat{T}_V)$ for the 25 000 samples selected from each population.

umn on the right use the jackknife estimator, \hat{T}_J , and the jackknife variance estimator, $\hat{V}(\hat{T}_J)$.

All three t -based intervals based on the ratio estimator exhibit the same tendency to fail far more frequently from below as was observed in Table 2 for the interval based on \hat{T}_V . For the interval of the form $\hat{T}_R \pm t_{1-\alpha/2}^{n-1} \sqrt{\hat{V}(\hat{T}_R)}$ the pattern and rate of failure to cover T mimics that observed in Table 2 quite closely. The interval that uses $\hat{V}_a(\hat{T}_R)$ in place of $\hat{V}(\hat{T}_R)$ tends to fail from below uniformly less often. The interval based on \hat{T}_J and $\hat{V}_j(\hat{T}_J)$ performs best, in that its tendency to fail from below most nearly matches the failure rate from above; even with these jackknifed intervals, however, there is still a pronounced tendency to fail from below.

The asymmetric failure rates with t -based intervals from the $n = 40$ samples are exacerbated when constructed with $n = 20$ samples and ameliorated with $n = 25G_1^2$ samples. For no sample size did Johnson's corrected intervals make much of a difference.

DISCUSSION

Positive skewness of a population affects confidence statements about the population total T for two reasons:

(1) it imparts skew to the sampling distribution of an estimator of T , and (2) a strong positive correlation results between the estimator of T and an estimator of its standard error. A suspicion of the latter consequence led us to examine the behavior of customary $(1 - \alpha)100\%$ confidence intervals in detail, as it is a topic that is completely absent from any traditional text on sampling methods.

For the moderately skewed populations we examined, the skewness of the sampling distributions of \hat{T}_V and \hat{T}_R was quite slight, as evidenced by the left panels of Figs. 2 and 4. Moreover, the overall coverage rate of 25 000 confidence intervals constructed during the simulation experiment was but slightly diminished from the nominal value of $(1 - \alpha)100\%$.

The strength of the positive correlation between the estimators of T_V and their corresponding estimators of standard error surprised us. As a result of this correlation, there was a marked effect on the rates of failure to cover T_V from below and from above. At the nominal 80% level, intervals that failed to cover the population total occurred nearly twice as often from below as from above; as the nominal $(1 - \alpha)100\%$ level increased, the asymmetry in failure became more pronounced. We

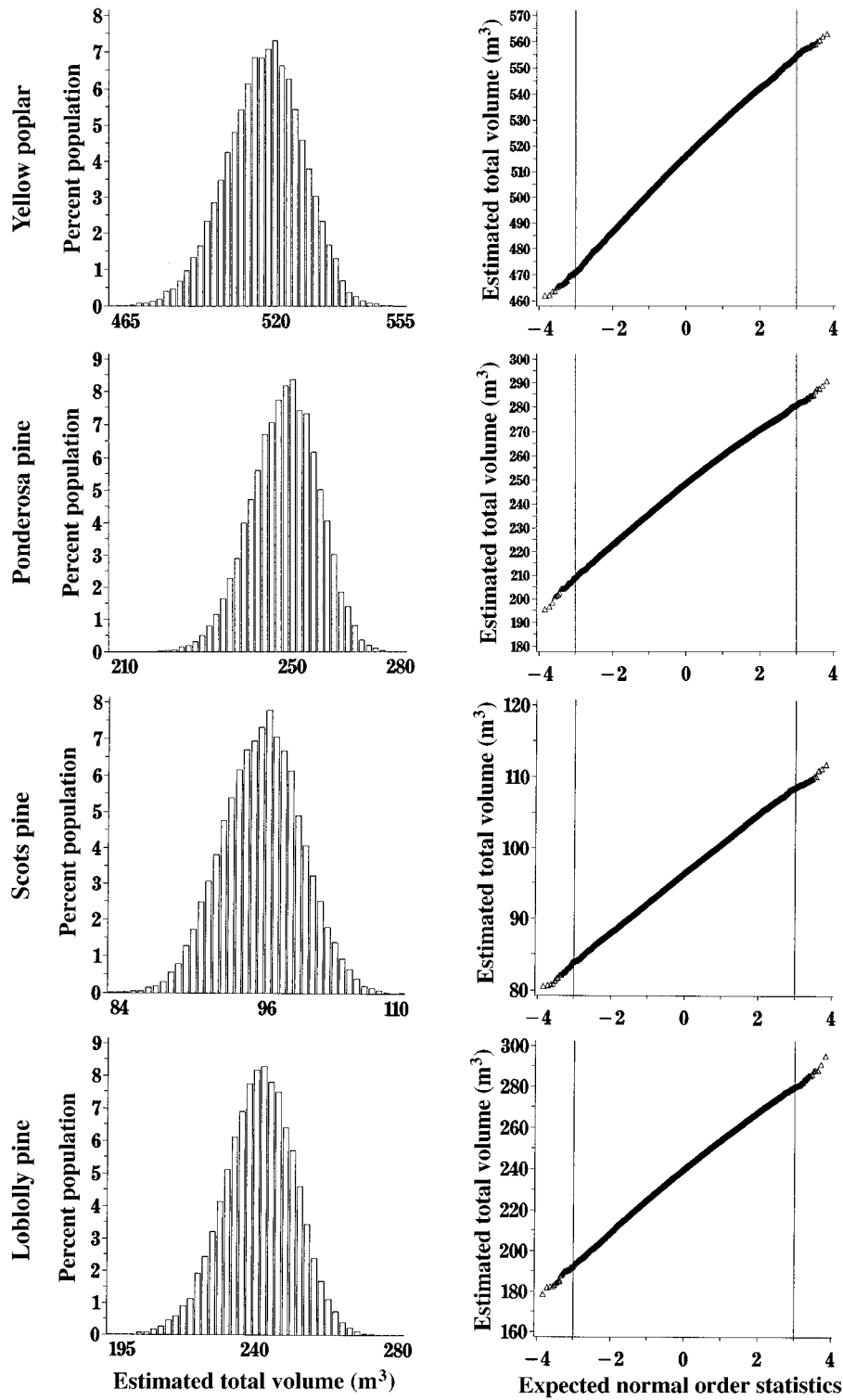


FIG. 4. Empirical sampling distributions of \hat{T}_R based on 25 000 Monte Carlo samples are shown on the left, and the corresponding normal probability plots are shown on the right.

TABLE 3. Coverage rates and failure rates of confidence intervals (given as percentages) for T_V constructed from samples of $n = 40$ trees and the customary and jackknifed ratio estimator of T_V .

Population	$\hat{T}_R \pm z_{(1-\alpha/2)} \sqrt{V(\hat{T}_R)}$ (z-based interval)		$\hat{T}_R \pm t_{1-\alpha/2}^{n-1} \sqrt{\hat{V}(\hat{T}_R)}$ (normal t -based interval)		$\hat{T}_R \pm t_{1-\alpha/2}^{n-1} \sqrt{\hat{V}_d(\hat{T}_R)}$ (alternative t -based interval)		$\hat{T}_J \pm t_{1-\alpha/2}^{n-1} \sqrt{\hat{V}(\hat{T}_J)}$ (jackknife t -based interval)	
	Below	(Cover) Above	Below	(Cover) Above	Below	(Cover) Above	Below	(Cover) Above
Nominal CI = 80%								
Yellow poplar	11.8	(79.4) 8.8	14.3	(77.8) 7.9	12.3	(77.9) 9.8	9.1	(81.7) 9.2
Ponderosa pine	12.4	(78.9) 8.7	16.6	(76.9) 6.4	14.8	(76.7) 8.4	8.5	(85.7) 5.8
Scots pine	12.2	(79.1) 8.8	18.0	(76.4) 5.5	16.6	(75.8) 7.6	13.3	(80.0) 6.7
Loblolly pine	11.7	(79.6) 8.7	14.7	(77.4) 7.9	12.7	(77.6) 9.7	9.7	(81.0) 9.4
Nominal CI = 90%								
Yellow poplar	6.4	(89.6) 4.0	9.1	(87.6) 3.3	7.0	(88.1) 4.8	5.0	(90.2) 4.1
Ponderosa pine	6.9	(89.3) 3.8	11.1	(86.7) 2.2	9.0	(87.2) 3.8	4.3	(93.9) 1.9
Scots pine	6.2	(89.4) 4.4	13.5	(84.8) 1.7	11.7	(85.0) 3.2	9.2	(88.3) 2.5
Loblolly pine	6.4	(89.5) 4.1	9.3	(87.6) 3.1	7.2	(88.0) 4.8	5.3	(90.4) 4.3
Nominal CI = 95%								
Yellow poplar	3.6	(94.9) 1.5	6.1	(92.7) 1.2	4.3	(93.5) 2.2	2.8	(95.4) 1.8
Ponderosa pine	3.9	(94.6) 1.6	7.7	(91.6) 0.7	5.4	(93.0) 1.6	2.2	(97.2) 0.6
Scots pine	3.2	(94.6) 2.1	10.3	(89.2) 0.5	8.6	(90.1) 1.3	6.2	(92.9) 0.9
Loblolly pine	3.6	(94.6) 1.8	6.3	(92.5) 1.2	4.3	(93.3) 2.4	2.9	(95.2) 2.0

Notes: All rates are based on the coverage of intervals from 25 000 samples selected in the simulation.

suspect that these results will be surprising to others who must deal with positively skew distributions.

The impact of these results on the practice and interpretation of sampling-based estimates is difficult to predict or prescribe, although one might be justified in viewing one-sided confidence intervals with skepticism. But, beyond that, it is simply the realization that the risk of erroneous inference is not balanced between overestimation and underestimation.

It is hoped that the anomalous behavior highlighted here may lead to further investigation into its detection in a practical setting and development of a corrective procedure. For example, to achieve an approximate $\alpha/2$ failure rate from below and above, an interval that is asymmetric around the estimator of T might be devised. A transformation of the data may also be useful, although it would be exceedingly difficult to assess the design-based properties of an estimator based on transformed data, where the transformation itself was suggested by the particular sample drawn. This suggests that jackknifing or other forms of resampling might be useful, especially in light of the improvement evident in Table 3 with the jackknifed estimate \hat{T}_J and corresponding jackknifed estimate of variance. Efron and Tibshirani (1993: 163–165) discuss the failure of bootstrap t -intervals when the variance of an estimator, say θ , is a function of θ . They outline a five-step algorithm of nested bootstrapping to stabilize the variance and estimate a confidence interval in this situation. Perhaps this procedure may be applied fruitfully to counter the ill effects of population skewness, too.

Finally, results identical to those exhibited in Tables 2 and 3 are produced when the focus of interest is on an estimator of the mean $\mu_V = T_V/N$.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the suggestions to improve this manuscript that were made by Mary Christman, Andrew Robinson, and Steve Stehman.

LITERATURE CITED

Bebbington, A. C. 1975. A simple method of drawing a sample without replacement. *Applied Statistics* **24**:136.

Campbell, C., and M. Meyer. 1978. Some properties of t confidence intervals for survey data. Pages 437–442 in *Proceedings of the American Statistical Association, Section on Survey Research Methods*.

Cochran, W. G. 1977. *Sampling techniques*. John Wiley and Sons, New York, New York, USA.

Dalén, J. 1986. Sampling from finite populations: actual coverage probabilities for confidence intervals on the population mean. *Journal of Official Statistics* **2**:13–24.

Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. CBMS-NSF Regional Conference Series in Applied Mathematics, Volume 38. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, USA.

Efron, B., and R. J. Tibshirani. 1993. *An introduction to the bootstrap*. Chapman and Hall, New York, New York, USA.

Fuller, W. A. 1991. Simple estimators for the mean of skewed populations. *Statistica Sinica* **1**:137–158.

Johnson, N. J. 1978. Modified t tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association* **73**:536–544.

Knox, R. G., R. K. Peet, and N. L. Christensen. 1989. Population dynamics in loblolly pine stands: changes in skewness and size inequality. *Ecology* **70**:1153–1166.

Neyman, J., and E. S. Pearson. 1928. On the use and interpretation of certain test criteria for purposes of statistical inference. Part I. *Biometrika* **20A**:175–240.

Overton, W. S., and S. V. Stehman. 1994. Variance estimation in the EMAP strategy for sampling discrete ecological resources. *Environmental and Ecological Statistics* **1**:133–152.

Rao, P. S. R. S., and J. N. K. Rao. 1971. Small sample results for ratio estimators. *Biometrika* **58**:625–630.

Rivest, L.-P. 1994. Statistical properties of Winsorized means for skewed distributions. *Biometrika* **81**:373–383.

Rivest, L.-P., and D. Hurtubise. 1995. On Searls' Winsorized mean for skewed populations. *Survey Methodology* **21**: 107–116.

Royall, R. M., and W. G. Cumberland. 1985. Conditional coverage properties of finite population confidence intervals. *Journal of the American Statistical Association* **80**: 355–359.

Särndal, C.-E., B. Swensson, and J. Wretman. 1992. Model assisted survey sampling. Springer-Verlag, New York, New York, USA.
 Schreuder, H. T., T. G. Gregoire, and G. B. Wood. 1993. Sampling methods for multiresource forest inventory. Wiley, New York, New York, USA.

Weiner, J., and O. T. Solbrig. 1984. The meaning and measurement of size hierarchies in plant populations. *Oecologia* **61**:334–336.
 Weiner, J., and S. C. Thomas. 1986. Size variability and competition in plant monocultures. *Oikos* **47**:211–222.

APPENDIX A

To provide assurance that the Monte Carlo simulation of 25 000 samples was sufficiently accurate for the purposes of this study, the following check statistics were computed:

$$MC_1 = \left(\frac{\bar{T}_v - T_v}{T_v} \right) 100\%$$

$$MC_2 = \left(\frac{\sqrt{\sigma_{\hat{T}_v}^2} - \sqrt{V(\hat{T}_v)}}{T_v} \right) 100\%$$

$$MC_3 = \left(\frac{(25\ 000)^{-1} \sum \sqrt{\hat{V}(\hat{T}_v)} - \sqrt{V(\hat{T}_v)}}{\sqrt{V(\hat{T}_v)}} \right) 100\%$$

where \bar{T}_v and $\sigma_{\hat{T}_v}^2$ are the computed mean and variance of the empirical distribution of \hat{T}_v generated from the Monte Carlo simulation, and $(25\ 000)^{-1} \sum \sqrt{\hat{V}(\hat{T}_v)}$ is the computed mean of the Monte Carlo distribution of the estimates $\hat{V}(\hat{T}_v)$. For $M = N!/n!(N - n)!$ distinct samples, when selected without replacement, the MC_1 and MC_2 check statistics should be identically zero, and MC_3 should be very close to zero. The aim was to ensure that the simulation provided values of these check statistics sufficiently close to zero that one can ignore

TABLE A1. Statistics indicating the accuracy of the 25 000-sample simulation study.

Population	n = 20 samples			n = 40 samples		
	MC ₁	MC ₂	MC ₃	MC ₁	MC ₂	MC ₃
Yellow poplar	-0.1	0.0	-0.1	0.1	0.0	0.2
Ponderosa pine	0.4	0.0	0.6	0.0	0.0	-0.1
Scots pine	0.1	0.0	0.3	0.0	0.1	-0.5
Loblolly pine	0.1	0.0	0.2	-0.1	0.0	-0.3

Notes: All table entries are percentage deviations. MC_1 is the percentage deviation of the observed mean of \hat{T}_v from T_v ; MC_2 is the deviation of the observed standard error of \hat{T}_v from its analytical value, $\sqrt{V(\hat{T}_v)}$, expressed as a percentage of T_v ; and MC_3 is the percentage deviation of the average estimated standard error from $\sqrt{V(\hat{T}_v)}$.

the sampling error due to the simulation itself when interpreting the results. The results in Table A1 show that the greatest deviation from zero was 0.6%, and in most cases it was far less.

APPENDIX B

Symmetrization of a population involved the following steps. The original population was reflected about the mean value, and then the reflected population was shifted to the right so that its minimum value coincided with the minimum value of the original population. The original population was shifted to the right, also, so that its minimum value after shifting was just greater than the maximum value of the re-

flected and shifted population. The shifted original population was combined with the shifted reflected population. Every second value of this combined population was deleted, so that the size of the combined, symmetric population was the same as the size of the original population. Finally, each value of this population was multiplied by a constant to make its total, T_v , identical to the the value of T_v in the original population.