Journal of Official Statistics
Vol. 2, No. 1, 1986, pp. 13–24

Sampling From Finite Populations: Actual Coverage Probabilities for Confidence Intervals on the Population Mean

Jörgen Dalén[1]

Abstract: The normal and the Student's t approximations for the distribution of the sample mean under simple random sampling from a finite population are empirically compared, resulting in support for the use of the latter approximation. A simple rule of thumb for estimating the population mean is proposed. The rule is of the type $n > KG_2^2$, where n is the sample size, K is a constant, and G_2 is the standardized absolute third moment of the population. It is derived empirically through extensive studies of dichotomous populations with different degrees of skewness by looking at the actual coverage probabilities of the standard confidence intervals for these populations. The rule is designed so that a nominal 95 % confidence interval on the population mean can be assumed to be correct α % of the time in an average sense defined in the paper by assigning different values to K for five levels of α from 85 to 94.5 %. The rule is tested and empirically verified by means of Monte-Carlo experiments for critical sample sizes on populations based on fixed percentiles of well-known parametric distributions.

Key words: t distribution; skewness; central limit theorem.

1. Introduction

In survey sampling the prevailing strategy for estimating a finite population parameter θ is based upon an (approximately) unbiased point estimator $\hat{\theta}$ and an (approximately) unbiased variance estimator $\hat{V}(\hat{\theta})$. Then a central limit theorem is evoked for the assumption that $\hat{\theta}$ is approximately normally distributed, and it is stated that the interval

$$\hat{\theta} \pm 1.96 \{\hat{V}(\hat{\theta})\}^{1/2}$$

covers the true value θ with a probability of approximately 95 %. Sometimes 1.96 is exchanged for the corresponding value taken from the Student's t table with the appropriate degrees of freedom.

In an individual survey, however, it is not easy to establish the accuracy of this approximation. It depends on a number of factors such as the type of estimator and design used, the underlying population, and the sample size. Consequently, there is a great need to increase our knowledge of the coverage properties of the standard procedures for calculating confidence intervals in different set-ups, and to work out simple rules of thumb useful for the survey practitioner.

[1] Senior Statistician, Department of Enterprise Statistics, Statistics Sweden.

In sampling practice, the standard approximation frequently fails. One example of this is sampling from populations of enterprises with very skewed variables such as production, employment, investment, export, or import. Another example is small area estimation where most observations are set at zero.

In this paper we will study the case of estimating the population mean by the sample mean under simple random sampling. For this case, Erdös and Rényi (1959) and Hájek (1960) have developed conditions under which the sampling distribution converges to normality.

Stenlund and Westlund (1975, 1976), Barrett and Goldsmith (1976), and Hägglund (1978) studied this problem by means of Monte-Carlo experiments.

For populations in which the principal deviation from normality consists of a marked positive skewness, Cochran (1977) suggested the simple rule

$$n > 25G_1^2,$$

where n is the sample size and G_1 the usual measure of population skewness defined below. According to Cochran, "this rule is designed so that a 95 % confidence probability statement will be wrong not more than 6 % of the time."

Robinson (1978) gave an asymptotic Edgeworth-type expansion for the sum of a simple random sample without replacement from a finite population. The crucial quantities in this expansion are skewness and kurtosis. He showed that, subject to a condition ensuring that the population distribution is "almost continuous," the absolute difference between the distribution function of the sample sum and the asymptotic expansion is bounded by a term containing the absolute fifth moment of the population distribution.

Perhaps it would be possible to base a rule of thumb on this expansion, although it would have to be quite complicated, as it has to take into account the skewness and the kurtosis as well as the absolute fifth moment of the population distribution. It would also have to exclude the (lattice) case where the population is clustered around too few values.

We use a simpler approach, inspired by Cochran's rule and by Höglund (1978), who has derived the following remainder term estimate (slightly manipulated algebraically to serve our purpose):

$$\left| F(t) - \Phi\left[\frac{t-n\mu}{\sigma\sqrt{n(1-f)}}\right]\right| \leq \frac{CG_2}{\sqrt{n(1-f)}}, \qquad (1.1)$$

where

F is the distribution function of the sum of a sample of n units among the N population units $(x_1, x_2, \ldots, x_N)$,
Φ is the standard normal distribution function,
μ is the population mean,
σ is the population standard deviation,
$f = n/N$,
C is an absolute constant (Quine (1985) shows that $C \leq 145$) and

$$G_2 = \frac{\sum_{j=1}^{N} |x_j-\mu|^3}{N\sigma^3}.$$

From above, we have

$$G_1 = \frac{\sum_{j=1}^{N} (x_j-\mu)^3}{N\sigma^3}.$$

We notice three things about (1.1):

i) The deviation from normality is bounded by a term containing the factor G_2, the standardized absolute third moment.
ii) The formula is symmetric in n and $(N-n)$, indicating that the accuracy of the normal approximation could be expected to be

equally good for these two sample sizes. In fact, as pointed out by Plane and Gordon (1982), the sampling distributions of the sample mean of n and $(N–n)$ units are mirror images of each other except for a scale change. For this reason, $N/2$ is the sample size where the sampling distribution is closest to the normal.

iii) If we wish to have an upper limit to the absolute difference (called ε) of formula (1.1), we obtain the condition $CG_2\sqrt{n(1-f)} < \varepsilon \leftrightarrow n(1-f) > C^2G_2^2/\varepsilon^2$ or, if we consider large populations and set $K=C^2/\varepsilon^2$ then $n > KG_2^2$. This provides a theoretical argument for a rule similar to Cochran's, although the population skewness is replaced by G_2. The constant K is dependent only on the maximum error allowed in the approximation. Of course, in constructing two-sided confidence intervals we are interested in the difference between the deviations in symmetric pairs of percentiles of the distribution, usually 2.5 and 97.5, and therefore ε is not necessarily equal to the difference between the nominal and actual coverage probability of the confidence interval.

The above arguments provide the logical foundation for the empirical investigations presented in this paper. Here we calculate the exact coverage probabilities of confidence intervals based on the normal and the t distribution for dichotomous populations, where these probabilities are based on a simple hypergeometric distribution. No other distributions are known for which these probabilities are easily calculated for arbitrary sample sizes and degrees of skewness. Moreover, there is strong reason to believe that this distribution, due to its extreme lattice character, represents more or less the worst case. This was actually proved by Esséen (1956) in the i.i.d. case. The t distribution is studied together with the normal because it is recommended by many textbook authors, although a solid theoretical argument based on a limit theorem is lacking.

The structure of the type of rule of thumb that we investigate is therefore

$$n > K_\alpha G_2^2, \tag{1.2}$$

the interpretation being that if we have an exact value for G_2 and are prepared to allow an actual coverage probability of α, we must choose a sample size greater than $K_\alpha G_2^2$. We study the degree to which the K_α:s are stable for different degrees of skewness of the dichotomous population and for finite realizations of some continuous parametric distributions.

For extremely skewed populations $G_2 \approx G_1$ and then this rule coincides with Cochran's. But, at the other extreme, for symmetric populations $G_1=0$ and Cochran's rule is reduced to $n > 0$ and is therefore generally unsuitable. For this reason G_1 is not used in the empirical investigations below. On the contrary, $G_2 \geqslant 1$ for all populations, with equality if and only if the population is dichotomous and symmetric as shown in Dalén (1985). For reference, $G_2=4/\sqrt{2\pi}\approx 1.6$ for the normal distribution and $\sqrt{27}/4 \approx 1.3$ for the uniform distribution.

2. The Dichotomous Population

For the dichotomous population studied, the following notations are used:

Value	Number of units in the population	Number of units in the sample
0	$N-M$	$n-m$
1	M	m
Total	N	n

The population has the following characteristics:

population mean = $\mu = P$,
population variance = $\sigma^2 = P - P^2$,
$G_1 = (1\text{–}2P)/(P\text{–}P^2)^{0.5}$ and
$G_2 = (1\text{–}2P+2P^2)/(P\text{–}P^2)^{0.5}$,
where $P = M/N$.

Notice that $G_2 = G_1 + 2P^{1.5}/(1\text{–}P)^{0.5}$ so that $\lim_{P \to 0} (G_2\text{–}G_1) = 0$ and that $G_1 = 0$ and $G_2 = 1$ when $P = 0.5$.

The sample has the following characteristics:
sample mean = $\overline{X} = m/n$ and
sample variance = $s^2 = (m\text{–}m^2/n)/(n\text{–}1)$.

A nominal 95 % confidence interval for μ based on the sample outcome now becomes

$$\overline{X} - t_{0.975}\, s \sqrt{(1\text{–}f)/n} < \mu < \overline{X} + t_{0.975}\, s \sqrt{(1\text{–}f)/n} ,$$

where $t_{0.975}$ is either 1.96 or the corresponding quantity from the t distribution with $(n\text{–}1)$ degrees of freedom.

(Since we are interested in how bad the approximation would be at worst, the continuity correction is not used. If it was, the constants K_α needed would be much lower but would be more difficult to generalize to other types of populations.)

Now, let I_m be the indicator of this confidence interval statement as a function of the sample outcome. That is:

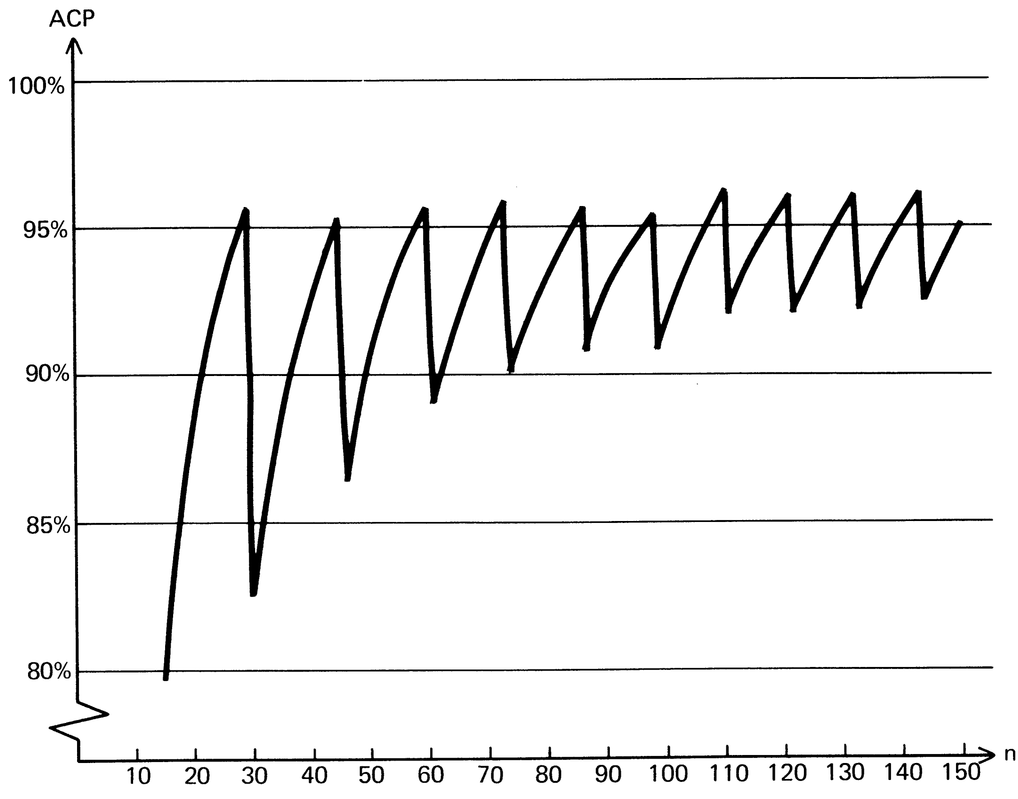

Diagram 1. Actual coverage probabilities. N = 300, P = 0.1, t approximation used

$$I_m = \begin{cases} 1 \text{ for those } m \text{ where the confidence interval statement is true} \\ 0 \text{ for those } m \text{ where the confidence interval statement is false.} \end{cases}$$

The *actual coverage probability* (ACP) is now defined as the probability for a sample of a certain size n from our population to produce a true confidence interval statement, that is

$$\text{ACP}(N, M, n) = \sum_{m=0}^{n} I_m\, p(m),$$

$$\text{where } p(m) = \frac{\binom{M}{m}\binom{N-M}{n-m}}{\binom{N}{n}}$$

according to the hypergeometric distribution.

Computer programs were written where these probabilities for various combinations of N, n, and P were computed. The programming language was SIMULA, and the IMSL procedures MDBIN, MDHYP, and MDSTI were used.

In Diagram 1, a typical example is given of how the ACP varies with n up to $N/2$. We see that the ACP does not increase monotonously with n. Typically there are intervals of increase (shorter and shorter as n increases), followed by downward jumps. This is of course due to the discontinuous character of the population studied. Up to 85–90 % the increase is rapid, but then oscillations occur around a mean, which comes closer and closer to 95 %.

3. Average ACP

ACP is a measure of the goodness of the normal or t approximation. If the nominal confidence level is 95 %, we consider the approximation to be good if we can count on a coverage probability α sufficiently close to 95 %.

However, it is not possible in an individual case to have a "guaranteed" ACP. This is partly because we do not know the population characteristics exactly, but also because of the oscillations in the ACP-level as n varies as shown in diagram 1. A device intended to deal with the latter problem is the concept of an *average ACP.*

Definition 1: α is an average ACP for a certain sample size n in a certain population if

$$\sum_{j=0}^{s} \text{ACP}(n+j)/(s+1) \geqslant \alpha$$

for all integers s such that $0 \leqslant s \leqslant N/2-n$.

Definition 2: For a certain population the sample size n_α required for an average ACP of α is the smallest n for which α is an average ACP.

These two definitions also give a unique value for the constant K in (1.2) for a certain population, namely

$$K_\alpha = n_\alpha / G_2^2 .$$

In Tables 1a–1e, values of these constants are presented for various combinations of N and P including the binomial case ($N=\infty$) and for five α-levels: 85, 90, 93, 94, and 94.5 %.

Two comments regarding the calculation of these tables should be made:

i) Definition 1 could not be applied exactly in the binomial case, since N is infinite. Instead, we had to choose a maximum sample size up to which we calculated the ACP and which was equated to $N/2$ in definition 1. This sample size was in all cases larger than $100G_2^2$.
ii) In those cases where $N/2-n_\alpha < 50$, we have put brackets around the value of the constant. This is because those values may be considered accidental from a global point of view. (The number 50 is, of course, to a certain extent arbitrary.)

Table 1a. Constants for an Average ACP of 85 %

P	G_1	G_2	Normal approximation $N=$ 500	1000	2000	5000	∞	t approximation $N=$ 500	1000	2000	5000	∞
0.01	9.849	9.851	1.6	1.8	1.9	1.9	2.0	1.6	1.8	1.9	1.9	2.0
0.02	6.857	6.863	1.8	1.9	2.0	2.0	2.0	1.8	1.9	2.0	2.0	2.0
0.03	5.510	5.521	1.9	2.0	2.0	2.1	2.1	1.9	2.0	2.0	2.1	2.1
0.04	4.695	4.711	2.0	2.1	2.1	2.1	2.1	2.0	2.1	2.1	2.1	2.1
0.05	4.130	4.152	2.1	2.1	2.1	2.2	2.2	2.1	2.1	2.1	2.2	2.2
0.1	2.667	2.733	2.5	2.5	2.5	2.5	2.5	2.4	2.4	2.4	2.5	2.5
0.15	1.960	2.086	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8	2.8
0.2	1.500	1.700	3.5	3.5	3.5	3.5	3.5	3.1	3.1	3.1	3.1	3.1
0.25	1.155	1.443	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4	3.4
0.3	0.873	1.266	3.7	3.8	3.8	3.8	3.7	3.7	3.8	3.8	3.8	3.7
0.35	0.629	1.143	4.6	4.6	4.6	4.6	4.6	3.8	3.8	3.8	3.8	3.8
0.4	0.408	1.061	5.3	5.3	5.3	5.3	5.3	4.4	4.4	4.4	4.4	4.4
0.45	0.201	1.015	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9	3.9
0.5	0	1	4	4	4	4	4	4	4	4	4	4

Table 1b. Constants for an Average ACP of 90 %

P	Normal approximation $N=$ 500	1000	2000	5000	∞	t approximation $N=$ 500	1000	2000	5000	∞
0.01	1.9	3.5	3.8	3.9	4.0	1.9	3.5	3.8	3.9	4.0
0.02	3.6	3.9	4.0	4.1	4.2	3.6	3.9	4.0	4.1	4.2
0.03	3.9	4.0	4.2	4.3	4.3	3.9	4.0	4.2	4.3	4.3
0.04	4.1	4.2	4.3	4.4	4.4	4.1	4.2	4.3	4.4	4.4
0.05	4.3	4.4	4.5	4.5	4.5	4.3	4.4	4.5	4.5	4.5
0.1	5.1	5.2	5.2	5.2	5.1	5.1	2.9	3.1	3.1	5.1
0.15	5.7	5.7	5.7	6.0	6.0	3.4	3.4	3.4	3.4	3.4
0.2	6.6	6.6	6.6	6.6	6.6	3.8	3.8	3.8	3.8	3.8
0.25	7.2	7.2	7.2	7.2	7.2	4.3	4.3	4.3	4.3	4.3
0.3	7.5	7.5	7.5	7.5	7.5	4.4	4.4	4.4	4.4	4.4
0.35	4.6	6.1	6.1	8.4	8.4	4.6	4.6	4.6	4.6	4.6
0.4	5.3	5.3	5.3	5.3	5.3	4.4	4.4	4.4	4.4	4.4
0.45	4.9	7.8	7.8	7.8	7.8	4.9	4.9	4.9	4.9	4.9
0.5	8	8	8	8	8	5	5	5	5	5

Table 1c. Constants for an Average ACP of 93 %

P	Normal approximation $N=$ 500	1000	2000	5000	∞	t approximation $N=$ 500	1000	2000	5000	∞
0.01	–	(4.9)	7.9	9.9	11.7	–	(4.9)	7.9	9.9	11.7
0.02	(5.2)	8.2	10.0	11.8	12.1	(5.1)	8.2	10.0	10.4	12.1
0.03	(8.0)	10.1	10.7	10.9	12.5	(6.9)	8.8	10.7	10.9	11.0
0.04	8.6	10.6	10.9	11.3	11.3	8.6	9.3	10.9	11.3	11.3
0.05	9.2	11.0	11.3	11.6	11.7	9.2	9.7	9.9	11.6	11.6
0.1	12.8	13.1	13.1	13.3	13.4	11.0	9.5	9.6	11.5	9.8
0.15	14.9	13.1	13.1	13.1	13.1	10.8	8.7	10.8	10.8	11.0
0.2	12.1	12.5	17.0	14.5	17.0	9.7	7.3	7.3	10.0	10.0
0.25	13.4	16.3	16.3	16.3	16.3	7.7	7.7	7.7	8.2	8.2
0.3	14.4	15.0	15.0	15.0	15.0	8.7	8.7	8.7	8.7	8.7
0.35	15.3	18.4	18.4	18.4	18.4	5.4	5.4	5.4	5.4	5.4
0.4	12.4	17.8	19.5	19.5	19.5	6.2	6.2	6.2	6.2	6.2
0.45	17.5	17.5	17.5	17.5	17.5	4.9	4.9	4.9	4.9	4.9
0.5	17	19	19	19	19	11	8	8	9	9

Table 1d. Constants for an Average ACP of 94 %

P	Normal approximation N= 500	1000	2000	5000	∞	t approximation N= 500	1000	2000	5000	∞
0.01	–	(5.1)	–	18.6	22.5	–	(5.1)	–	18.6	22.5
0.02	(5.2)	–	16.4	20.0	23.3	(5.2)	–	16.4	19.9	23.2
0.03	(8.2)	(16.2)	18.6	22.1	24.0	(8.2)	13.9	18.6	22.1	21.3
0.04	(11.3)	17.3	20.7	22.7	23.1	(11.3)	16.0	19.6	20.0	23.1
0.05	(14.4)	16.9	20.1	23.4	23.8	(14.4)	16.9	19.0	19.4	23.8
0.1	19.4	24.8	25.3	23.8	25.8	19.4	20.1	20.5	22.4	22.5
0.15	24.1	23.2	26.9	25.3	27.6	18.6	17.2	19.3	21.4	19.8
0.2	31.8	28.4	30.8	30.8	26.3	17.0	19.4	15.2	22.1	17.6
0.25	25.9	24.0	31.7	36.5	36.5	11.0	13.9	19.2	14.4	16.8
0.3	30.6	28.7	36.8	33.7	31.8	8.7	11.9	11.9	15.0	15.0
0.35	29.1	35.2	30.6	26.8	43.7	13.0	10.0	10.0	10.0	10.0
0.4	43.5	30.2	28.4	28.4	30.2	6.2	6.2	6.2	6.2	6.2
0.45	26.2	26.2	36.9	36.9	34.9	5.8	5.8	5.8	5.8	5.8
0.5	36	37	29	34	34	14	9	9	9	9

Table 1e. Constants for an Average ACP of 94.5 %

P	Normal approximation N= 500	1000	2000	5000	∞	t approximation N= 500	1000	2000	5000	∞
0.01	–	(5.2)	–	25.2	46.3	–	(5.2)	–	25.2	46.3
0.02	(5.3)	–	21.1	36.6	48.7	(5.3)	–	(21.1)	35.6	45.0
0.03	–	(16.4)	(31.8)	39.8	46.5	–	(16.4)	29.5	39.8	46.5
0.04	–	(22.5)	33.4	43.7	49.0	–	(22.5)	32.2	38.6	45.1
0.05	–	(28.2)	40.3	47.9	47.6	–	(27.0)	34.8	37.5	43.5
0.1	(32.8)	35.7	49.0	48.1	53.3	(28.5)	34.4	38.1	42.0	48.5
0.15	36.5	44.8	49.4	51.5	53.8	31.7	37.7	38.4	36.8	40.4
0.2	44.3	50.9	57.8	60.2	58.5	24.2	42.9	45.3	33.2	41.9
0.25	46.6	51.8	56.6	59.5	57.6	35.5	33.6	29.3	25.0	25.0
0.3	61.8	53.1	44.3	54.3	64.3	23.1	21.2	21.2	29.3	26.2
0.35	57.4	51.3	56.7	62.0	62.0	13.0	13.0	13.0	16.1	16.1
0.4	53.3	47.0	56.8	60.4	62.1	15.1	13.3	13.3	13.3	13.3
0.45	44.6	81.5	62.1	57.3	57.3	5.8	8.7	8.7	8.7	8.7
0.5	62	75	65	80	69	14	9	9	9	9

A summary of the tables is given by the span for each level:

α	Normal	t
85	1.6– 5.3	1.6– 4.4
90	1.9– 8.4	1.9– 5.1
93	7.9–19.5	4.9–12.1
94	16.4–43.7	5.8–23.8
94.5	25.2–81.5	5.8–46.5

There are some interesting features of the tables:

i) For the higher α-levels and less skewed populations, the t approximation gives much smaller constants. This provides an empirical argument for the use of this approximation instead of the normal one.

ii) For very skewed populations the convergence rate of the t approximation is "rapid in the beginning and slow in the end" in the sense that we get relatively small constants for low α-levels and large ones for high α-levels. The opposite holds for less skewed populations. For $P=0.5$ the constant only increases from 4 to 9 in most populations as α increases from 85 to 94.5, while for $P=0.01$ the increase in the binomial case is from 2 to 46.3.

4. Almost Continuous Populations

The K_α-values obtained in the dichotomous case were also tried on populations of an "almost continuous" type. The K-values chosen were 3, 5, 11, 20, and 40, corresponding roughly to the five α-levels.

It is to be expected that the convergence rate is more rapid for such populations than for the dichotomous population with its pronounced lattice character. If this is correct, the α-levels should generally be exceeded if we choose the above K-values.

The populations used were based on fixed percentiles of the continuous theoretical distributions below. For each of these four distributions, six different finite populations were generated with different degrees of skewness by taking the percentiles from 0.001 to 0.999 with intervals of 0.002, making the population size 500.

The distributions are:

I) The beta distribution with the probability density function

$$f(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)}; 0 \leqslant x \leqslant 1, a>0, b>0$$

where $B(a,b)$ is the beta function, and

$$\gamma = \text{coefficient of skewness} = \frac{2(b-a)(a+b+1)^{1/2}}{(a+b+2)(ab)^{1/2}};$$

II) The lognormal distribution with the probability density function

$$f(x) = (\sigma x)^{-1}(2\pi)^{-1/2} \exp\{-(\log x-\theta)^2/2\sigma\};$$

$$x > 0, \sigma > 0$$

and

$$\gamma = \{\exp(\sigma^2)+2\}\{\exp(\sigma^2)-1\}^{1/2};$$

III) The power function with the distribution function

$$F(x) = (x/\theta)^c; 0 < x < \theta, \theta > 0, c > 0$$

and

$$\gamma = 2(1-c)(2+c)/(3+c)c^{1/2};$$

IV) The Weibull distribution with the distribution function

$$F(x) = 1 - \exp\{-\theta x^c\}; x \geqslant 0, \theta > 0, c > 0$$

and

$$\gamma = \frac{\Gamma(1+3/c)-3\Gamma(1+2/c)\,\Gamma(1+1/c)+2\Gamma^3(1+1/c)}{\{\Gamma(1+2/c)-\Gamma^2(1+1/c)\}^{3/2}},$$

where $\Gamma(x)$ is the gamma function.

In all cases γ stands for the coefficient of skewness corresponding to G_1 above, that is

$$\gamma = E(x-\mu)^3/\{E(x-\mu)^2\}^{3/2}.$$

A common feature of all these distributions is that for some value(s) of the involved parameter(s) γ can take on at least any value > 0.

The reference used for these distributions is Patel et al. (1976).

For each population five different sample sizes were chosen to correspond as closely as possible to the K-values above. This means that n was chosen so that $n \geqslant KG_2^2 > n-1$ for values of K of 3, 5, 11, 20, and 40 respectively.

For every sample size, 1 000 simple random samples without replacement were drawn. For each sample, the population mean was estimated and a confidence interval based on the sample standard deviation and the t distribution was calculated. The number of cases when this interval covered the true population mean was counted. This number divided by 1 000 became our estimated actual coverage probability (EACP). EACP is of course

Table 2a. EACPs for 1 000 Random Samples from Populations Based on the Beta Distribution. (b=1 coincides with a uniform distribution.)

K	α-level (%)	$b=1/a=1$, $G_2=1.299$	n	$b=1/a=1.756$, $G_2=1.592$	n	$b=1/a=2.390$, $G_2=1.978$	n	$b=1/a=4.083$, $G_2=2.946$	n	$b=1/a=6.16$, $G_2=3.882$	n	$b=1/a=8.509$, $G_2=4.714$	n
3	85	94.2	6	90.2	8	90.9	12	89.8	27	88.5	46	89.3	67
5	90	94.6	9	92.6	13	90.9	20	90.9	44	91.7	76	91.6	112
11	93	94.8	19	93.7	28	93.3	44	94.1	96	94.2	166	94.6	245
20	94	95.1	34	95.5	51	93.9*	79	94.5	174				
40	94.5	94.7	68	95.0	102	94.6	157						

Table 2b. EACPs for 1 000 Random Samples from Populations Based on the Lognormal Distribution

K	α-level (%)	$\sigma=0.1$, $G_2=1.597$	n	$\sigma=0.4$, $G_2=1.982$	n	$\sigma=0.72$, $G_2=2.943$	n	$\sigma=0.92$, $G_2=3.883$	n	$\sigma=1.06$, $G_2=4.695$	n	$\sigma=1.3$, $G_2=6.331$	n
3	85	95.9	8	92.7	12	92.1	26	89.6	46	90.6	67	89.5	121
5	90	94.7	13	93.6	20	91.4	44	92.8	76	91.8	111	91.7	201
11	93	95.9	28	95.6	44	94.4	96	95.2	166	94.5	243		
20	94	94.8	51	95.5	79	93.9*	174						
40	94.5	95.4	102	96.2	158								

Table 2c. EACPs for 1 000 Random Samples from Populations Based on the Power Function Distribution

K	α-level (%)	$c=0.316$, $G_2=1.590$	n	$c=0.1933$, $G_2=1.981$	n	$c=0.0905$, $G_2=2.940$	n	$c=0.0540$, $G_2=3.877$	n	$c=0.0376$, $G_2=4.698$	n	$c=0.0213$, $G_2=6.319$	n
3	85	89.6	8	88.5	12	87.7	26	88.8	46	87.0	67	88.2	120
5	90	92.3	13	91.0	20	91.1	44	92.0	76	92.0	111	90.9	200
11	93	93.7	28	93.3	44	93.1	96	92.9*	166	94.9	243		
20	94	94.0	51	95.0	79	94.5	173						
40	94.5	94.7	102	95.1	157								

Table 2d. EACPs for 1 000 Random Samples from Populations Based on the Weibull Distribution

K	α-level (%)	$c=2.15$, $G_2=1.585$	n	$c=1.24$, $G_2=1.984$	n	$c=0.795$, $G_2=2.923$	n	$c=0.615$, $G_2=3.967$	n	$c=0.53$, $G_2=4.831$	n	$c=0.425$, $G_2=6.554$	n
3	85	94.4	8	92.8	12	90.7	26	90.4	48	88.9	71	91.0	129
5	90	94.1	13	93.8	20	92.0	43	92.0	79	91.6	117	91.0	215
11	93	95.0	28	94.3	44	93.8	94	93.2	174				
20	94	94.5	51	93.7*	79	95.3	171						
40	94.5	95.2	101	94.1*	158								

stochastic in this case with a standard error of 0.7 % to 1.1 % when the ACP ranges from 95 % to 85 %.

In Tables 2a–2d, the outcome of these Monte-Carlo trials is presented in terms of the EACP for a certain combination of population and sample size. We see, as expected, that in almost all cases the α-levels obtained from the studies of the dichotomous population are exceeded, for $\alpha = 85$ and 90 % by large margins. The convergence rate up to 90–92 % in general seems to be rapid. In only 5 cases out of 90 are the presupposed levels not obtained (those cases are indicated with an asterisk). The EACPs are in these cases 0.1–0.4 % below the expected level. One case is for $\alpha = 93$ % (0.1 below), three cases are for $\alpha = 94$ % (0.1–0.3 below) and one case is for $\alpha = 94.5$ % (0.4 below). The deviations may well be entirely due to the stochastic effect of the Monte-Carlo trials.

5. Conclusions

Our empirical investigations into the problem of how large the sample size must be to allow a calculation of a standard 95 % confidence interval for a simple random sample from a finite population support the following tentative conclusions:

i) When the difference is of any significance, the confidence interval should be based on the Student's t distribution with $n-1$ degrees of freedom, making the convergence rate more rapid.

ii) A rule of thumb of the Cochran type (1.2) is useful to the practicing statistician, if he has reasonably good knowledge of G_2. A choice of $K = 20$ should in most cases allow him to count on an ACP of 94 % for a nominal 95 % confidence interval. For "almost continuous" populations, a K greater than 3 should be enough for an ACP of around 90 %. For some symmetric populations, i.e., those close to uniform, even more liberal limits will do.

The rule of thumb could be used a priori to assist a decision on sample size. If our knowledge of G_2 is insufficient before the survey is conducted, the rule could be used to evaluate the quality of a standard confidence interval based on the sample data after the sample is drawn.

6. Some Comments for Practical Application

Practical application of a rule of thumb such as (1.2) raises a number of questions, as commented below.

a) *G_2 is not known.* In practice no population parameters are known exactly and G_2 is no exception. Estimating G_2 from the sample is difficult. No unbiased estimator is known and the corresponding sample quantity

$$g_2 = \sqrt{n}\sum_{i=0}^{n}|x_i-\bar{X}|^3 / \{\sum_{i=0}^{n}(x_i-\bar{X})^2\}^{3/2}$$
$$\leqslant (n^2-2n+2)/n\,\sqrt{n-1} < \sqrt{n}$$

and therefore underestimates G_2 with a probability of one as soon as $n \leqslant G_2^2$, as shown in Dalén (1985).

Moreover, if the population consists of two subsets A and B where B contains a few large-value units with a small probability of showing up in a sample of size n, and G_2 calculated over $A \cup B$ is much greater than G_2' calculated over A, then in most sample outcomes we would in a sense estimate G_2' rather than G_2 and our rule of thumb based on g_2 instead of G_2 would be seriously misleading.

It is therefore necessary to know more about G_2 than what can be inferred from a sample. If, for example, we know that the range of population values is not much larger than the range of sample values we would be on safer ground using g_2 or a similar estimator.

b) *Stratified samples.* In the presence of a skewed population, estimation by the sample mean under simple random sampling is certainly not the best sampling strategy. In such situations the prevailing strategy at central statistical offices is stratified random sampling using the weighted mean with the stratum sizes as weights. However, due to the lack of a sufficiently good auxiliary variable we sometimes end up with very skewed subpopulations in many strata. It then becomes an issue when it is reasonable to use the normal approximation in stratified samples.

Some empirical studies of this problem have been made, but due to the many dimensions involved (number of strata, stratum sizes, sample sizes, variances, and degrees of skewness in each stratum), results are difficult to present systematically. There are indications that a rule like

$$n > K_\alpha \Sigma\, w_i G_{2i}^2 \,,$$

where summation is over strata, where G_{2i} is G_2 in stratum i, and w_i are weights such that $\Sigma w_i = 1$, would work satisfactorily. If a Neyman allocation is used, $w_i = N_i\sigma_i/\Sigma N_i\sigma_i$ seems to work in many cases. (N_i is the size and σ_i^2 the variance of stratum i.)

c) *Alternative confidence intervals.* There is no universally applicable method for constructing confidence intervals for the mean in finite populations when the standard procedure fails. No method can handle the case where there is a large unit of unknown size with a small probability of showing up in a sample.

In some situations alternative methods are available, however. In the dichotomous case there are methods based on the binomial or hypergeometric distribution. Cochran (1977) describes these methods with examples.

Johnson (1978) considers a procedure where the t variable is modified according to a Cornish-Fisher expansion of the sample mean. The procedure is shown to give improved confidence intervals and tests for distributions as asymmetric as χ^2 with two degrees of freedom for sample sizes as small as 13.

A novel method in this area is that of resampling from the empirical distribution function, called the bootstrap. Efron (1981) presents two methods – the percentile method and the bootstrap t – which use this idea for constructing confidence intervals, and he points out the similarity between the results from Johnson's t and the bootstrap t. A shortcoming of the bootstrap techniques is the large amount of computing power they require. There is a rapid development in this area with much research. Among recently published papers mention could be made of Rao and Wu (1984), Bickel and Freedman (1984), Efron and Tibshirani (1985), and Abramovitch and Singh (1985).

6. Acknowledgements

This paper has been part of a Statistics Sweden project on sampling and estimation led by Bengt Swensson. Swensson, Carl-Erik Särndal, Jan Hagberg, Thomas Höglund and the referees have also contributed to this paper with helpful comments and suggestions.

7. References

Abramovitch, L. and Singh, K. (1985): Edgeworth Corrected Pivotal Statistics and the Bootstrap. The Annals of Statistics, Vol. 13, No. 1, pp. 116–132.

Barrett, J.P. and Goldsmith, L. (1976): When Is n Sufficiently Large? The American Statistician, Vol. 20, No 2, pp. 67–70.

Bickel, P.J. and Freedman, D.A. (1984): Asymptotic Normality and the Bootstrap in Stratified Sampling. The Annals of Statistics, Vol. 12, No. 2, pp. 470–482.

Cochran, W.G. (1977): Sampling Techniques. Third edition, Wiley.

Dalén, J. (1985): Bounds on Standardized Moments in Samples and Finite Populations. Research Report 1985:4, Department of Statistics, University of Stockholm.

Efron, B. (1981): Nonparametric Standard Errors and Confidence Intervals. The Canadian Journal of Statistics, Vol. 9, No. 2, pp. 139–172.

Efron, B. and Tibshirani, R. (1985): The Bootstrap Method for Assessing Statistical Accuracy. Technical Report No. 101, Division of Biostatistics, Stanford University, Stanford, California.

Erdös, P. and Rényi, A. (1959): On the Central Limit Theorem for Samples from a Finite Population. Publ. Math. Inst. Acad. Sci. 4, pp. 49–57.

Esséen, C.G. (1956): A Moment Inequality with an Application to the Central Limit Theorem. Scandinavian Actuarial Journal, 1–2, pp. 160–170.

Hájek, J. (1960): Limiting Distributions in Simple Random Sampling from a Finite Population. Publ. Math. Inst. Hung. Acad. Sci. 5, pp. 361–374.

Hägglund, G. (1978): Monte Carlo Studies of the Normal Approximation in Stratified Sampling. Unpublished manuscript, University of Uppsala (In Swedish).

Höglund, T. (1978): Sampling from a Finite Population. A Remainder Term Estimate. Scandinavian Journal of Statistics, 5, pp. 69–71.

Johnson, N.J. (1978): Modified *t* Tests and Confidence Intervals for Asymmetrical Populations. Journal of the American Statistical Association, Vol. 73, No. 363, pp. 536–544.

Patel, J.K., Kapadia, C.H., and Owen, D.B. (1976): Handbook of Statistical Distributions. Marcel Dekker, Inc.

Plane, D.R. and Gordon, K.R. (1982): A Simple Proof of the Non-applicability of the Central Limit Theorem to Finite Populations. The American Statistician, Vol. 36, No. 3, Part 1, pp. 175–176.

Quine, M.P. (1985): Remainder Term Estimates in a Conditional Central Limit Theorem for Integer-valued Random Variables, Journal of the Australian Mathematical Society, Series A, 39, pp. 75–85.

Rao, J.N.K. and Wu, C.F.J. (1984): Bootstrap Inference for Sample Surveys. American Statistical Association. Proceedings of the Section on Survey Research Methods, pp. 106–112.

Robinson, J. (1978): An Asymptotic Expansion for Samples from a Finite Population. The Annals of Statistics, Vol. 6, No. 5, pp. 1005–1011.

Stenlund, H. and Westlund, A. (1975): A Monte Carlo Study of Simple Random Sampling from a Finite Population. Scandinavian Journal of Statistics, 2, pp. 106–108.

Stenlund, H. and Westlund, A. (1976): On the Asymptotic Normality of the Mean Estimator. Scandinavian Journal of Statistics, 3, pp. 127–131.

Received June 1984
Revised November 1985