or, transposing

$$\hat{N} = \frac{CM}{R} \tag{2.1}$$

where $\hat{N}$ = Estimate of population size at time of marking,* and the other terms are as defined above

This formula is the "Petersen estimate" of population size and has been widely used because it is intuitively clear. Unfortunately, formula (2.1) produces a *biased* estimator of population size, tending to overestimate the actual population. This bias can be large for small samples, and several formulas have been suggested to reduce this bias. Seber (1982) recommends the estimator

$$\hat{N} = \frac{(M + 1)(C + 1)}{(R + 1)} - 1 \tag{2.2}$$

which is unbiased if $(M + C) > N$ and nearly unbiased if there are at least seven recaptures of marked animals ($R > 7$). This formula assumes sampling *without* replacement (see page 263) in the second sample, so any individual can only be counted once.

In some ecological situations, the second sample of a Petersen series is taken *with* replacement so that a given individual can be counted more than once. For example, animals may be merely observed at the second sampling and not captured. For these cases, the size of the second sample (C) can be even larger than total population size (N) because individuals might be sighted several times. In this situation we must assume that the chances of sighting a marked animal are on the average equal to the chances of sighting an unmarked animal. The appropriate estimator from Bailey (1952) is

$$\hat{N} = \frac{M(C + 1)}{(R + 1)} \tag{2.3}$$

which differs only very slightly from equation (2.2) and is nearly unbiased when the number of recaptures (R) is 7 or more.

2.1.1 Confidence Intervals

How reliable are these estimates of population size? To answer this critical question, a statistician constructs *confidence intervals* around the estimates. A *confidence interval* is a range of values that is expected to include the true population size a given percentage of the time. Typically the given percentage is 95%, but you can construct 90% or 99% confidence intervals, or any range you wish. The high and low values of a confidence interval are called the *confidence limits*. Clearly, we want confidence intervals to be as small as possible, and the statistician's job is to recommend confidence intervals of minimal size consistent with the assumptions of the data at hand.

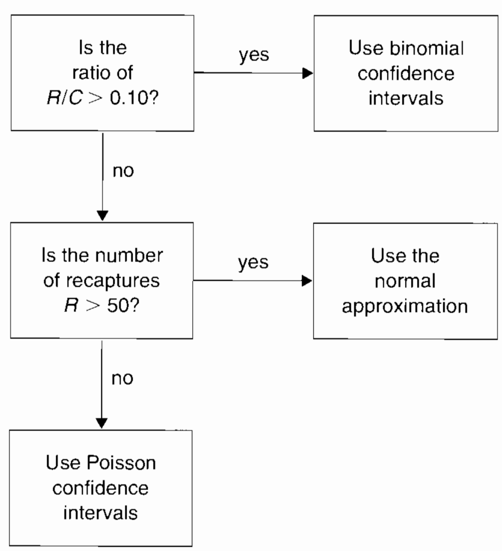
Confidence intervals are akin to gambling. You can state that the chances of flipping a coin and getting "heads" is 50%, but after the coin is flipped, it is either "heads" or

* A ^ over a variable means "an estimate of."

"tails." Similarly, after you have estimated population size by the Petersen method and calculated the confidence interval, the true population size (unfortunately not known to you) will either be inside your confidence interval or outside it. You cannot know which, and all the statistician can do is tell you that *on the average* 95% of confidence intervals will cover the true population size. Alas, you only have one estimate, and *on the average* does not tell you whether your one confidence interval is lucky or unlucky.

Confidence intervals are an important guide to the precision of your estimates. If a Petersen population estimate has a very wide confidence interval, you should not place too much faith in it. If you wish, you can take a larger sample next time and narrow the confidence limits. But remember that even when the confidence interval is narrow, the true population size may *sometimes* be outside the interval. Figure 2.1 illustrates the variability of Petersen population estimates from artificial populations of known size, and shows that some random samples by chance produce confidence intervals that do not include the true value.

Several techniques of obtaining confidence intervals for Petersen estimates of population size are available, and the particular one to use for any specific set of data depends upon the size of the population in relation to the samples we have taken. Seber (1982) gives the following general guide:



Poisson Confidence Intervals We discuss the Poisson distribution in detail in Chapter 4 (Section 4.2.1). We are concerned here only with the mechanics of determining confidence intervals for a Poisson variable.

Table 2.1 provides a convenient listing of values for obtaining 95% confidence intervals based on the Poisson distribution. An example will illustrate this technique. If I mark

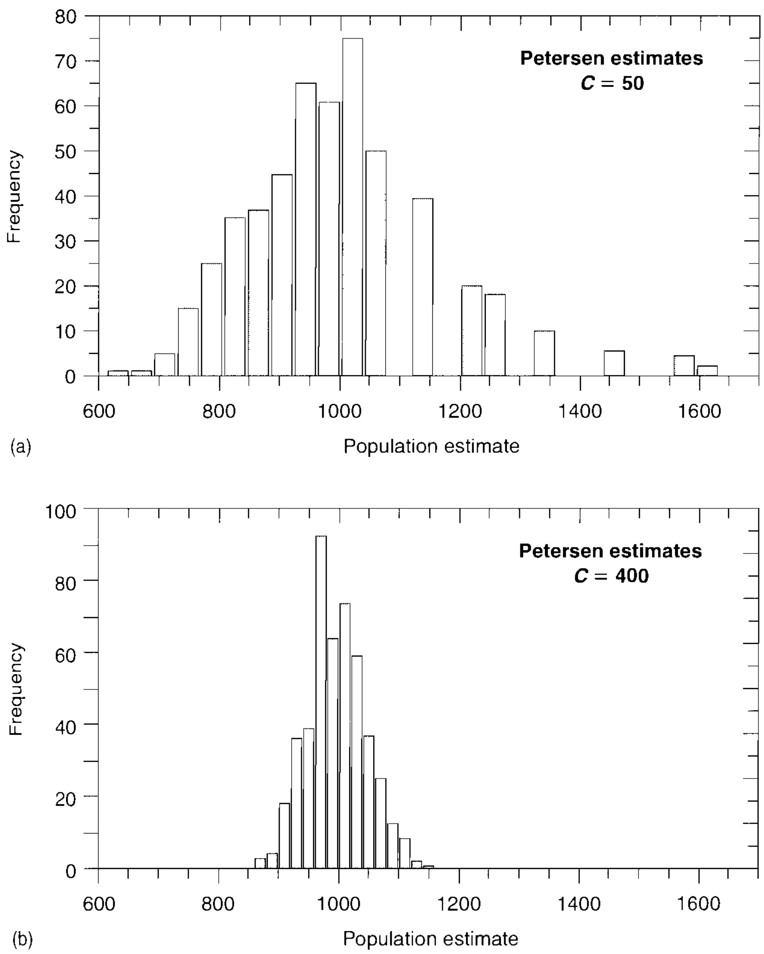


Figure 2.1 Petersen population estimates for an artificial population of $N = 1000$. Five hundred replicate samples were drawn. In both cases $M = 400$ individuals were marked in the first sample. (a) Samples of $C = 50$ were taken repeatedly for the second sample. A total of 13 estimates out of 500 did not include the known population size of 1000 (estimates below 705 or above 1570). (b) Samples of $C = 400$ were taken for the second sample. A total of 22 estimates out of 500 did not include the known population size (estimates below 910 or above 1105). Note the wide range of estimates of population size when the number of animals recaptured is small.

600 (M) and then recatch a total of 200 (C) animals, 13 (R) of which are marked, we have from Table 2.1:

$$\text{Lower 95\% confidence limit of } R \text{ when } R \text{ is } 13 = 6.686$$

$$\text{Upper 95\% limit of } R \text{ when } R \text{ is } 13 = 21.364$$

TABLE 2.1 CONFIDENCE LIMITS FOR A POISSON FREQUENCY DISTRIBUTION. Given the number of organisms observed (x), this table provides the upper and lower limits from the Poisson distribution. Do not use this table unless you are sure the observed counts are adequately described by a Poisson distribution.

	95%		99%			95%		99%	
x	Lower	Upper	Lower	Upper	x	Lower	Upper	Lower	Upper
0	0	3.285	0	4.771	46	34.05	60.24	29.90	65.96
1	0.051	5.323	0.010	6.914	47	34.66	61.90	31.84	66.81
2	0.355	6.686	0.149	8.727	48	34.66	62.81	31.84	67.92
3	0.818	8.102	0.436	10.473	49	36.03	63.49	32.55	69.83
4	1.366	9.598	0.823	12.347	50	37.67	64.95	34.18	70.05
5	1.970	11.177	1.279	13.793	51	37.67	66.76	34.18	71.56
6	2.613	12.817	1.785	15.277	52	38.16	66.76	35.20	73.20
7	3.285	13.765	2.330	16.801	53	39.76	68.10	36.54	73.62
8	3.285	14.921	2.906	18.362	54	40.94	69.62	36.54	75.16
9	4.460	16.768	3.507	19.462	55	40.94	71.09	37.82	76.61
10	5.323	17.633	4.130	20.676	56	41.75	71.28	38.94	77.15
11	5.323	19.050	4.771	22.042	57	43.45	72.66	38.94	78.71
12	6.686	20.335	4.771	23.765	58	44.26	74.22	40.37	80.06
13	6.686	21.364	5.829	24.925	59	44.26	75.49	41.39	80.65
14	8.102	22.945	6.668	25.992	60	45.28	75.78	41.39	82.21
15	8.102	23.762	6.914	27.718	61	47.02	77.16	42.85	83.56
16	9.598	25.400	7.756	28.852	62	47.69	78.73	43.91	84.12
17	9.598	26.306	8.727	29.900	63	47.69	79.98	43.91	85.65
18	11.177	27.735	8.727	31.839	64	48.74	80.25	45.26	87.12
19	11.177	28.966	10.009	32.547	65	50.42	81.61	46.50	87.55
20	12.817	30.017	10.473	34.183	66	51.29	83.14	46.50	89.05
21	12.817	31.675	11.242	35.204	67	51.29	84.57	47.62	90.72
22	13.765	32.277	12.347	36.544	68	52.15	84.67	49.13	90.96
23	14.921	34.048	12.347	37.819	69	53.72	86.01	49.13	92.42
24	14.921	34.665	13.793	38.939	70	54.99	87.48	49.96	94.34
25	16.768	36.030	13.793	40.373	71	54.99	89.23	51.78	94.35
26	16.77	37.67	15.28	41.39	72	55.51	89.23	51.78	95.76
27	17.63	38.16	15.28	42.85	73	56.99	90.37	52.28	97.42
28	19.05	39.76	16.80	43.91	74	58.72	91.78	54.03	98.36
29	19.05	40.94	16.80	45.26	75	58.72	93.48	54.74	99.09
30	20.33	41.75	18.36	46.50	76	58.84	94.23	54.74	100.61
31	21.36	43.45	18.36	47.62	77	60.24	94.70	56.14	102.16
32	21.36	44.26	19.46	49.13	78	61.90	96.06	57.61	102.42
33	22.94	45.28	20.28	49.96	79	62.81	97.54	57.61	103.84
34	23.76	47.02	20.68	51.78	80	62.81	99.17	58.35	105.66
35	23.76	47.69	22.04	52.28	81	63.49	99.17	60.39	106.12
36	25.40	48.74	22.04	54.03	82	64.95	100.32	60.39	107.10
37	26.31	50.42	23.76	54.74	83	66.76	101.71	60.59	108.61
38	26.31	51.29	23.76	56.14	84	66.76	103.31	62.13	110.16
39	27.73	52.15	24.92	57.61	85	66.76	104.40	63.63	110.37
40	28.97	53.72	25.83	58.35	86	68.10	104.58	63.63	111.78
41	28.97	54.99	25.99	60.39	87	69.62	105.90	64.26	113.45
42	30.02	55.51	27.72	60.59	88	71.09	107.32	65.96	114.33
43	31.67	56.99	27.72	62.13	89	71.09	109.11	66.81	114.99
44	31.67	58.72	28.85	63.63	90	71.28	109.61	66.81	116.44
45	32.28	58.84	29.90	64.26	91	72.66	110.11	67.92	118.33

TABLE 2.1 *Continued*

	95%		99%			95%		99%	
x	Lower	Upper	Lower	Upper	x	Lower	Upper	Lower	Upper
92	74.22	111.44	69.83	118.33	97	78.73	116.93	73.20	124.16
93	75.49	112.87	69.83	119.59	98	79.98	118.35	73.62	125.70
94	75.49	114.84	70.05	121.09	99	79.98	120.36	75.16	127.07
95	75.78	114.84	71.56	122.69	100	80.25	120.36	76.61	127.31
96	77.16	115.60	73.20	122.78					

Source: Crow and Gardner 1959.

When $x > 100$ use the normal approximation:

95% confidence limits of x:

Lower limit $= x - 0.94 - 1.96\sqrt{x - 0.02}$

Upper Limit $= x + 1.94 + 1.96\sqrt{x + 0.98}$

99% confidence limits of x:

Lower limit $= x - 1.99 - 2.576\sqrt{x + 0.33}$

Upper Limit $= x + 2.99 + 2.576\sqrt{x + 1.33}$

and we obtain the 95% confidence interval for estimated population size (sampling without replacement) by using these values of R in equation (2.2):

$$\text{Lower 95\% confidence limit on } \hat{N} = \frac{(601)(201)}{21.364 + 1} - 1 = 5402$$

$$\text{Upper 95\% confidence limit on } \hat{N} = \frac{(601)(201)}{6.686 + 1} - 1 = 15{,}716$$

Normal Approximation Confidence Intervals This method is essentially a "large sample" method that obtains a confidence interval on the fraction of marked animals in the second catch (R/C). It should be used only when R is above 50. The confidence interval for (R/C) is defined by the following formula:

$$\frac{R}{C} \pm \left\{ z_\alpha \left[\sqrt{\frac{(1 - f)(R/C)(1 - R/C)}{(C - 1)}} \right] + \frac{1}{2C} \right\} \tag{2.4}$$

where f = fraction of total population sampled in the second sample $= \dfrac{R}{M}$

$\dfrac{1}{2C}$ = correction for continuity

z_α = standard normal deviate for $(1 - \alpha)$ level of confidence
= 1.96 (for 95% confidence limits)
= 2.576 (for 99% confidence limits)

For large samples and a large population size, both the *finite population correction* $(1 - f)$ and the correction for continuity are negligible, and this formula for the normal approximation to the binomial simplifies to

$$\frac{R}{C} \pm z_\alpha \sqrt{\frac{(R/C)(1 - R/C)}{(C - 1)}} \tag{2.5}$$

The constant z_α defines $100\,(1 - \alpha)$ percent confidence limits, and values can be substituted from tables of the standard normal distribution (z); see Zar (1996, app19). For example, for 80% confidence limits, replace z_α with the constant 1.2816.

One example will illustrate this method. If I mark 1800 animals (M) and catch at the second sampling a total of 800 (C), of which 73 (R) are already tagged, then from formula (2.4) for 95% confidence limits,

$$\frac{73}{800} \pm \left\{ 19.6 \left[\sqrt{\frac{(1 - 73/1800)(73/800)(1 - 73/800)}{(800 - 1)}} \right] + \frac{1}{2(800)} \right\}$$

$$= 0.09125 \pm 0.020176$$

and the 95% confidence interval for R/C is 0.07107 to 0.111426. To obtain a 95% confidence interval for the estimated population size, we use these limits for R/C in equation (2.1):

$$\hat{N} = \frac{CM}{R}$$

$$\text{Lower 95\% confidence limit on } \hat{N} = \frac{1}{0.111426}(1800) = 16{,}154$$

$$\text{Upper 95\% confidence limit on } \hat{N} = \frac{1}{0.07107}(1800) = 25{,}326$$

Binomial Confidence Intervals Binomial confidence intervals for the fraction of marked animals (R/C) can be obtained most easily graphically from Figure 2.2. The resulting confidence interval will be approximate but should be adequate for most ecological data. For example, suppose I mark 50 birds (M), and then capture 22 (C) birds of which 14 (R) are marked. The fraction of marked animals (R/C) is 14/22, or 0.64. Move along the x-axis (*Sample Proportion*) to 0.64, and then move up until you intercept the first sample size line of C, or 22. Then read across to the y-axis (*Population Proportion*) to obtain 0.40, the lower 95% confidence limit for R/C. Now repeat the procedure to intercept the second sample size line of C, or 22. Reading across again, you find on the y-axis 0.83, the upper 95% confidence limit for R/C.

These confidence limits can be converted to confidence limits for population size (N) by the use of these limits for R/C in formula (2.1), exactly as described above (page 26). We use these limits for R/C in equation (2.1):

$$\hat{N} = \frac{CM}{R}$$

$$\text{Lower 95\% confidence limit on } \hat{N} = \frac{1}{0.83}(50) = 60$$

$$\text{Upper 95\% confidence limit on } \hat{N} = \frac{1}{0.40}(50) = 125$$

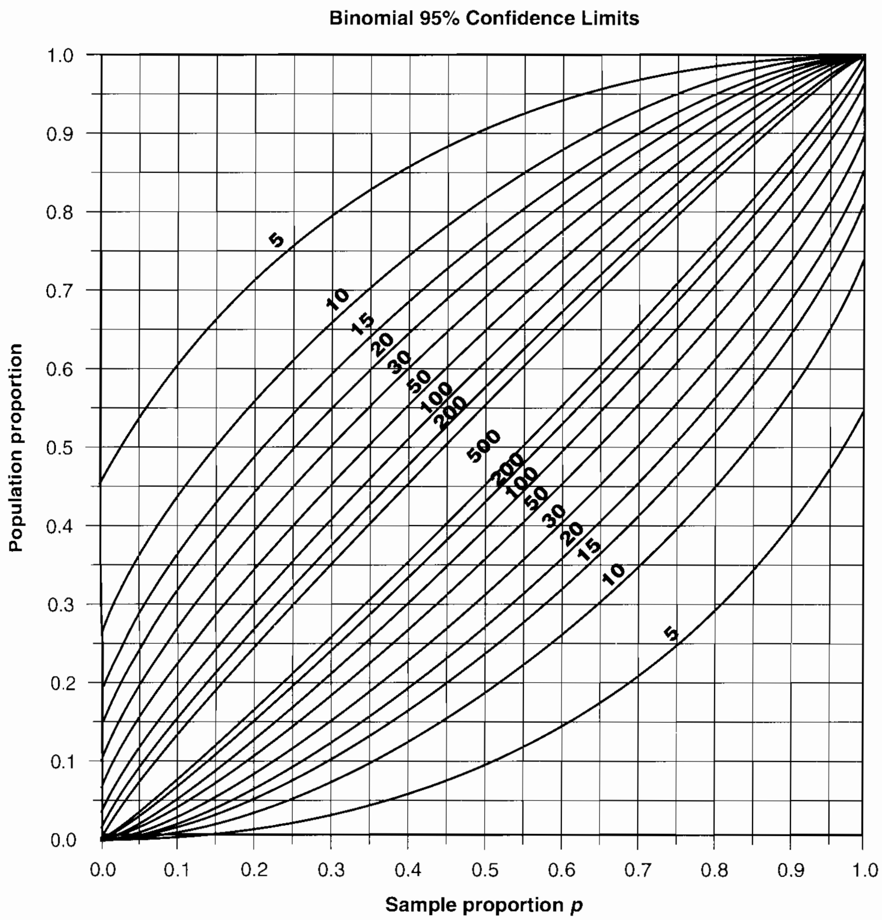


Figure 2.2 Upper and lower 95% confidence limits for a population proportion. Confidence limits are read off the y-axis for an observed value of p on the x-axis. Sample sizes are marked on the contour lines.

Alternatively, binomial confidence limits can be calculated using program-group EXTRAS (see Appendix 2), which uses the formulas given in Zar (1996, 524). Binomial confidence intervals can also be read from extensive tables, like those of Burnstein (1971), but for most ecological applications the precision of extensive tables of the binomial distribution is not required. For the bird example above, the tables provide the exact 95% confidence limits of 0.407 to 0.828, compared with the slightly less accurate 0.40 to 0.83 from Figure 2.2.

Program-group MARK-RECAPTURE (see Appendix 2) computes the Petersen estimate of population size and the appropriate confidence interval according to the recommendations of Seber (1982). Box 2.1 illustrates how the Petersen calculations are done.