The Siren call of the Schaefer Estimator

Carl J. Schwarz
Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6
cschwarz@cs.sfu.ca www.stat.sfu.ca/~cschwarz

Abstract:

Mark-recapture methods are commonly used to estimate the size of fish populations. The simplest estimator is the Petersen, but it can be biased if tagging or recovery probabilities are heterogeneous. Schaefer (1951) proposed an estimator that stratified the tagging and recovery probabilities. Conditions for consistency of this estimator are known, but neither the size of the bias nor the variance of the estimator are known. This investigation showed that the Schaefer estimator and Petersen estimator are essentially equivalent - hence there is no reason to prefer the Schaefer over the simpler Petersen.

Petersen Estimator:

In a simple mark-recapture experiment, fish are tagged with individually numbered tags and released back to the population. A second sample is taken, and the number of fish and the number of marked fish recorded. A basic Petersen estimator is formed as:

$$\hat{N}_p = \frac{n_\cdot^c \times n_\cdot^r}{m_{\cdot\cdot}}$$

where

$n_\cdot^c$ is the number of fish initially captured, tagged, and released

$n_\cdot^r$ is the number of fish recovered at the second sample; and

$m_{\cdot\cdot}$ is the number of marked fish recovered in the second sample.

Consistency of the Petersen estimator:

It is well known that the Petersen estimator is consistent only certain conditions - the most common being:

(1) all fish have an equal probability of capture in the first sample;
or (2) all fish have an equal probability of recovery in the second sample;
or (3) the probabilities of capture and recovery are independent over fish.

The approximate relative bias of the Petersen can be expressed as:

$$RB_p \cong -C(tagging, recovery) \times \frac{\sqrt{V(tagging)V(recovery)}}{E(tagging \times recovery)}$$

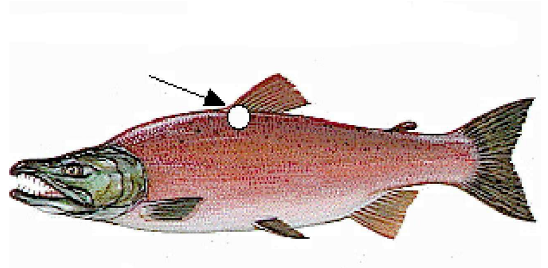
where C, V, and E represent the correlation, variance and expectation of the tagging and recovery probabilities taken over the fish in the population. Conditions (1), (2), and (3) represent a correlation=0 giving “unbiased” estimates.

Precision of the Petersen estimator:

Standard methods show that the coefficient of variation of the Petersen estimator can be approximated by:

$$CV(\hat{N}_p) \cong \frac{1}{\sqrt{E(m_{\cdot\cdot})}} = \frac{1}{\sqrt{N \times E(tagging \times recovery)}}$$

Stratification:



A common problem in fisheries management is to estimate the number of salmon returning to spawn. As the adult fish pass a certain point on the river, they are sampled and tagged over an 8 week period. After reaching the spawning grounds, the fish die and carcasses line the banks. Regular searches of the carcasses over a 9 week period record the number of carcasses searched and the number of marks recovered.

Here is an example considered by Schaefer (1951) where the stratification takes place by week of tagging or week of recovery:

Release week	Tagged	Recovery week 1	2	3	4	5	6	7	8	9
1	15	1	0	2	0	0	0	0	0	0
2	59	1	3	7	0	0	0	0	0	0
3	410	1	11	33	24	5	1	0	1	0
4	695	0	5	29	79	52	3	2	7	3
5	773	0	0	11	67	77	2	16	7	3
6	335	0	0	0	14	25	3	10	6	2
7	59	0	0	0	0	0	0	1	5	0
8	5	0	0	0	0	0	0	1	0	0
Recoveries =>		19	132	800	2848	3476	644	1247	930	376

For example, 59 fish were tagged and released in week 2. Of these 59 tagged fish, 1 was recovered in week 1, 3 in week 2, 7 in week 3 of the recovery periods. A total of 19, 132, and 800 carcasses were examined in recovery weeks 1, 2 and 3 respectively.

The Petersen estimate is formed as:

$$n_\cdot^c = 2{,}351 \quad = 15 + 59 + 410 + \ldots + 5$$
$$n_\cdot^r = 10{,}472 = 19 + 132 + 800 + \ldots + 376$$
$$m_{\cdot\cdot} = 520 = 1 + 0 + 2 + 0 + \ldots.\ 1 + 0 + 0$$

giving

$$\hat{N}_p = \frac{2{,}351 \times 10{,}472}{520} = 47{,}345 \quad se(\hat{N}_p) = 1{,}780 \quad cv(\hat{N}_p) = 4\%$$

The estimate is suspect because the tagging and recovery effort varies over the course of the run and there is obvious incomplete mixing of the tagged fish over the entire run.

Schaefer Estimator

Schaefer (1951) proposed a stratified estimator based on the following representation of the stratified data:

$$\begin{array}{c|cccc|}
n_1^c & m_{11} & m_{12} & \cdots & m_{1t} \\
n_1^c & m_{21} & m_{22} & \cdots & m_{2t} \\
 & \vdots & & & \vdots \\
n_s^c & m_{s1} & m_{s2} & \cdots & m_{st} \\
\hline
 & n_1^r & n_2^r & \cdots & n_t^r
\end{array}$$

where

n_i^c are the number of fish tagged and released in week *i*

n_j^r are the number of fish recovered in week *j*

m_{ij} are the number of fish released in week *i* and recovered in week *j*.

Schaefer’s estimator has the form:

$$\hat{N}_s = \sum_{i=1}^{s}\sum_{j=1}^{t} \frac{n_i^c n_j^r m_{ij}}{m_{i\cdot} m_{\cdot j}}$$

$$\hat{N}_s = 47{,}886 = \frac{15 \times 19 x 1}{3 \times 3} + \frac{15 \times 132 \times 0}{3 \times 19} + \ldots + \frac{5 \times 376 \times 0}{1 \times 8}$$

Both the Schaefer and Petersen estimator are very close. Coincidence? Perhaps the bias in the Petersen is small enough that it can be ignored?

Consistency of the Schaefer estimator:

Chapman and Junge (1956) showed that the Schaefer estimator is consistent under the same conditions as the Petersen.

Bias of the Schaefer estimator:

From a Taylor-series expansion and long-and-tedious algebra, the relative bias of the Schaefer estimator can be approximated as:

$$RB_s \cong -C(tagging, recovery)\frac{\sqrt{V(tagging)V(recovery)}}{E[tagging]E[recovery]}$$

which is very similar to that of the Petersen.

Precision of the Schaefer estimator:

A Taylor-series expansion shows that the coefficient of variation of the Schaefer estimator can be approximated as:

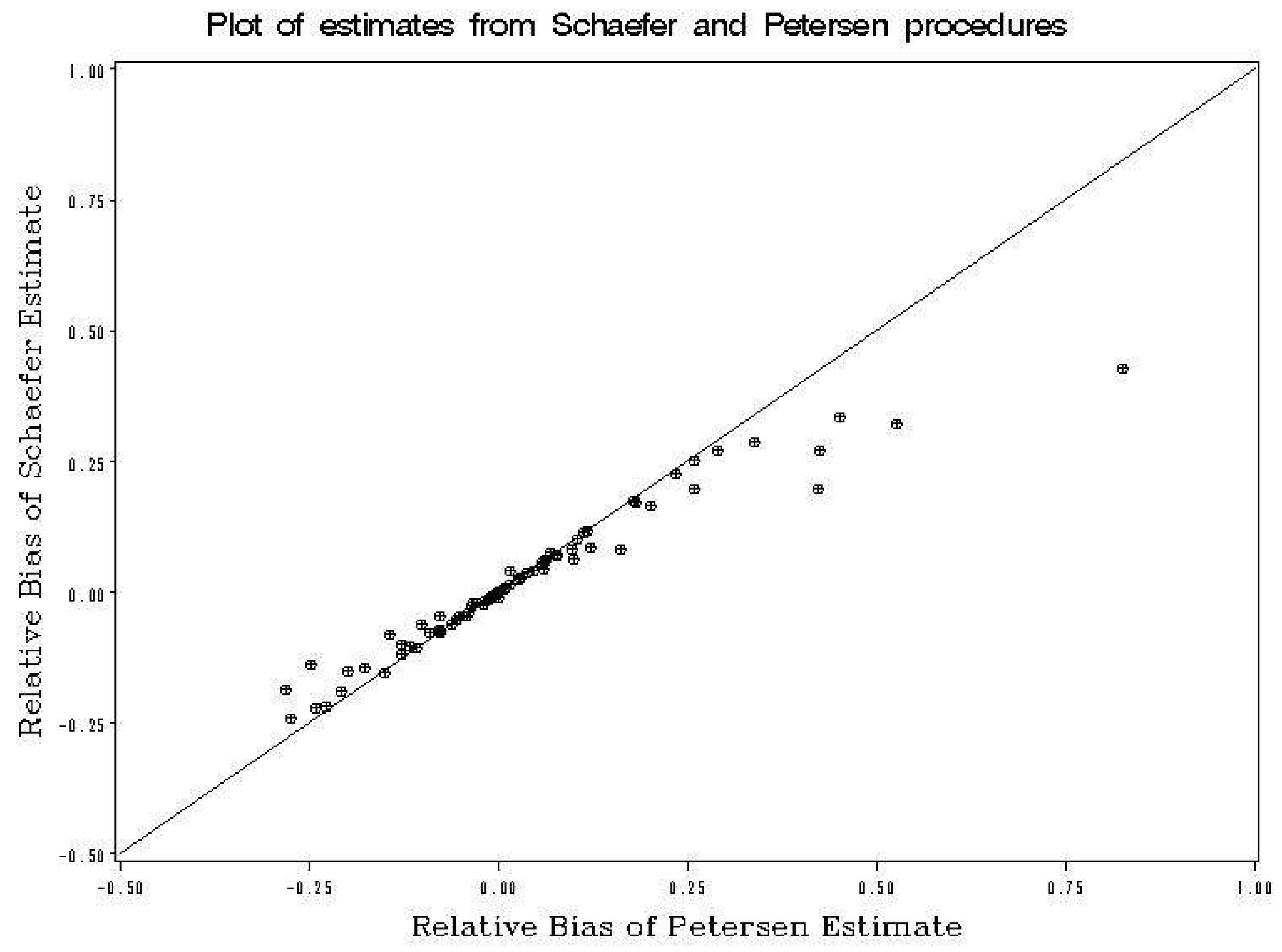
$$CV(\hat{N}_s) \cong \frac{1}{\sqrt{NE[tagging]E[recovery]}}$$

which is similar to that of the Petersen.

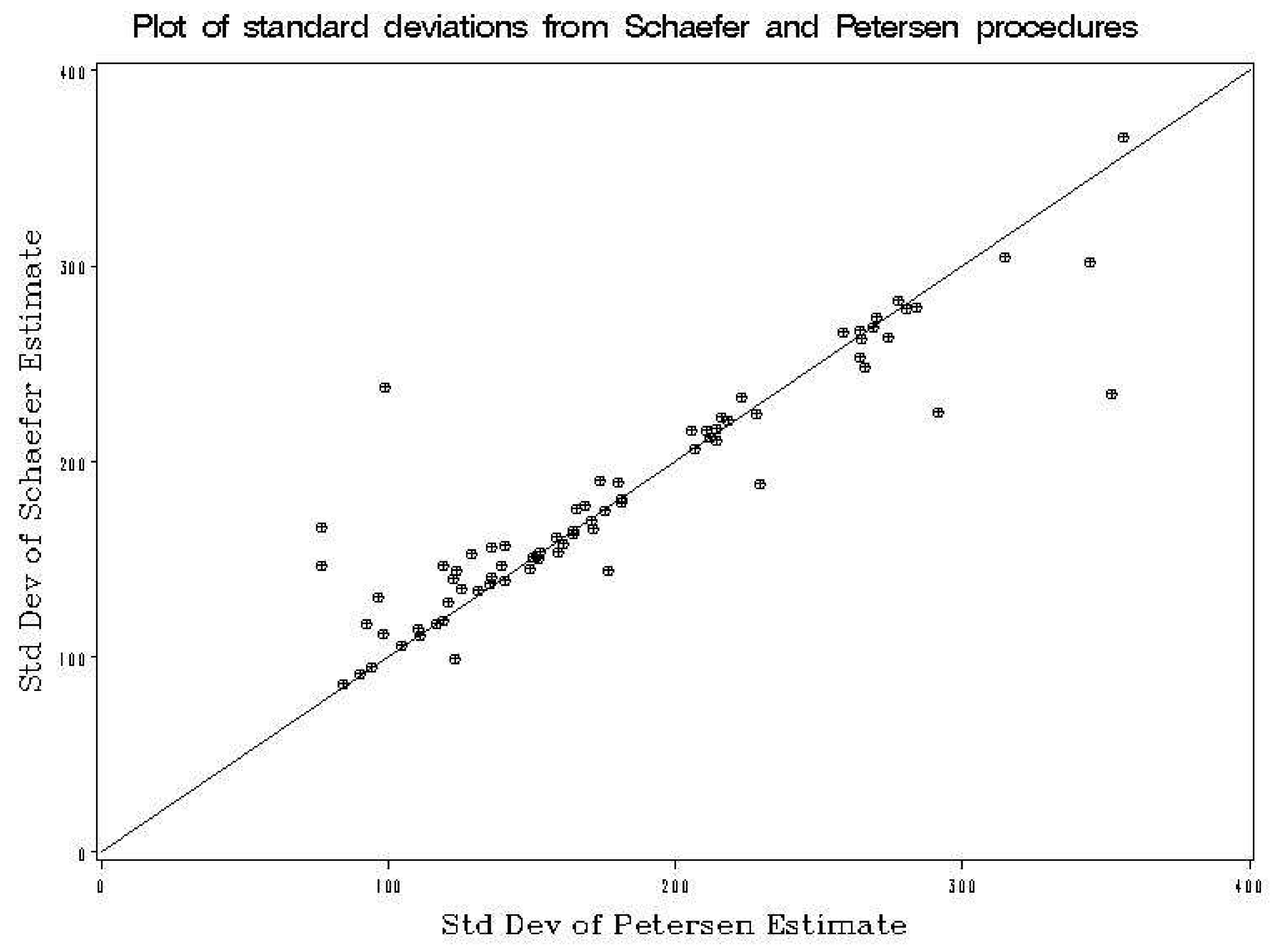
A simple simulation experiment:

A simulation experiment was run to verify that the large sample results obtained from the Taylor series expansion hold in small to moderate sized populations.

Bivariate (tagging,recovery) probabilities were generated with correlations ranging from -0.9 to 0.9; with a range of average capture and recovery probabilities, and a range of coefficients of variation around these average values. For each simulation, the capture and recovery indicator variables were generated using a Bernoulli process. Then both the Petersen and the Schaefer estimator (using a 3x3 stratification based on the capture probabilities) were computed. The relative bias and the standard deviations were computed over the simulations for each combination of population parameters.



And a plot of the standard deviations also shows they are similar.



Implications and Conclusions:

(1) Despite its more complex appearance and attempt to deal with heterogeneity in capture or recovery probabilities the Schaefer estimator is essentially equivalent to the Petersen estimator
  (a) It is requires the same assumptions as the Petersen to be consistent
  (b) It has a very similar size of bias for most studies
  (c) It has similar precision as the Petersen
(2) Consequently there is no good reason to continue to use the Schaefer estimator.
(3) The stratified-Petersen (Darroch, 1961; Plante et al, 1998; Schwarz and Taylor, 1998) remains unbiased under these circumstances. Its estimate is just over 57,000 fish (se 4,000 fish).

References:

Arnason, A.N., Kirby, C.W., Schwarz, C.J. and Irvine, J.R. (1996). Computer analysis of data from stratified mark-recovery experiments for estimating salmon escapement and other population. Canadian Technical Report of Fisheries and Aquatic Sciences 2106.

Chapman, D.G. and Junge, C.O. (195). The estimation of the size of a stratified animal population. Annals of Mathematical Statistics 27, 375-389.

Darroch, J.N. (1961). The two-sample capture-recapture census when tagging and sampling are stratified. Biometrics 48, 241-260.

Plante, N., Rivest, L.-P., and Tremblay, G. (1998). Stratified capture-recapture estimation of the size of a closed population. *Biometrics* **54**, 47-60.

Schaefer, M.B. (1951). Estimation of the size of animal populations by marking experiments. U.S.Fish and Wildlife Bulletin 69, 191-203.

Schwarz, C. J. and Taylor, C. G. (1997). The use of the stratified-Petersen estimator in fisheries management with an illustration of estimating the number of pink salmon (Oncorhynchus gorbuscha) that return to spawn in the Fraser River. Canadian Journal of Fisheries and Aquatic Sciences, 55, 281-296