Ratio estimator for the Sightability Model

Carl James Schwarz

2020-04-07

Contents

1 Introduction 1
2 Implementation 1
3 References 2

1 Introduction

This vignette documents the development of the ratio estimator in the SightabilityModel package (Fieberg, 2012).

The theory for this extension is found in Wong (1996). A portion of her thesis is appended to this document.

2 Implementation

Equation 2.4.2 is the ratio estimator formed as the ratio of estimates of the total for variables $a$ and $b$ (e.g. bulls:cow ratio) that are adjusted using the inverse of sightability ($\Theta_{ij}$) for group $j$ in primary sampling unit $i$. The delta-method is used to find the variance of the ratio as a function of the variances of the estimates of the numerator and denominator and the covariance between the two estimates (Equation 2.4.4). The variance of the numerator and denominator can be obtained using the Sight.Est() function in the SightabilityModel package. So that all remains is to estimate the covariance between the estimate of the numerator and denominator ($\text{Cov}(\hat{\tau}_a, \hat{\tau}_b)$).

The first Equation 2.4.5 gives the formula for the covariance of the numerator and denominator. In order to estimate the covariance, we need estimators for each term, i.e.

$$\text{Cov}(\hat{\tau}_a, \hat{\tau}_b) = \text{estimator}(E[\hat{\tau}_a \hat{\tau}_b]) - \text{estimator}(\tau_a \tau_b)$$

The estimator $(E[\hat{\tau}_a \hat{\tau}_b])$ is simply $\hat{\tau}_a \hat{\tau}_b$. This then gives:

$$\text{Cov}(\hat{\tau}_a, \hat{\tau}_b) = \hat{\tau}_a \hat{\tau}_b - \hat{\tau}_a \hat{\tau}_b$$

where $\hat{\tau}_a \hat{\tau}_b$ is the estimator of $\tau_a \tau_b$. NOTICE THE MINUS SIGN between the two terms.

The population value for the second term above ($\tau_a \tau_b$) is given by the equation between the two Equations 2.4.5 which consists of three terms.
This is combined in the second Equation 2.4.5 and Equation 2.4.6. Note that the equation has an error, where the minus sign above has been dropped, so the proper form of the two estimating equations is:

\[ \hat{\text{Cov}}(\hat{\tau}_a, \hat{\tau}_b) = \hat{\tau}_a \hat{\tau}_b - [\text{term}2 + \text{term}3 + \text{term}4] \]

rather than simply adding the 4 terms together.

This corrections was validated using simulation.

3 References


2.4 Ratio estimator

Samuel et al. (1992) developed a modified Horvitz-Thompson estimator for a population age or sex ratio. Given the same assumptions as section 2.1, define
\[ \tau_a = \text{population size of } a \text{ type animals}, \]
\[ \tau_b = \text{population size of } b \text{ type animals}, \]
\[ \hat{\tau}_{a\pi} = \text{population size estimator of } a \text{ type animals when sighting probabilities are known}, \]
\[ \hat{\tau}_{b\pi} = \text{population size estimator of } b \text{ type animals when sighting probabilities are known}, \]
\[ \hat{\tau}_{aLR} = \text{population size estimator of } a \text{ type animals when sighting probabilities are unknown and are estimated via a logistic model}, \]
\[ \hat{\tau}_{bLR} = \text{population size estimator of } b \text{ type animals when sighting probabilities are unknown and are estimated via a logistic model}, \]
\[ a_{ij} = \text{the number of } a \text{ type animals in the } j\text{th group of } i\text{th primary unit}, \]
\[ b_{ij} = \text{the number of } b \text{ type animals in the } j\text{th group of } i\text{th primary unit}. \]

and with the notation defined in the last section, an estimator for the ratio \( R = \tau_a / \tau_b \), using (2.1.3) when sighting probabilities \( g_{ij} \) are specified, is
\[
\hat{R}_\pi = \frac{\hat{\tau}_{a\pi}}{\hat{\tau}_{b\pi}} = \frac{\sum_{i=1}^{N} \frac{L_i}{\tau_a} \sum_{j=1}^{M_i} Z_{ij} a_{ij} \Theta_{ij}}{\sum_{i=1}^{N} \frac{L_i}{\tau_b} \sum_{j=1}^{M_i} Z_{ij} b_{ij} \Theta_{ij}}.
\]

(2.4.1)

Using (2.1.4) when the \( \Theta_{ij} = 1/g_{ij} \)'s are unknown and are estimated via a logistic model, the ratio estimator is
\[
\hat{R}_{LR} = \frac{\hat{\tau}_{aLR}}{\hat{\tau}_{bLR}} = \frac{\sum_{i=1}^{N} \frac{L_i}{\tau_a} \sum_{j=1}^{M_i} Z_{ij} a_{ij} \hat{\Theta}_{ij}}{\sum_{i=1}^{N} \frac{L_i}{\tau_b} \sum_{j=1}^{M_i} Z_{ij} b_{ij} \hat{\Theta}_{ij}}.
\]

(2.4.2)

By applying the delta method,
\[
\frac{\hat{\tau}_a}{\hat{\tau}_b} \approx \frac{\tau_a}{\tau_b} + \frac{1}{\tau_b} (\hat{\tau}_a - \tau_a) - \frac{1}{\tau_b^2} (\hat{\tau}_b - \tau_b),
\]

(2.4.3)
In (2.4.3), \( \hat{r}_a, \hat{r}_b \) are \( \hat{r}_{a, \pi}, \hat{r}_{b, \pi} \) (see (2.1.2)) if sighting probabilities are known and they are \( \hat{r}_{a, LR}, \hat{r}_{b, LR} \) (see (2.1.4)) if sighting probabilities are unknown. Thus, we obtain an approximate variance of a ratio estimator as

\[
\text{var}(\hat{R}) = R^2 \left[ \frac{\text{var}(\hat{r}_a)}{\tau_a^2} + \frac{\text{var}(\hat{r}_b)}{\tau_b^2} - \frac{2\text{cov}(\hat{r}_a, \hat{r}_b)}{\tau_a \tau_b} \right].
\] (2.4.4)

To obtain an (approximate) estimator of \( \text{var}(\hat{R}) \) in (2.4.4), we replace the ratio the population sizes, variances and covariances of the population size estimator by their respective unbiased estimators. We have already derived (approximate and/or asymptotic) unbiased estimators for the ratio, the population sizes (see (2.1.2) and (2.1.4)) and variances of the population size estimator (see (2.2.2) and (2.2.13)). Therefore, to extend the results of Samuel et. al. (1992), we only need to find an unbiased estimator for the covariance term. To do so, we write

\[
\text{cov}(\hat{r}_a, \hat{r}_b) = E[\hat{r}_a \hat{r}_b] - E[\hat{r}_a]E[\hat{r}_b] = E[\hat{r}_a \hat{r}_b] - \tau_a \tau_b.
\] (2.4.5)

Then, the expression \( \hat{\tau}_a \hat{\tau}_b - \hat{\tau}_{ab} \), where \( \hat{\tau}_{ab} \) is an unbiased estimator of \( \tau_a \tau_b \), is an unbiased estimator of the covariance term in (2.4.5). To find an unbiased estimator of \( \tau_a \tau_b \), we note that

\[
\tau_a \tau_b = \left( \sum_{i=1}^{N} \sum_{j=1}^{M_i} a_{ij} \right) \left( \sum_{i=1}^{N} \sum_{j=1}^{M_i} b_{ij} \right) = \sum_{i=1}^{N} \sum_{j=1}^{M_i} a_{ij} b_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{M_i} b_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{M_i} a_{ij} b_{ij} + \sum_{i=1}^{N} \sum_{j=1}^{M_i} a_{ij} b_{ij}.
\]

Hence, an unbiased estimator for the covariance term in (2.4.5) is

\[
\text{cov}(\hat{r}_a, \hat{r}_b) = \hat{r}_{a, \pi} \hat{r}_{b, \pi} + \sum_{i=1}^{N} \frac{I_i}{\pi_i} \sum_{j=1}^{M_i} \frac{Z_{ij} a_{ij} b_{ij}}{g_{ij}} + \sum_{i=1}^{N} \frac{I_i}{\pi_i} \sum_{j \neq j'}^{M_i} \frac{Z_{ij} Z_{ij'} a_{ij} b_{ij'}}{g_{ij} g_{ij'}} + \sum_{i \neq i'}^{N} \frac{I_i I_{i'}}{\pi_{ii'}} \sum_{j=1}^{M_i} \sum_{j=1}^{M_{i'}} \frac{Z_{ij} Z_{ij'} a_{ij} b_{ij'}}{g_{ij} g_{ij'}}
\] (2.4.5)

for the case of known \( g_{ij} \)'s.

Change this + sign to a minus sign
For the case of unknown \( \Theta_{ij} \)’s estimated via a logistic model, an unbiased estimator for the covariance term in (2.4.5) is

\[
\text{cov}(\hat{\tau}_a, \hat{\tau}_b) = \hat{\tau}_a \hat{\tau}_b L^2 R^2 + \sum_{i=1}^{N} \sum_{j=1}^{M} \frac{I_i}{\pi_i} \sum_{j=1}^{M_j} Z_{ij} a_{ij} b_{ij} \hat{\Theta}_{ij} \\
+ \sum_{i=1}^{N} \sum_{j=1}^{M_j} Z_{ij} Z_{ij'} a_{ij} b_{ij'} [-\text{cov}(\hat{\Theta}_{ij}, \hat{\Theta}_{ij'}) + \hat{\Theta}_{ij} \hat{\Theta}_{ij'}] \\
+ \sum_{i \neq i'} \frac{I_i I_{i'}}{\pi_{ii'}} \sum_{j=1}^{M_j} \sum_{j'=1}^{M_{j'}} Z_{ij} Z_{ij'} a_{ij} b_{ij'} [-\text{cov}(\hat{\Theta}_{ij}, \hat{\Theta}_{ij'}) + \hat{\Theta}_{ij} \hat{\Theta}_{ij'}].
\] (2.4.6)