Plant-Capture Methods for Estimating Population Size from Uncertain Plant Captures Supplementary Materials

Yiran Wang, Martin Lysy, and Audrey Béliveau*

Department of Statistics and Actuarial Science, University of Waterloo

Waterloo, ON, Canada

Contents

A Derivation of MLE for Model $\mathcal{M}_{basic}$

Based on the joint likelihood of the parameters of interest $\boldsymbol{\gamma}$ described in Section 3.1, it is easy to get the log-likelihood as

$$\begin{aligned}
&l(\boldsymbol{\gamma}; y, m^{yes}, m^{mb}, m^{no}) \\
&= m^{yes} \log\{p^c(1-p^{mb})\} + m^{mb} \log p^{mb} + m^{no} \log\{(1-p^c)(1-p^{mb})\} \\
&\quad + \log\{(H+m^{mb})!\} - \log\{(H+m^{mb}-y+m^{yes})!\} \\
&\quad + (y-m^{yes}) \log p^c + (H+m^{mb}-y+m^{yes}) \log(1-p^c)
\end{aligned} \tag{A.1}$$

Taking partial derivatives on Equation (A.1) with respect to p^{mb} and p^c and letting them equal to 0, we arrive at the ML estimator for p^{mb} as $\frac{M^{mb}}{M^{yes}+M^{mb}+M^{no}} = \frac{M^{mb}}{M}$ and the ML estimator for p^c (with known H) as $\frac{Y}{H+M^{yes}+M^{mb}+M^{no}} = \frac{Y}{H+M}$.

To find the ML estimator for H, we consider the ratio

$$\frac{L(H+1)}{L(H)} = \frac{\begin{pmatrix} H+1+m^{mb} \\ y-m^{yes} \end{pmatrix}(1-p^c)^{H+1+m^{mb}-y+m^{yes}}}{\begin{pmatrix} H+m^{mb} \\ y-m^{yes} \end{pmatrix}(1-p^c)^{H+m^{mb}-y+m^{yes}}}$$
$$= \frac{H+1+m^{mb}}{H+1+m^{mb}-y+m^{yes}}(1-p^c) < 1 \text{ when } H > \frac{y-m^{yes}-m^{mb}p^c}{p^c} - 1.$$

This implies a ML estimator of H (when p^c is known) as $\lfloor \frac{Y-M^{yes}-M^{mb}p^c}{p^c} \rfloor$. The ML estimator for p^{mb} does not depend on the other two parameters, but the ML estimators for H and p^c depend on each other. Since the ML estimators need to satisfy all three expressions simultaneously, we can solve for H and p^c to yield the ML estimators for each parameter: $\hat{p^c} = \frac{M^{yes}}{M^{yes}+M^{no}}$, $\hat{p^{mb}} = \frac{M^{mb}}{M}$ and $\hat{H} = \lfloor \frac{Y-M^{yes}-M^{mb}\hat{p^c}}{\hat{p^c}} \rfloor$, where $\lfloor x \rfloor$ denotes the greatest integer less than or equal to x.

B Alternative Computational Methods for Bayesian Inference

We explored two alternative computational approaches to fit our models. These approaches are detailed in this section, and evaluated in a simulation study in Appendix C.

B.1 Bayesian Normal Approximation (BNA)

Bayesian normal approximation is an approach that constructs an approximate representation of the posterior distribution using a multivariate normal distribution. The mean of the distribution is approximated by the vector of posterior mode of the parameters, obtained via numerical optimization of the posterior distribution. The variance-covariance matrix of the distribution is defined as the inverse of the negative Hessian matrix of the log posterior density at the modes. A more comprehensive description of this technique is given in Gelman et al. [2013].

Similarly to the MLE method described in Section 3.1, we can express the posterior distribution of our proposed models in a general form $\pi(\boldsymbol{\gamma}|\boldsymbol{x}) \propto \pi(\boldsymbol{\gamma})P(\boldsymbol{X}=\boldsymbol{x}|\boldsymbol{\gamma}) = \pi(\boldsymbol{\gamma})\sum_{\boldsymbol{z}\in\Omega} P(\boldsymbol{X}=\boldsymbol{x},\boldsymbol{Z}=\boldsymbol{z}|\boldsymbol{\gamma})$, where $\boldsymbol{\gamma}$ denotes the model parameters with a prior distribution $\pi(\boldsymbol{\gamma})$, $\boldsymbol{X}$ represents the data, $\boldsymbol{Z}$ stands for latent variables, and Ω denotes the set of values that $\boldsymbol{Z}$ can take. We also apply a log transformation on the counts and a logit transformation on the probabilities to remove any constraints on their bounds, avoiding computational issues related to boundary constraints. The numerical method to apply this approach is the same as the MLE approach detailed in Section 3.1. Furthermore, the prior settings for the parameters remain the same with those described in Section 4.

B.2 Uncertainty Propagation Method (UP)

While inference for model $\mathcal{M}_{basic}$ is relatively straightforward using MCMC algorithms, the specification of models $\mathcal{M}_{id}$ and $\mathcal{M}_{class}$ using probabilistic programming languages can be relatively complicated because of the equality constraints in Equations (11) and (14), as discussed in Section 3.2. Instead of using the *dsum* function in JAGS, another simple solution is to employ an uncertainty propagation (UP) method to obtain an approximate posterior inference for H. To illustrate this method, we use model $\mathcal{M}_{id}$ as an example.

Initially, it is important to recognize that if p^c and H^c were observed, it would be feasible to construct an approximate representation of the posterior distribution $\pi(H|p^c, H^c)$ using a normal distribution due to the Bernstein–von Mises theorem, as follows:

$$H|H^c, p^c \sim N\left(\hat{H}_0, \frac{\hat{H}_0(1-p^c)}{p^c}\right), \tag{A.2.1}$$

where $\hat{H}_0 = H^c/p^c$ is a variant of the MLE for H [Rukhin, 1975] based on the binomial distribution in Equation (9). A proof of the asymptotic correspondence of $\hat{H}_0$ with the MLE is presented in the box below.

Proof

Suppose we have $H^c \sim \text{Binom}(H, p^c)$. Given p^c and H^c, we use the method introduced in the Appendix of the main paper to derive the MLE of H:

$$\begin{aligned}\frac{L(H+1)}{L(H)} &= \frac{\frac{(H+1)!}{H^c!(H+1-H^c)!}(p^c)^{H^c}(1-p^c)^{H+1-H^c}}{\frac{H!}{H^c!(H-H^c)!}(p^c)^{H^c}(1-p^c)^{H-H^c}} \\ &= \frac{H+1}{H+1-H^c}(1-p^c) < 1 \text{ when } H > \frac{H^c}{p^c} - 1,\end{aligned}$$

which leads to the ML estimator of H as $\lfloor \frac{H^c}{p^c} \rfloor$.

The approximate MLE, $\hat{H}_0$, has variance

$$\text{Var}(\hat{H}_0) = \frac{\text{Var}(H^c)}{(p^c)^2} = \frac{H(1-p^c)}{p^c},$$

which is estimated by $\frac{\hat{H}_0(1-p^c)}{p^c}$ in Equation (A.2.1). Hence, given observed values of p^c and H^c, we could sample from the approximate posterior distribution by sampling directly from Equation (A.2.1).

In practice, while p^c and H^c are not directly observable, we can sample from their posterior distribution conditional on the observed data $\boldsymbol{x}$, $\pi(H^c, p^c|\boldsymbol{x})$. This is possible because p^c and H^c are completely informed when Equation (9) is removed from our model. In fact, the role of Equation (9) is strictly to expand H^c into H via p^c, offering no insights into any model parameters other than H. To summarize, fitting model $\mathcal{M}_{id}$ without Equation (9) and marginalizing over $p^{i|c}, p^{mb}$ and $M^{mb,c}$ provides a posterior sample from $\pi(H^c, p^c|\boldsymbol{x})$. Once an MCMC sample is obtained (first step), the values can be plugged into Equation (A.2.1) to simulate posterior samples from H (second step), which result in the desired approximate posterior distribution $\pi(H|\boldsymbol{x})$. Essentially, uncertainty from the first step is propagated into the second step.

Instead of employing a two-step approach, the desired outcome can be achieved more straightforwardly in a single step by replacing Equation (9) in our model with the approximate normal distribution (A.2.1). This substitution simplifies the MCMC process and enhances computational efficiency. BUGS-based software can thus be used to directly sample from our approximate representation of $\pi(\boldsymbol{\gamma}|\boldsymbol{x})$, which can then be marginalized to sample from $\pi(H|\boldsymbol{x})$.

Finally, when implementing the proposed UP method in BUGS languages, the following grammatical nuance must be considered: Equations (11) and (14) need to be rearranged to shift H^c to the left side. This ensures that the software doesn't interpret H^c as an undefined node.

C Supplementary Tables

Table 1: Results of the simulation studies for $\mathcal{M}_{basic}$ using BNA. All the values are rounded to integers or 2 decimal points.

Method	M	Parameter	True Value	Median	SD	RBias	RRMSE	CP	LCI
		H	150	149	27	-0.01	0.18	0.91	109
BNA	15	p^c	0.7	0.72	0.11	0.03	0.15	0.99	0.43
		p^{mb}	0.2	0.24	0.10	0.19	0.49	0.94	0.39
		H	1,500	1,498	112	0.00	0.08	0.94	442
BNA	100	p^c	0.7	0.70	0.05	0.01	0.07	0.95	0.19
		p^{mb}	0.2	0.21	0.04	0.03	0.20	0.96	0.16

Table 2: Results of the simulation studies for $\mathcal{M}_{id}$ using BNA and UP method. All the values are rounded to integers or 2 decimal points.

Method	M	Parameter	True Value	Median	SD	RBias	RRMSE	CP	LCI
BNA	15	H	150	150	26	0.00	0.17	0.92	105
		p^c	0.7	0.71	0.11	0.02	0.14	0.98	0.42
		$p^{mb\|ni}$	0.2	0.28	0.15	0.39	0.73	0.93	0.55
		$p^{i\|c}$	0.8	0.79	0.04	-0.01	0.05	0.94	0.15
UP	15	$\hat{H}$	150	159	38	0.06	0.22	0.97	144
		$\hat{H}_0$	150	159	37	0.06	0.22	0.96	138
		p^c	0.7	0.68	0.11	-0.03	0.16	0.97	0.43
		$p^{mb\|ni}$	0.2	0.26	0.14	0.30	0.73	0.96	0.52
		$p^{i\|c}$	0.8	0.80	0.04	-0.01	0.05	0.95	0.14
BNA	100	H	1,500	1,500	105	0.00	0.07	0.93	414
		p^c	0.7	0.70	0.05	0.00	0.07	0.94	0.18
		$p^{mb\|ni}$	0.2	0.21	0.06	0.06	0.29	0.96	0.23
		$p^{i\|c}$	0.8	0.80	0.01	0.00	0.02	0.95	0.05
UP	100	$\hat{H}$	1,500	1,512	111	0.01	0.07	0.94	434
		$\hat{H}_0$	1,500	1,513	108	0.01	0.07	0.93	422
		p^c	0.7	0.70	0.05	-0.00	0.07	0.94	0.18
		$p^{mb\|ni}$	0.2	0.21	0.06	0.04	0.29	0.96	0.23
		$p^{i\|c}$	0.8	0.80	0.01	0.00	0.02	0.96	0.05

Table 3: Results of the simulation studies for $\mathcal{M}_{class}$ using BNA and UP method. All the values are rounded to integers or 2 decimal points.

Method	M	Parameter	True Value	Median	SD	RBias	RRMSE	CP	LCI
BNA	30	H	300	294	40	-0.02	0.12	0.99	218
		p_1^c (easy)	0.9	0.86	0.08	-0.04	0.08	0.94	0.33
		p_2^c (hard)	0.4	0.48	0.13	0.19	0.35	0.95	0.48
		$p^{mb\|ni}$	0.2	0.23	0.10	0.14	0.49	0.95	0.41
		$p^{i\|c}$	0.8	0.80	0.03	-0.00	0.03	0.96	0.10
UP	30	$\hat{H}$	300	327	103	0.09	0.21	0.96	323
		$\hat{H}_0$	300	328	100	0.09	0.21	0.95	312
		p_1^c (easy)	0.9	0.84	0.08	-0.06	0.09	0.94	0.32
		p_2^c (hard)	0.4	0.42	0.13	0.04	0.32	0.96	0.49
		$p^{mb\|ni}$	0.2	0.22	0.10	0.09	0.49	0.98	0.38
		$p^{i\|c}$	0.8	0.80	0.03	-0.00	0.03	0.94	0.10
BNA	100	H	1,500	1,486	124	-0.01	0.08	0.98	634
		p_1^c (easy)	0.9	0.90	0.04	-0.00	0.05	0.96	0.17
		p_2^c (hard)	0.4	0.42	0.08	0.06	0.20	0.95	0.29
		$p^{mb\|ni}$	0.2	0.21	0.06	0.06	0.29	0.96	0.23
		$p^{i\|c}$	0.8	0.80	0.01	-0.00	0.02	0.95	0.05
UP	100	$\hat{H}$	1,500	1,538	156	0.03	0.10	0.96	606
		$\hat{H}_0$	1,500	1,538	152	0.03	0.10	0.95	589
		p_1^c (easy)	0.9	0.89	0.04	-0.01	0.05	0.96	0.16
		p_2^c (hard)	0.4	0.41	0.08	0.00	0.19	0.96	0.30
		$p^{mb\|ni}$	0.2	0.21	0.06	0.04	0.29	0.95	0.23
		$p^{i\|c}$	0.8	0.80	0.01	-0.00	0.02	0.95	0.05

Table 4: Estimation of the homeless population size H using the Chapman-Bailey estimator. All the values are rounded to integers.

	Maybe as seen		Maybe as not seen	
City	Estimate	95% CI	Estimate	95% CI
Chicago	7	(3, 20)	42	(23, ∞)
New Orleans	63	(56, 72)	76	(65, 91)
Phoenix	96	(81, 118)	102	(85, 129)
New York	1,520	(1,368, 1,721)	1,670	(1,624, 2,233)
Los Angeles	257	(212, 335)	289	(231,402)

References

A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. *Bayesian data analysis*. CRC press, 3rd edition, 2013.

A. Rukhin. Statistical decision about the total number of observable objects. *Sankhyā: The Indian Journal of Statistics, Series A*, 37(4):514–522, 1975.