Plant-Capture Methods for Estimating Homeless Population Size From Uncertain Plant Captures

Yiran Wang Martin Lysy Audrey Béliveau
Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, ON, Canada

Introduction

Plant-capture methods estimate homeless population size H using a single sampling occasion and marked "plants" M, whose reported target status may be uncertain (Yes/Maybe/No) [1, 2, 3]. We develop three models to address uncertain plant captures, partial interview identification, and site-class heterogeneity.

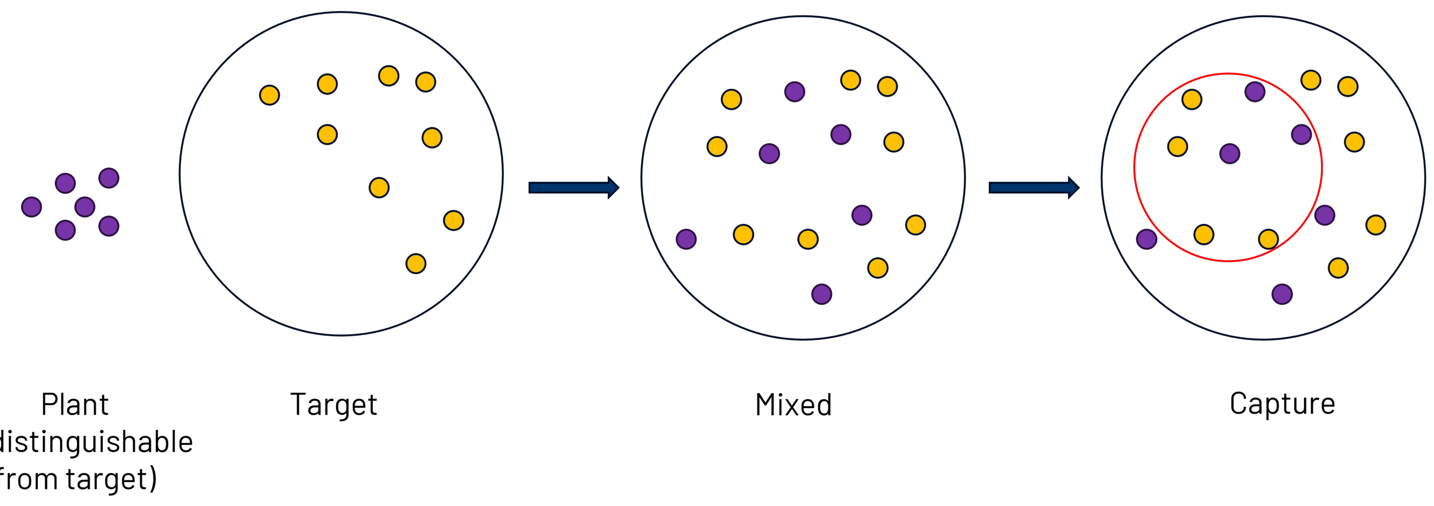

Figure 1. Schematic of plant-capture sampling: introducing indistinguishable plants (M) into the target population (H).

Model $\mathcal{M}_{basic}$: Basic Framework

$$(M^{yes}, M^{mb}, M^{no}) \mid M \sim \text{Multinom}(M; p^{yes}, p^{mb}, p^{no}) \quad (1)$$
$$M^{mb,c} \mid M^{mb} \sim \text{Binom}(M^{mb}, p^{c|mb}) \quad (2)$$
$$H^{c} \mid H \sim \text{Binom}(H, p^{c}) \quad (3)$$
$$Y = M^{yes} + M^{mb,c} + H^{c}. \quad (4)$$

- **Assumption:** $p^{c|mb} = p^{c} \implies p^{yes} = p^{c}(1-p^{mb}), p^{no} = (1-p^{c})(1-p^{mb})$.

	Captured	Not Captured
Maybe	$M^{mb,c}$	$M^{mb} - M^{mb,c}$
Not Maybe	M^{yes}	M^{no}

Table 1. Partitioning of plant counts under the basic model.

Model $\mathcal{M}_{id}$: Partial Identification

When some individuals are identified via interview:

$$(M^{i}, M^{yes}, M^{mb}, M^{no}) \mid M \sim \text{Multinom}(M; \boldsymbol{\theta}_{\text{II}}) \quad (5)$$
$$M^{mb,c} \mid M^{mb} \sim \text{Binom}(M^{mb}, p^{c|mb,ni}) \quad (6)$$
$$H^{c} \mid H \sim \text{Binom}(H, p^{c}) \quad (7)$$
$$H^{i} \mid H^{c} \sim \text{Binom}(H^{c}, p^{i|c}) \quad (8)$$
$$Y = M^{i} + M^{yes} + M^{mb,c} + H^{c}. \quad (9)$$

- **Assumption (non-identified):** $p^{c|mb,ni} = p^{c}(1-p^{i|c})/\{p^{c}(1-p^{i|c}) + (1-p^{c})\}$.

	Captured	Not Captured
Identified	M^{i}	
Maybe	$M^{mb,c}$	$M^{mb} - M^{mb,c}$
Not Maybe	M^{yes}	M^{no}

Table 2. Partitioning of observed and latent plant counts under capture status uncertainty.

Model $\mathcal{M}_{class}$: Site Heterogeneity

To account for the heterogeneous catchability between sites due to factors such as visual barriers, drug activities, time constraints, or enumerator behavior:

$$(M_k^{i}, M_k^{yes}, M_k^{mb}, M_k^{no}) \mid M_k \sim \text{Multinom}(M_k; \boldsymbol{\theta}_{\text{III},k}) \quad (10)$$
$$M_k^{mb,c} \mid M_k^{mb} \sim \text{Binom}(M_k^{mb}, p_k^{c|mb,ni}) \quad (11)$$
$$H_k^{c} \sim \text{Binom}(H_k, p_k^{c}) \quad (12)$$
$$H_k^{i} \sim \text{Binom}(H_k^{c}, p^{i|c}) \quad (13)$$
$$Y_k = M_k^{i} + M_k^{yes} + M_k^{mb,c} + H_k^{c}. \quad (14)$$

- **Assumption:** $p_k^{c|mb,ni} = p_k^{c}(1-p^{i|c})/\{p_k^{c}(1-p^{i|c}) + (1-p_k^{c})\}$, for $k = 1, \dots, K$.

Estimation Methods

Frequentist: Maximum Likelihood Estimation

For Model $\mathcal{M}_{class}$, the MLE of parameters of interest can be obtained directly. For Models $\mathcal{M}_{id}$ and $\mathcal{M}_{class}$, the models can be written in a general form:

$$L(X|\gamma) = \sum_{Z \in \Omega} L(X, Z|\gamma),$$

where γ represents parameters, X represents data, and Z represents latent variables. By summing over $Z \in \Omega$, we efficiently remove latent variables from the likelihood.

Bayesian: MCMC & Uncertainty Propagation

MCMC via JAGS [4]:

- **Priors:** $p \sim$ Uniform$(0, 1)$ for probabilities; $H \sim$ Log-normal$(0, 100)$.
- **Setup:** 3 chains, 30,000 iterations, 15,000 burn-in.

Bayesian Normal Approximation:

$$\pi(\gamma|x) \propto \pi(\gamma) \sum_{Z \in \Omega} L(X, Z|\gamma)$$

Uncertainty Propagation for H:

$$H|H^{c}, p^{c} \sim N\left(\hat{H}_0, \hat{H}_0(1-p^{c})/p^{c}\right),$$

where $\hat{H}_0 = H^{c}/p^{c}$ is a variant of the MLE for H [5].

Simulation Results

Model	Method	Avg	RBias	RRMSE	CP
$\mathcal{M}_{basic}$	MLE	149	-0.01	0.24	0.85
	MCMC	159	0.06	0.24	0.97
$\mathcal{M}_{id}$	MLE	150	0.00	0.22	0.88
	MCMC	159	0.06	0.22	0.97

Table 3. Small City Scenario ($M = 15, H = 150$).

Model	Method	Avg	RBias	RRMSE	CP
$\mathcal{M}_{basic}$	MLE	1,497	-0.00	0.08	0.93
	MCMC	1,513	0.01	0.08	0.94
$\mathcal{M}_{id}$	MLE	1,498	-0.00	0.07	0.93
	MCMC	1,512	0.01	0.07	0.94

Table 4. Large City Scenario ($M = 100, H = 1500$).

Application: 1990 S-Night Survey

Reconstructing homeless population size H for 5 U.S. cities using model $\mathcal{M}_{id}$ [6, 7]:

	MCMC		MLE	
City	Estimate	(95% CrI)	Estimate	(95% CI)
New York	1,709	(1,494, 2,056)	1,688	(1,450, 1,964)
New Orleans	70	(61, 87)	69	(58, 82)
Phoenix	102	(87, 135)	98	(80, 120)
Los Angeles	290	(233, 415)	282	(215, 372)
Chicago	37	(11, 156)	54	(13, 217)

Table 5. Estimated population size H with 95% intervals.

Discussion & Conclusions

Policy Implications

Accurate quantification of vulnerable populations is instrumental in shaping governmental policies and housing provisions. Our framework provides a robust tool for evaluating public health interventions despite data uncertainty.

Methodological Considerations

The MAR assumption is a crucial identifying restriction for "maybe" responses. We recommend applying $\mathcal{M}_{class}$ primarily when sufficient plant samples are available in each stratum to ensure estimator precision.

Future Directions

Future refinements could incorporate spatio-temporal components using GPS data from mobile applications to account for mobility. Additionally, hierarchical priors can relax strict MAR assumptions by treating capture probabilities as exchangeable.

References

[1] EM Laska and M Meisner. A plant–capture method for estimating the size of a population from a single sample. *Biometrics*, 49(1):209–220, 1993.

[2] United States Department of Housing and Urban Development. A guide to counting unsheltered homeless people, 2nd revision, 2008.

[3] National Law Center on Homelessness & Poverty. *Don't Count on It: How the HUD Point-in-Time Count Underestimates the Homelessness Crisis in America*. Washington, DC, 2017.

[4] M Plummer. JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, pages 1–10. Vienna, Austria, 2003.

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

[6] E Martin, EM Laska, K Hopper, and J Wanderling. Issues in the use of a plant–capture method for estimating the size of the street dwelling population. *Journal of Official Statistics*, 13(1):59, 1997.

[7] E Martin. Assessment of S-Night street enumeration in the 1990 census. *Evaluation Review*, 16(4):418–438, 1992.

Acknowledgements

The authors thank Professor Ruth King (The University of Edinburgh) for helpful discussions.

Notes

Current affiliation of Yiran Wang: Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada.