Exploring MCMC Samplers for N-Mixture Models

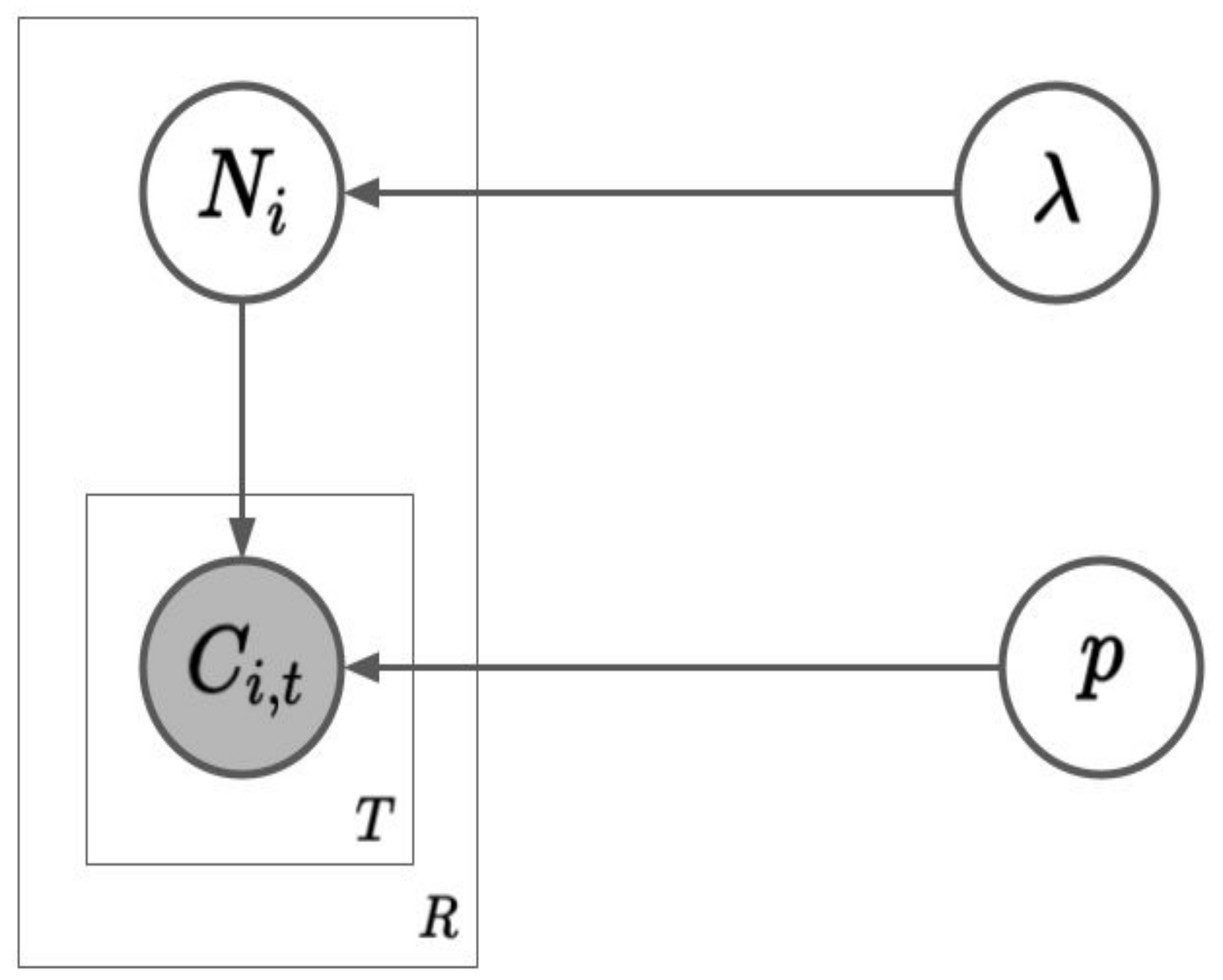
Angela Zhu and Brayden Edwards

Goal: learn the mechanistic behavior of Markov Chain Monte Carlo (MCMC) samplings algorithms used for N-Mixture models[1].

Plate Diagram of Model

True Abundance
N_i ~ Poisson(λ)

Expected Abundance
(static)



Observed Data
$C_{i,t}$ ~ Binomial(N_i, p)

Detection Probability
(static)

Problem Formulation

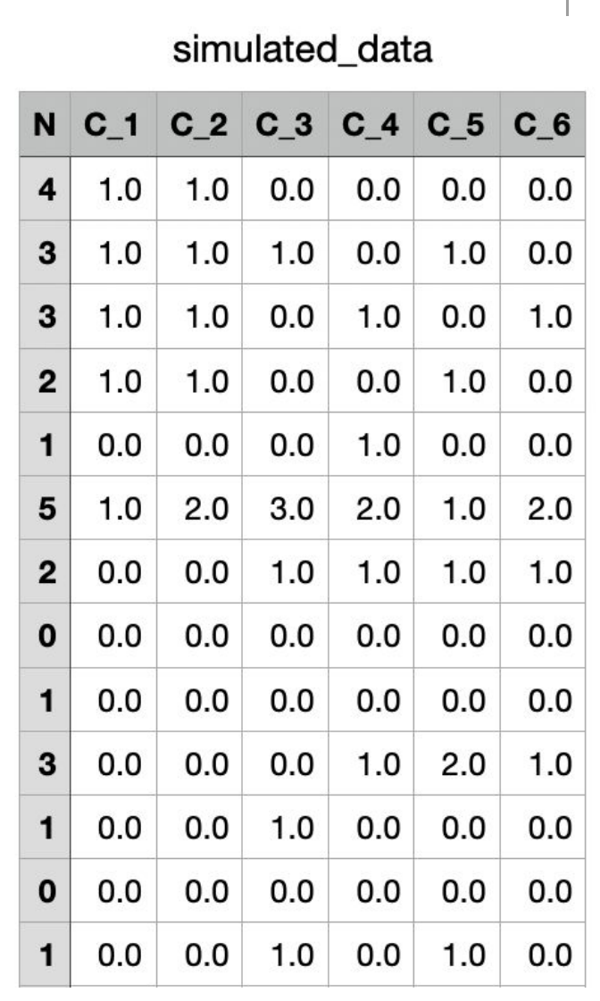
Step 1: Generate Simulation Data[2]

1. Set State Parameters → 2. Sample True Abundance → 3. Generate Observed Counts

Sites, **R** = 20
Visits, **J** = 6
λ = 5
p = 0.25

N_i ~ Poisson(λ)
For each site r

$C_{i,t}$ ~ Binomial(N_i, p)
For each site r and visit j



simulated_data

N	C_1	C_2	C_3	C_4	C_5	C_6
4	1.0	1.0	0.0	0.0	0.0	0.0
3	1.0	1.0	1.0	0.0	1.0	0.0
3	1.0	1.0	0.0	1.0	0.0	1.0
2	1.0	1.0	0.0	0.0	1.0	0.0
1	0.0	0.0	0.0	1.0	0.0	0.0
5	1.0	2.0	3.0	2.0	1.0	2.0
2	0.0	0.0	1.0	1.0	1.0	1.0
0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	1.0	2.0	1.0
1	0.0	0.0	1.0	0.0	0.0	0.0
0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	1.0	0.0	1.0	0.0

Step 2: Estimate Parameters with Metropolis Hastings

1. Generate Sample → 2. Calculate Acceptance Score → 3. Accept Sample → 4. Estimate Parameters

Sample X' | C_i

N= 2
p = 0.2
λ = 3

Given the joint distribution and proposal distribution,

$$\alpha = \min\left(1, \frac{\pi(x')\mathcal{T}(x' \to x)}{\pi(x)\mathcal{T}(x \to x')}\right)$$

Choose to accept or reject sample.

$\alpha \geq U$~ Uniform

Use samples to estimate parameters
N, λ, P

Method

Method	Proposal for p and λ	Proposal for N_i	Key Property
1.Uniform Proposal	p' ~ Uniform(0,1) λ' ~ Uniform(1, S)	N_i ~ Uniform(1,S)	Symmetric proposal.
2 Uniform Priors only	p' ~ Uniform(0,1) λ' ~ Uniform(1, S)	N_i ~ Poisson(λ)	Incorporates ecological structure.
3.Random Walk (RW)	p' ~ N(p, σ^2_p) λ' ~ N(λ, σ^2_λ)	N'_i~ Poisson(N_i)	Smaller, local steps rather than global jumps.
4. RW with Component-Wise Updates	p' ~ N(p, σ^2_p), λ' ~ N(λ, σ^2_λ)	N'_i ~ Poisson(N_i)	Each sample updates one parameter at a time.
5. RW with Truncated Normals	p' ~ N_truncated(p, σ^2_p) λ' ~ N_truncated(λ, σ^2_λ)	N'_i ~ Poisson(N_i)	Accounts for truncation in acceptance probability.
Baseline NIMBLE	Random Walk Sampling	Slice Sampling	Bayesian inference package. 3 chains, 40k iterations, 10k burn-in.

Experiments on Simulated Data

Baseline parameters: R = 20 sites & T = 6 visits. True Population 107.

- All methods exhibit a λ-p tradeoff: overestimating lambda leads to underestimating p
- Method 4 achieves the lowest population error (AE=292.), outperforming NIMBLE (AE = 32.72).

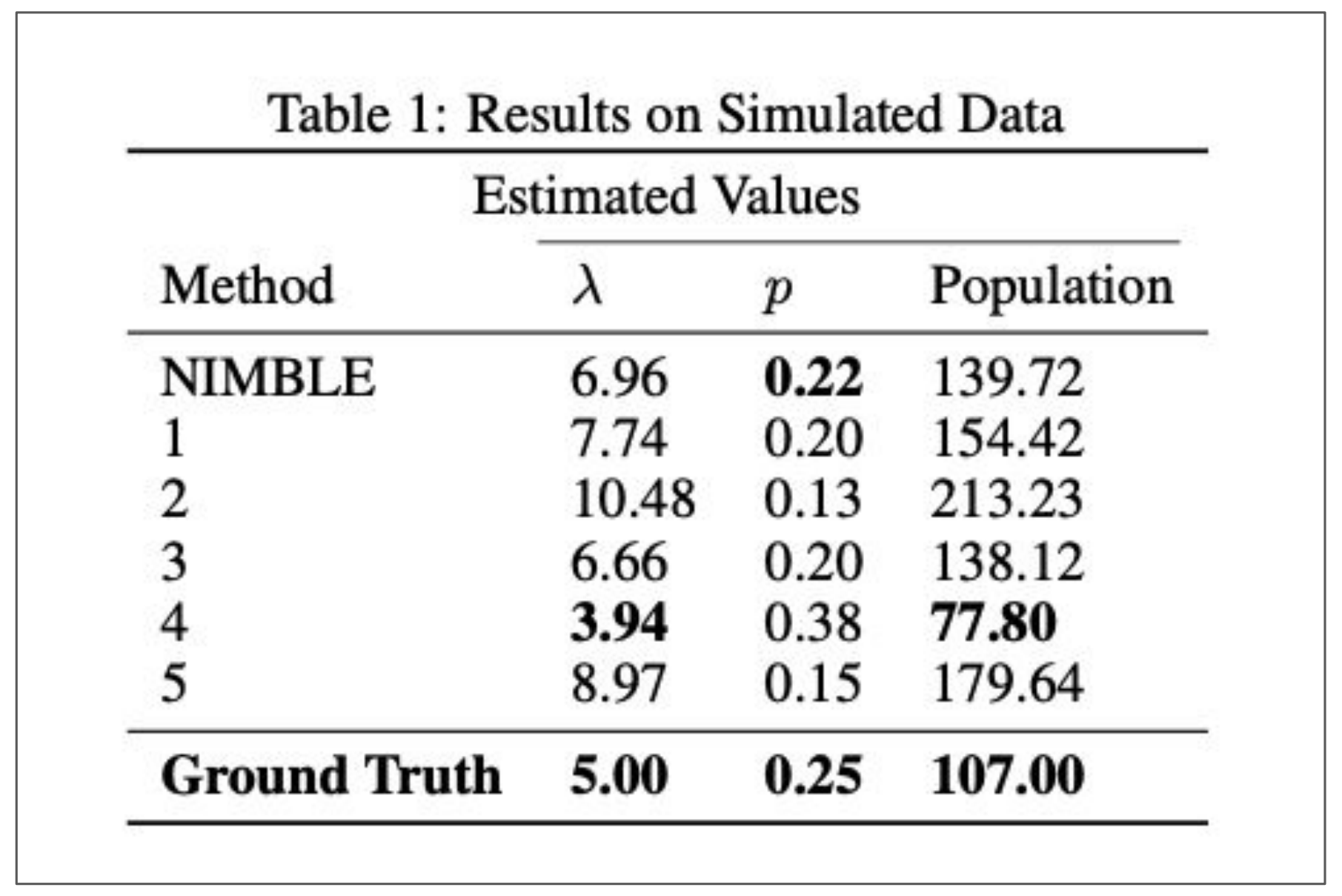


Table 1: Results on Simulated Data

	Estimated Values		
Method	λ	*p*	Population
NIMBLE	6.96	**0.22**	139.72
1	7.74	0.20	154.42
2	10.48	0.13	213.23
3	6.66	0.20	138.12
4	**3.94**	0.38	**77.80**
5	8.97	0.15	179.64
Ground Truth	**5.00**	**0.25**	**107.00**

Experiments on Real World Data

Mallard data from the 2005 Swiss Survey[3] of common birds.
Parameters: 191 sites, 3 visits, *85.69% zero observations.*

- Custom methods overestimate population. Our methods assume global λ and p
- Best: Method 1 makes broad, uniform sweeps, better explore the posterior under sparse conditions, while RW gets stuck in low density regions

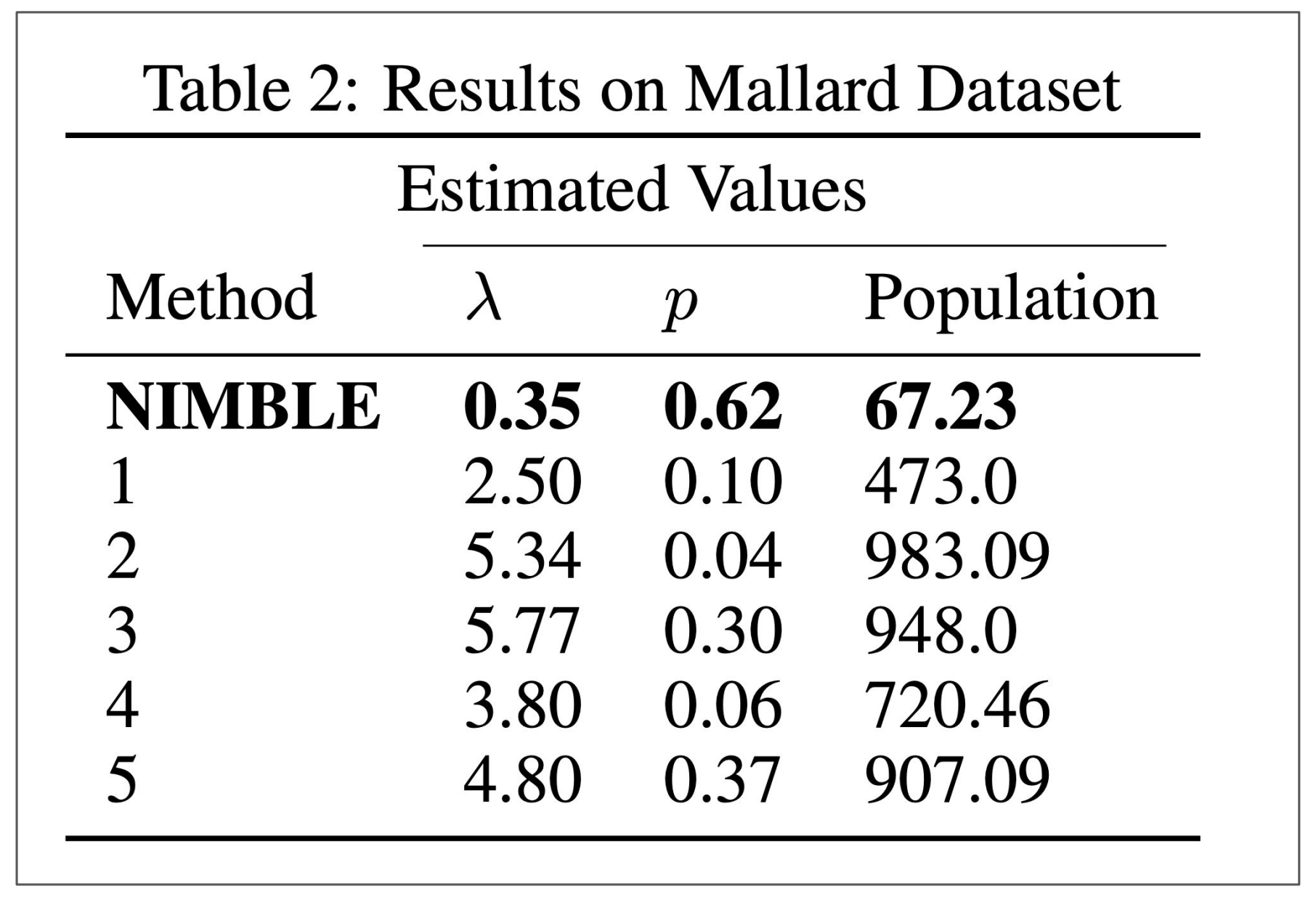
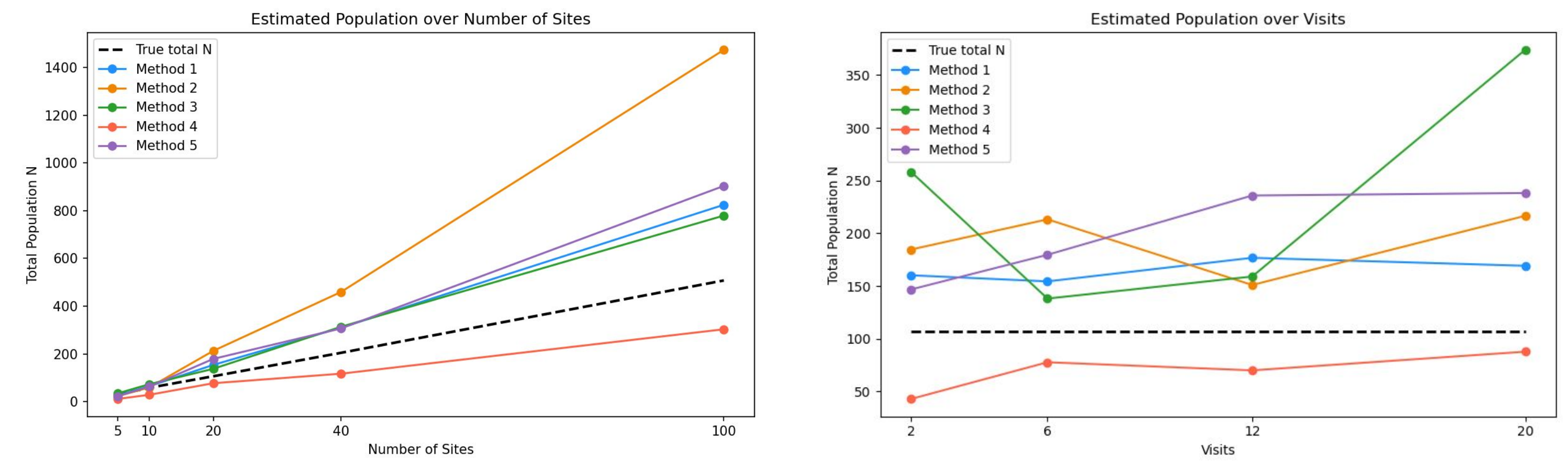


Table 2: Results on Mallard Dataset

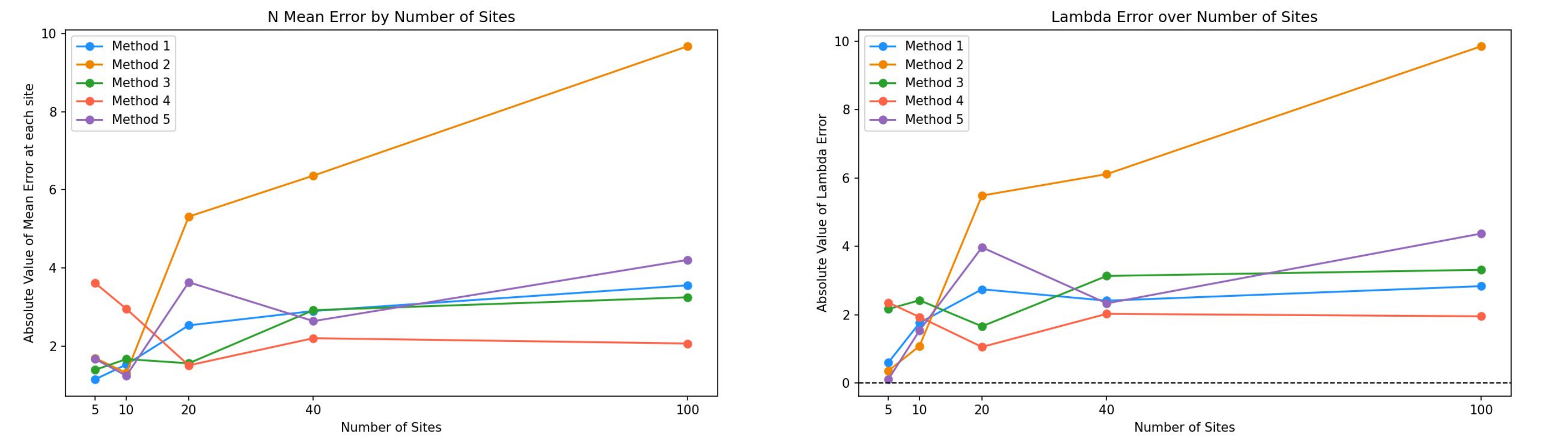
	Estimated Values		
Method	λ	*p*	Population
NIMBLE	**0.35**	**0.62**	**67.23**
1	2.50	0.10	473.0
2	5.34	0.04	983.09
3	5.77	0.30	948.0
4	3.80	0.06	720.46
5	4.80	0.37	907.09

Results

Population Estimates while varying Sites vs Visits

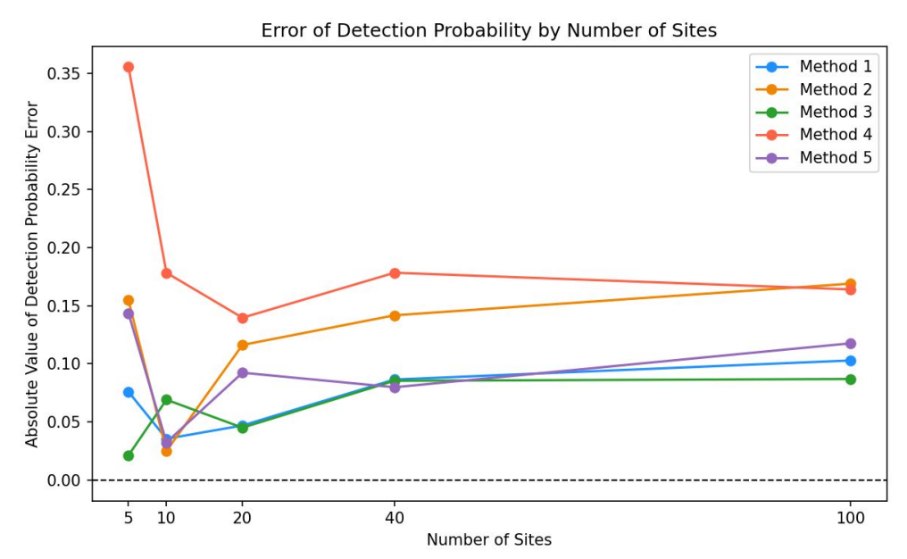


Error in λ Correlated with Error in N (Total Population)



Overall Findings

- Performance depends heavily on data and hyperparameters
- More visits → better estimates? convergence to true total abundance did not occur, likely due to a sampling limitation.



Best Overall: Method 4 component wise updates improve acceptance rate and stability. Future direction: component-wise version of Method 5.

Conclusion

Ecological Research Connection.

Sparse data is common in ecological studies, but models want more data, specifically more detections.

- Detections are more informative than non-detections, truncation of search space

Future Directions

1. Update Sampling Methodology
   - Tailoring to sparse data: Adding truncations to uniform proposals "prior knowledge"
   - More advanced MCMC: Slice, Adaptive MH, Hamiltonian
2. Adding State and Detection Covariates to λ and P

References

1. Royle, J. Andrew. "N-mixture models for estimating population size from spatially replicated counts." *Biometrics* 60.1 (2004).
2. Kéry, M. & Royle, J.A. (2016). *Applied Hierarchical Modeling in Ecology*, Vol. 1. Academic Press. Ch. 6.2–6.9.
3. Kéry, Marc, J. Andrew Royle, and Hans Schmid. "Modeling avian abundance from replicated counts using binomial mixture models." *Ecological applications* 15.4 (2005).

Code and Paper are at https://github.com/angelazhu2/nmixture_model

Note = this is not novel work, we are reproducing work that has already been done