Online Supplement

Greater Than the Sum of its Parts:
Computationally Flexible Bayesian Hierarchical Modeling

S1. Deterministic Bayesian inference

To this point we have really only considered using the posterior mode and Hessian derived covariance matrix for the first stage normal approximation. For a large amount of data in the first stage, this might be just fine. But, we might consider using the posterior mean and variance of $\boldsymbol{\theta}_i$ instead. If MCMC is used in the first stage, we calculate the sample mean and covariance matrix. If this is the case, one might use the two-stage MCMC procedures proposed and discussed by Hooten et al. (2021), Lunn et al. (2013), and Goudie et al. (2019). If the number of parameters is relatively small, say ≤ 6, we propose using a deterministic sampling procedure for approximating the posterior mean and variance (see Johnson et al. 2011).

The deterministic procedure proceeds as follows for each i,.

1. Maximize $[\boldsymbol{\theta}|\mathbf{y}]$, to obtain the posterior mode, $\hat{\boldsymbol{\theta}}$ and covariance matrix, $\hat{\mathbf{S}}$
2. Form the Eigen decomposition of the covariance matrix $\hat{\mathbf{S}} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}'$
3. Explore the principle axes of $[\boldsymbol{\theta}|\mathbf{y}]$ via the parameterization $\boldsymbol{\theta}^{(j)} = \hat{\boldsymbol{\theta}} + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\mathbf{z}$, where $\mathbf{z}$ is successively incremented from $\mathbf{0}$ one entry at a time by step length, say δ_z until $\log[\hat{\boldsymbol{\theta}}|\mathbf{y}] - \log[\boldsymbol{\theta}^{(j)}|\mathbf{y}] > \delta_\pi$.
4. Repeat by successively incrementing each element of $\mathbf{z}$ by $-\delta_z$.

5. Finally, form a grid with every combination of the dimensionality $\boldsymbol{\theta}^{(j)}$ entries saved in the previous two steps and retain with the same δ_π criterion.

6. Form weights $w_j \propto [\boldsymbol{\theta}^{(j)}|\mathbf{y}]$

The main benefit of the deterministic approach is that it avoids having to select a proposal distribution and assess chain convergence. The drawbacks of this approach, however, are the unknown number of likelihood evaluations and the potential coarseness in high density areas. As the number of parameters becomes large this method quickly succumbs to the curse of dimensionality.

S2. Additional details and results for Example II

S2.1. Model Details

Ring-recovery models are based on marking animals and releasing them back into the wild. In subsequent years, the marked animals are then recovered after they have died. In this study lapwings were released and recovered annually from 1963–1997. The likelihood for this type of model is a product-multinomial where numbers of ring-recoveries in the following years are multinomial distributed with cell probabilities

$$p_{it} = \begin{cases} (1-\phi_{t-1})\lambda_t & \text{for } t = i+1 \\ (1-\phi_{t-1})\lambda_t \prod_{t'=I}^{t-2} \phi_{t'} & \text{for } i+2 < t \leq T \\ 1 - \sum_{t'=i}^{J} p_{it'} & \text{for } t = T+1 \end{cases},$$

where p_{it} is the probability that an animal ringed in year i is recovered in year t. The last year of the study is denoted with T, ($t = T+1$ is for animals never recovered), λ_t is the probability of recovering an animal in year t given it died over the previous year, and ϕ_t is the probability of surviving from year t to $t+1$.

To assess the influence of weather on survival and trends in recovery, the ring-recovery parameters are modeled with

$$\begin{gathered} \text{logit}(\phi_{yt}) = \eta_{y0} + \eta_{y1}x_t, \qquad \text{logit}(\phi_{at}) = \eta_{a0} + \eta_{a1}x_t, \\ \text{logit}(\lambda_t) = \gamma_{\lambda 0} + \gamma_{\lambda 1}t, \end{gathered} \tag{1}$$

where ϕ_{yt} is the survival of first year birds in year t (for animals ringed in year $t-1$), ϕ_{at} is adult female survival in year t, and x_t is the number of days below freezing in year t.

The census index data consist of noisy measures of female adult lapwing abundance in the study area, $\mathbf{y}_2$. These abundance measurements exist from 1965–1998. To make inference about population dynamics we use the state-space model of Besbeas et al. (2002),

$$
\begin{aligned}
N_{y,t+1} &= \phi_{yt}\rho_t N_{at} + \epsilon_{yt}; \qquad \varepsilon_{yt} \sim \mathrm{N}(0, \phi_{yt}\rho_t N_{at}), \\
N_{a,t+1} &= \phi_{at}(N_{yt} + N_{at}) + \epsilon_{at}; \qquad \varepsilon_{at} \sim \mathrm{N}(0, \phi_{at}(1-\phi_{at})N_{at}), \\
y_t &\sim \mathrm{N}(N_{at}, \sigma^2),
\end{aligned} \tag{2}
$$

where N_{yt} is the number of yearling females in year t, N_{at} is the number of adult females in year t, ρ_t is the rate at which female offspring are produced in year t, and y_t are the noisy observations of adult female abundance. The survival parameters are the same as the ring-recovery model, but the production is modeled on the log scale using

$$\log(\rho_t) = \gamma_{\rho 0} + \gamma_{\rho 1} t.$$

Adult birds are the only component of the bivariate abundance state that is (indirectly) observed, so, there is little information in these data to inform all survival and production parameters. The IPM melds the information in the two data sets to allow inference to be made on all the parameters of the joint model,

$$\boldsymbol{\eta} = (\eta_{y0}, \eta_{y1}, \eta_{a0}, \eta_{a1})',$$

$$\boldsymbol{\gamma}_1 = (\gamma_{\lambda 0}, \gamma_{\lambda 1})', \text{ and } \boldsymbol{\gamma}_2 = (\gamma_{\rho 0}, \gamma_{\rho 1}, \log\sigma, \log N_{y0}, \log N_{a0})'.$$

Random unit-level $\boldsymbol{\theta}$ parameters are not traditionally used, but see the analysis in Section 4.2 for an alternate version.

S2.2 Tabulated parameter estimates

Table S1: Full results for Example II: Integrated data modelThe B&M 2019 results were taken from Table 2 of Besbeas and Morgan (2019) and represent estimates from MCMC analysis of the full model. In the 3-stage RE model, $\boldsymbol{\theta}_i \sim \mathrm{N}(\boldsymbol{\eta}, \sigma_\theta \mathbf{I})$.

Parameter	B&M 2019	2-stage	3-stage	3-stage RE
η_{y0}	0.523 (0.067)	0.519 (0.068)	0.519 (0.068)	0.511 (0.070)
η_{y1}	-0.023 (0.007)	-0.024 (0.007)	-0.024 (0.007)	-0.03 (0.019)
η_{a0}	1.521 (0.070)	1.500 (0.068)	1.500 (0.068)	1.496 (0.070)
η_{a1}	-0.028 (0.005)	-0.028 (0.005)	-0.028 (0.005)	-0.017 (0.014)
$\gamma_{\lambda 0}$	-4.563 (0.035)	-4.566 (0.035)	-4.566 (0.035)	-4.568 (0.035)
$\gamma_{\lambda 1}$	-0.584 (0.064)	-0.582 (0.065)	-0.582 (0.065)	-0.573 (0.065)
$\gamma_{\rho 0}$	-1.178 (0.091)	-1.175 (0.087)	-1.115 (0.086)	-1.128 (0.091)
$\gamma_{\rho 1}$	-0.425 (0.076)	-0.388 (0.081)	-0.428 (0.077)	-0.456 (0.087)
$\log \sigma$	5.049 (0.136)	4.884 (0.127)	5.052 (0.138)	4.965 (0.152)
$\log N_{y0}$	5.966 (0.546)	6.039 (0.447)	5.985 (0.547)	6.001 (0.494)
$\log N_{a0}$	7.015 (0.135)	7.019 (0.114)	7.016 (0.137)	7.019 (0.126)
$\log \sigma_\theta$				-4.062 (0.519)

References

Besbeas, P., Freeman, S. N., Morgan, B. J., and Catchpole, E. A. (2002). Integrating mark–recapture–recovery and census data to estimate animal abundance and demographic parameters. *Biometrics*, 58:540–547.

Goudie, R. J., Presanis, A. M., Lunn, D., De Angelis, D., and Wernisch, L. (2019). Joining and splitting models with Markov melding. *Bayesian Analysis*, 14:81–109.

Hooten, M. B., Johnson, D. S., and Brost, B. M. (2021). Making recursive Bayesian inference accessible. *The American Statistician*, 75:185–194.

Johnson, D. S., London, J. M., and Kuhn, C. E. (2011). Bayesian inference for animal space use and other movement metrics. *Journal of Agricultural, Biological, and Environmental Statistics*, 16(3):357–370.

Lunn, D., Barrett, J., Sweeting, M., and Thompson, S. (2013). Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(4):551–572.