Heidi Swanson, Martin Lysy, Michael Power, Ashley Stasko, Jim Johnson, and James Reist. 2014. A new probabilistic method for quantifying n-dimensional ecological niches and niche overlap. *Ecology*

APPENDICES

Appendix C. Sampling algorithms for Bayesian posteriors and discussion of posteriors.

Appendix C. Sampling Algorithms for Bayesian posteriors and discussion of priors

Let $X = (X_1, \dots, X_N)$ be N iid observations of a multivariate Normal distribution,

$$X_i = (X_{i1}, \dots, X_{iN}) \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \Sigma),$$

with unknown μ and Σ to be estimated from the data. Here we provide two specifications of the prior distribution $\pi(\mu, \Sigma)$ which lead to simple and effective sampling algorithms for the posterior distribution $p(\mu, \Sigma \,|\, X)$.

C.1. Independent μ and Σ. It is possible to take this one step further by specifying a noninformative prior for Σ: $\pi(\Sigma) = \pi(\Sigma \,|\, \mu) \propto |\Sigma|^{(\nu+n+1)/2}$. The following prior assumes that, in the absence of any data, knowing μ has no bearing on the uncertainty about Σ and vice versa. In distributional terms, it is written as

$$\begin{aligned} \mu &\sim \mathcal{N}(\lambda, V) \\ \Sigma \,|\, \mu &\sim \mathcal{W}^{-1}(\Psi, \nu), \end{aligned}$$

meaning that μ is *a priori* Normally distributed independently of Σ, which in turn follows an n dimensional inverse Wishart distribution:

$$f(\Sigma \,|\, \Psi, \nu) = \frac{|\Psi|^{\nu/2}}{2^{\nu n/2}\Gamma_n(\frac{\nu}{2})} |\Sigma|^{-(\nu+n+1)/2} \exp\left\{-\tfrac{1}{2}\mathrm{tr}(\Psi\Sigma^{-1})\right\},$$

where

$$\Gamma_n(a) = \pi^{n(n-1)/4} \prod_{i=1}^{n} \Gamma\left(a + \tfrac{1-i}{2}\right)$$

is the multivariate gamma function. We shall refer to this prior as the Normal Independent Inverse Wishart (NIIW) prior and denote it $\pi_1(\mu, \Sigma)$. This prior is parametrized by the mean and variance of μ as well as the mean of Σ, which is $\Psi/(\nu - n - 1)$. Note that the variance and covariance of the inverse Wishart distribution are

$$\begin{aligned} \operatorname{var}(\Sigma_{ij}) &= \frac{(\beta+1)\Psi_{ij}^2 + (\beta-1)\Psi_{ii}\Psi_{jj}}{\beta(\beta-1)^2(\beta-3)} \\ \operatorname{cov}(\Sigma_{ij}, \Sigma_{kl}) &= \frac{2\Psi_{ij}\Psi_{kl} + (\beta-1)(\Psi_{ik}\Psi_{jl} + \Psi_{il}\Psi_{kj})}{\beta(\beta-1)^2(\beta-3)}, \end{aligned}$$

where $\beta = \nu - n$.

Sampling from the posterior distribution $p_1(\mu, \Sigma \mid X)$ for the NIIW choice of prior can be accomplished with a Gibbs sampler:

1. Pick any starting value (μ_0, Σ_0). Now suppose that the algorithm has been run for $m-1$ steps yielding the values $(\mu_0, \Sigma_0), (\mu_1, \Sigma_1), \ldots, (\mu_{m-1}, \Sigma_{m-1})$.

2. Sample μ_m from the conditional distribution $p(\mu \mid \Sigma_{m-1}, X)$, which is multivariate Normal:

$$\mu \mid \Sigma, X \sim \mathcal{N}\left(B\lambda + (1-B)\bar{X}, (1-B)\tfrac{1}{N}\Sigma\right),$$

   where $\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ and $B = \frac{1}{N}\Sigma(V + \frac{1}{N}\Sigma)^{-1}$.

3. Sample Σ_m from the conditional distribution $p(\Sigma \mid \mu_m, X)$, which is inverse Wishart:

$$\Sigma \mid \mu, X \sim \mathcal{W}\left(N(\mu - \bar{X})(\mu - \bar{X})' + S + \Psi, N + \nu\right),$$

where $S = \sum_{i=1}^{n}(X_i - \bar{X})(X_i - \bar{X})'$. It is standard practice to simulate draws from an inverse Wishart distribution by way of Bartlett's decomposition (e.g., Mardia et al., 1979).

4. We have now produced a sample (μ_m, Σ_m) which is dependent on the previous one. Repeat steps 2 and 3 for M iterations and discard e.g., the first 10%. Although they are correlated, the samples obtained by this Markov chain Monte Carlo (MCMC) algorithm have all the properties of an iid sample from $p_1(\mu, \Sigma | X)$ as long as M is large enough.

C.2. Dependent μ and Σ. Another useful prior specification for μ and Σ is

$$\Sigma \sim \mathcal{W}^{-1}(\Psi, \nu)$$
$$\mu | \Sigma \sim \mathcal{N}(\lambda, \kappa^{-1}\Sigma).$$

This distribution is commonly referred to as a Normal Inverse Wishart (NIW) and we label it $\pi_2(\mu, \Sigma)$. In this case, the prior mean and variance of μ are

$$E[\mu] = \lambda, \qquad \mathrm{var}(\mu) = \frac{\Psi}{\kappa(\beta - 1)}.$$

However, even though μ and Σ are *a priori* dependent they are uncorrelated: $\mathrm{cov}(\mu_i, \Sigma_{jk}) = 0$.

Under the NIW prior specification, the posterior distribution $p_2(\mu, \Sigma | X)$ is also NIW:

$$\Sigma | X \sim \mathcal{W}^{-1}\left(\Psi + S + \frac{N\kappa}{N+\kappa}(\bar{X} - \lambda)(\bar{X} - \lambda)', N + \nu\right)$$
$$\mu | \Sigma, X \sim \mathcal{N}\left(\frac{N\bar{X} + \kappa\lambda}{N+\kappa}, (N+\kappa)^{-1}\Sigma\right).$$

The primary advantage of the NIW prior over the one with independent μ and Σ is that it admits a

Monte Carlo algorithm producing iid samples from the posterior distribution $p_2(\mu,\Sigma\,|\,X)$. That is, we can draw M iid samples $\Sigma_1,\ldots,\Sigma_M$ from the inverse Wishart distribution, and for each value of Σ_m, we then draw μ_m from the appropriate Normal.

C.3. Noninformative Priors. An interesting connection between the NIW and NIIW priors occurs by letting $\kappa \to 0$ in the former and $V \to \infty$ in the latter. The limiting distribution is a so-called Lebesgue or "flat" prior on μ: $\pi(\mu) \propto 1$. This is not a proper probability distribution in the sense that $\int \pi(\mu)\,\mathrm{d}\mu = \infty$. However, the posterior distributions $p_i(\mu,\Sigma\,|\,X)$, $i = 1,2$ are both proper and in fact one and the same NIW:

$$
\begin{aligned}
\Sigma\,|\,X &\sim \mathcal{W}^{-1}\left(\Psi + S, N + \nu - 1\right) \\
\mu\,|\,\Sigma, X &\sim \mathcal{N}\left(\bar{X}, \Sigma/N\right).
\end{aligned}
\tag{C.1}
$$

It is possible to take this one step further by specifying a noninformative prior for Σ, $\pi(\Sigma\,|\,\mu) \propto |\Sigma|^{-(\nu+n+1)/2}$. This is proportional to an inverse Wishart distribution with $\Psi = 0$. This is the only scale-invariant prior on Σ in the sense that if $\Omega = A\Sigma A'$, then $\pi(\Omega) \propto |\Omega|^{-(\nu+n+1)/2}$. Using the moment formulas of the NIW distribution, it is easy to check that settting $\nu = n + 1$ in (C.1) with $\Psi = 0$ ensures that the posterior means of μ and Σ coincide with their usual frequentist estimates: $E[\mu\,|\,X] = \bar{X}$ and $E[\Sigma\,|\,X] = S/(N-1)$.

C.4. Non-Conjugate Priors. The use of the Wishart distribution in the prior specifications above greatly simplifies posterior calculations. However, the Wishart distribution imposes certain undesirable constraints on the elements of Σ (e.g., Gelman and Hill, 2006). Several alternatives allowing greater modeling flexibility at the expense of computational tractability are available in the statistical literature, for instance: Leonard and Hsu (1992); Barnard et al. (2000); Liechty et al. (2004); OMalley and Zaslavsky (2008); Huang and Wand (2013).

While each of these priors has its own merit, differences in prior specification diminish in significance as the sample size becomes large. To investigate the impact of the prior on inferential results, we used the point estimates $\bar{X}$ and $S/(N-1)$ from the fish data (see Supplementary Material for raw data) as the true values of μ and Σ to simulate $N = 70$ iid multivariate Normals (the actual sample size for each species). Posterior parameter estimates from the default prior are plotted with their true values in Figure C1 for the Arctic Cisco data. For this particular simulation, the posterior means are quite close to the true parameter values.

For a more in-depth evaluation of the default prior, 1000 datasets $X_1, \ldots, X_{1000}$ of size $N = 70$ were simulated from the Arctic Cisco parameter values. For each parameter $\theta \in \{\mu_i, \Sigma_{ij}\}$, two metrics were computed. The first is the (relative) root mean-square error (RMSE):

$$\mathrm{RMSE}(\theta) = \sqrt{\frac{1}{1000} \sum_{i=1}^{1000} \frac{(\hat{\theta}_i - \theta)^2}{\theta^2}},$$

where $\hat{\theta}_i$ is the posterior mean of θ for dataset i. Results are shown in Table C1. The second metric is the true coverage of the posterior 95% credible intervals:

$$\mathrm{COV}(\theta) = \frac{1}{1000} \sum_{i=1}^{1000} \delta_i, \qquad \delta_i = \begin{cases} 1 & \theta \in \{L_i, U_i\}, \\ 0 & \theta \notin \{L_i, U_i\}, \end{cases}$$

where L_i and U_i are the 2.5% and 97.5% quantiles of $p(\theta \mid X_i)$.

The coverage probabilities of the 95% credible intervals out of the 200 replications for each parameter are given in Table C2. Nearly identical results for the other three species of fish indicate that the default noninformative prior described in Section C.3 is adequate for a sample size of $N = 70$ with the given sets of true parameter values. Practitioners are encouraged to carefully consider their choice of prior when sample sizes are low and/or number of dimensions is high.

TABLE C1: Relative RSME ($\times$100%) over 1000 simulated datasets from the Arctic Cisco parameters values.

μ		
δ^{15}N	δ^{13}C	δ^{34}S
0.768	0.539	1.02

Σ			
	δ^{15}N	δ^{13}C	δ^{34}S
δ^{15}N	2.07	3.79	4.91
δ^{13}C	3.79	3.34	8.74
δ^{34}S	4.91	8.74	4.57

TABLE C2: True coverage of 95% credible intervals for 1000 simulated datasets from the Arctic Cisco parameters values.

μ		
δ^{15}N	δ^{13}C	δ^{34}S
94.7	95.9	94.6

Σ			
	δ^{15}N	δ^{13}C	δ^{34}S
δ^{15}N	94.4	95.1	93.1
δ^{13}C	95.1	94.6	93.5
δ^{34}S	93.1	93.5	94.4

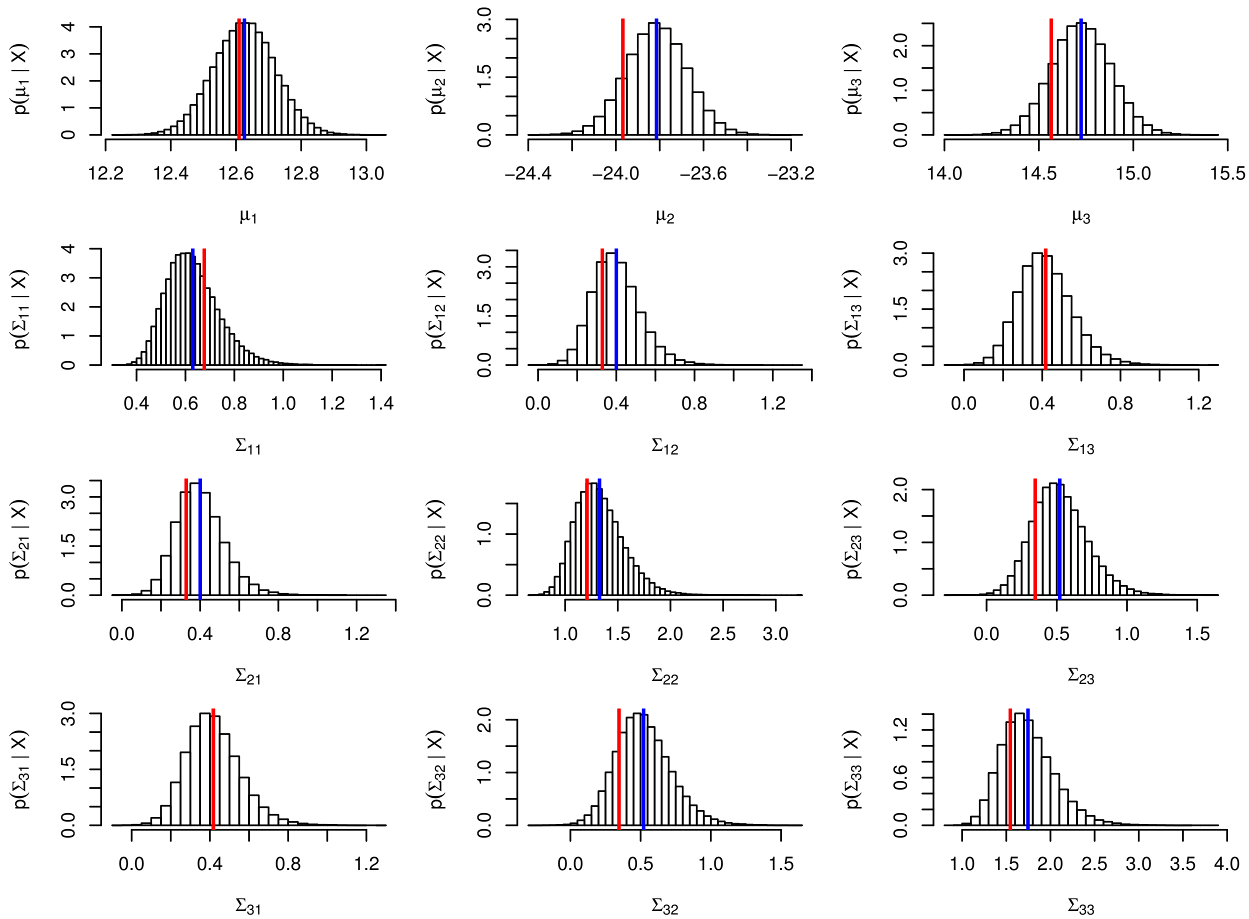

FIG. C1: Posterior parameter distribution with the default noninformative prior for a simulated data from the Arctic Cisco parameter values. The blue lines correspond to posterior means and the red lines are the true parameter values.

REFERENCES.

Barnard, J., McCulloch, R., and Meng, X.L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4): 1281–1312.

Gelman, A. and Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge University Press, New York.

Huang, A. and Wand, M.P. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, 8(2): 439–452.

Leonard, T. and Hsu, J.S.J. (1992). Bayesian inference for a covariance matrix. *The Annals of Statistics*, 20(4): 1669–1696.

Liechty, J.C., Liechty, M.W., and Müller, P. (2004). Bayesian correlation estimation. *Biometrika*, 91(1): 1–14.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). *Multivariate Analysis*. Probability and mathematical statistics. Academic Press, London.

OMalley, A.J. and Zaslavsky, A.M. (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association*, 103(484): 1405–1418.