## Bayesian Overview

* Bayes’ Theorem:
* **Evidence –** P(B) likelihood, averaged over the prior.
* **Prior –** P(A) probability distribution of the parameter before the evidence is taken into account.
  + Is updated based on the data.
* **Likelihood –** P(B|A) relative number of ways that the data can occur within the bounds of the prior.
* **Posterior distribution –** P(A|B) probability distribution of the parameter, given the prior and the data.
* **Parameter –** unknown. Estimated from the data.

## Priors

* **Informative –** specific, definite information about a variable.
* **Regularizing –** keeps inferences within a reasonable range, but he data does most of the informing.
  + Use this kind for myctophids.
  + We have some idea about what the parameters might be from previous studies, so can keep within a range, but we want the data do most of the informing.
* **Diffuse prior –** equal probabilities to all outcomes.

## Multilevel Models

1. Adjust for repeated sampling.
   * Many individuals per species.
2. Adjust for imbalance in sampling.
   * Different number of individuals per species.
3. Avoids averaging.
   * Individuals per species rather than species averages.

## MCMC

* **Monte Carlo simulations** - approximates a parameter by repeatedly generating random numbers.
* **Markov Chains** – sequence of event, where the next event is predicted by the current event.
* Basic method:

1. Pick a random potential parameter value.
2. Test how likely the parameter value is to explain the data, given the prior.
3. The parameter then informs the next parameter to be picked.

* Samples directly from posterior without assuming a shape.
  + I.E. doesn’t assume normality.
* **Gibbs Sampling –** adjusts the distribution of a proposed parameter value, depending on the current parameter values.
  + Used in BUGS and JAGS (and therefore MixSIAR).
  + Uses **conjugate priors** (prior and posterior distributions are from the same family, e.g. both Gaussian).
* **Hamiltonian Monte Carlo –** sends a frictionless particle to glide over the posterior, turning when the gradient of the posterior specifies.
  + Used in Stan.
  + Doesn’t require conjugate priors.
  + More efficient than Gibbs sampling.

## MixSIAR

* Uses JAGS (Just Another Gibbs Sampler).
* Mixing models that estimate proportion of sources to an isotopic mixture.
* Advantages of Bayesian:
  + Takes uncertainty into account for source values and discrimination.
  + Incorporates prior information.

### Priors

* Use the uninformative priors.

### MCMC Parameters

* **“Normal”**
  + Chain length = 100,000
  + Burn in = 50,000
    - Proportion of the chain that may not converge on the target distribution yet (discarded).
  + Thinning parameter = 50
  + Number of chains = 3

### Error Structure

* Error structure is “**process only**” because we only have one mixture data point.
  + Error comes from consumers sampling in different areas of the diet and DIC distribution.

### Diagnostics

* **Gelman-Rubin diagnostic** is the same as R-hat.
  + Needs to be a close to 1 to ensure convergence.
* **Geweke diagnostic** is a z-score.
  + MixSIAR tells you how many variables were outside +/- 1.96 in each chain.
* From PDF:
  + Deviance – Deviance information criterion.
    - Can ignore for our purposes.
  + ILR – Isometric log ratio transformation.
    - Used in the analysis of compositional data.
  + Loglik – log likelihood.
    - P(B|A)
  + p.fac and p.global – probability densities of each source.
    - Are the same for M because we are only looking at one individual at a time.
    - [1,1] is DIC and [1,2] is diet.
    - P(A|B)

### Outputs

* Posterior plot of contributions of each source.

## Body Mass Temperature Model

* M ~ a + a\_Var[Species] + b\_W\*Weight + b\_T\*Temp\_est
* Priors:
  + A, b\_W & b\_T ~ dnorm(0, 1)
  + Sigma & sigma\_Species ~ dexp(1)
* Adapt delta 🡪 size of steps that stan takes when sampling.
  + 0.999 to avoid divergent iterations.

### Test Run

* 3 chains, 1000 iter, 500 warmup.
* Errors:
  + 4 transitions after warmup that exceeded maximum treedepth.
  + 2 chains where the estimated Bayesian Fraction of Missing information was low.
  + ---
  + 809 transitions after warmup that exceeded maximum treedepth.
  + 2 chains where estimated Bayesian Fraction of missing information was low.
* Treedepth.
  + “When the maximum allowed tree depth is reached it indicates that NUTS is terminating prematurely to avoid excessively long execution time.”
  + NUTS 🡪 No U Turn Sampler (a variant of HMC used by Stan).
  + Try increasing max\_treedepth to 15.
    - Test 1 – worked!
    - Test 2 – worked!
* Bayesian Fraction of Missing information.
  + “This implies that the adaptation phase of the Markov Chains did not turn out well and those chains likely did not explore the posterior distribution efficiently.”
  + Try with longer chain (2000, 1000 iter).
    - Test 1 – 1 chain where BFMI was low.
    - Test 2 – 3 chains where BFMI was low.
  + E-BFMI should be above 0.2.
    - Chain 2 = 0.136.
    - Chain 3 = 0.056.
  + Try changing sigma to Cauchy prior.
    - Seems to have helped?
      * BFMI good.
      * Treedepth and iterations good.
    - Still some “banana” shape between b\_T and sigma.
* Tried long model.
  + 4000 + divergent iterations.
  + BFMI = 0.82 🡪 low.
* Try with longer warmup (5e5) and adapt delta = 0.9999.
  + Still getting errors, even worse than above.
* Try with JAGS number of iterations.
  + 100,000 iter 50,000 warmup.
  + N\_eff good.
  + BFMI low.
  + 60 divergent iterations.
  + 19 iterations exceeded max treedepth.
* Reparameterise model.
  + <https://mc-stan.org/docs/2_18/stan-users-guide/reparameterization-section.html>
  + Test with 3 chains. Iter = 1000, warmup = 500.

|  |  |  |
| --- | --- | --- |
| **Feature** | **Regular** | **Reparameterized – with sigma** |
| Error messages | None | “Stan model 'M\_est ~ dnorm(mu, sigma)' does not contain samples.” |
| N\_eff | Small for b\_T (71) and sigma(31). | Small for b\_T (52) and sigma (40). |
| R\_hat | High for sigma (1.08). | All below 1.01. |
| BFMI | 2 chains where BFMI was between 0.2 and 0.15. | Ok |
| Treedepth | OK | Ok |
| Divergent iterations | 0 | 0 |
| Pairs plot | All parameters fairly unimodal.  Correlation between b\_T and sigma. | All parameters fairly unimodal.  Correlation between b\_T and sigma. |
| Parameters | * A = 0.1835 * B\_W = 0.0016 * B\_T = -0.0034 * Sigma\_species = 0.338 * Sigma = 0.109 | * A = 0.1819 * B\_W = 0.0017 * B\_T = -0.0032 * Sigma\_species = 0.0341 * Sigma = 0.0114 |

* Try with default iter and warmup – regular.
  + Should have twice the number of effective samples.
    - Still low n\_eff for sigma (22) and b\_T (17).
    - Correspondingly high Rhat (1.16 and 1.13).
    - Correlation of 0.78.
* Email Andrew Jackson.
  + Give outline of the problem first rather than full on.

### Andrew Suggestions

* Try Z-scoring body mass and temperature.
  + It worked!

## Results – Body Mass and Temperature Model

* Run on default settings.
* Diagnostics:
  + Pairs plot, looking happy.
  + E-BFMI good.
  + Treedepth all ok.
  + No divergent iterations.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Parameter** | **Mean** | **SD** | **Lower 95%** | **Upper 95%** | **N\_eff** | **R\_hat** |
| a | 0.1835 | 0.0175 | 0.1501 | 0.2185 | 170 | 0.9993 |
| b\_W | 0.0020 | 0.0045 | -0.0070 | 0.0106 | 1500 | 1.0000 |
| b\_T | -0.0038 | 0.0025 | -0.0088 | 0.0011 | 289 | 0.9994 |
| a\_Var\_ELN | 0.0280 | 0.0177 | -0.0055 | 0.0643 | 175 | 0.9995 |
| a\_Var\_ELC | -0.0110 | 0.0182 | -0.0462 | 0.0228 | 174 | 0.9993 |
| a\_Var\_GYR | 0.0164 | 0.0181 | -0.0165 | 0.0544 | 174 | 0.9994 |
| a\_Var\_GYN | -0.0391 | 0.0201 | -0.0772 | 0.0004 | 205 | 0.9996 |
| a\_Var\_KRA | 0.0103 | 0.0183 | -0.0225 | 0.0486 | 198 | 0.9993 |
| a\_Var\_PRM | -0.0130 | 0.0183 | -0.0440 | 0.0269 | 193 | 0.9993 |
| sigma\_Species | 0.0354 | 0.0186 | 0.0118 | 0.0715 | 330 | 1.0027 |
| sigma | 0.0125 | 0.0016 | 0.0090 | 0.0154 | 402 | 0.9994 |