phyloflows: Performing MCMC diagnostic checks

Xiaoyue Xi and Oliver Ratmann

2019-09-10

This vignette describes how to run a number of diagnostics on **phyloflows** MCMC output, obtained with the function `phyloflows:::source.attribution.mcmc`. Please work through the vignette *phyloflows: Estimating transmission flows under heterogeneous sampling – a first example* before you go ahead here.

Getting started

We continue our "First_Example". The following code chunk contains all code needed, up to running **phyloflows** MCMC routine. The only change is that the number of iterations is now $50,000$. The MCMC should take about 2 minutes to run.

```
require(data.table)
require(phyloflows)

data(twoGroupFlows1, package="phyloflows")
dobs <- twoGroupFlows1$dobs
dprior <- twoGroupFlows1$dprior
control <- list(seed=42, mcmc.n=5e4, verbose=0)
mc <- phyloflows:::source.attribution.mcmc(dobs, dprior, control)
```

MCMC: diagnostics

phyloflow comes with a function to calculate standard MCMC diagnostics. You can

1. Make trace plots for all model parameters;
2. or make trace plots for the model parameters with smallest effective sample size. This may be useful to avoid generating very large pdf files that you won t be able to open anyway.
3. Make trace plots for values of the log likelihood and log posterior density.
4. Calculate acceptance rates.
5. Remove a burn-in period.
6. Calculate effective sample sizes.
7. Calculate summary statistics (mean, median, quantiles) of the marginal posterior densities.
8. Plot marginal posterior densities for the model parameters with smallest effective sample size.
9. Make autocorrelation plots for the model parameters with smallest effective sample size.

The syntax is as follows. Look up the help page for the diagnostics function for a full explanation of the control arguments.

```
outfile.base <- file.path(getwd(),'twoGroupFlows1_mcmc')
control <- list( burnin.p=0.05,
                 regex_pars='*',
                 credibility.interval=0.95,
                 pdf.plot.all.parameters=TRUE,
                 pdf.plot.n.worst.case.parameters=1,
                 pdf.height.per.par=1.2,
```

```
                outfile.base=outfile.base)
phyloflows:::source.attribution.mcmc.diagnostics(mc=mc, control=control)
#>
#> Using MCMC output specified as input...
#> Collecting parameters...
#> Plotting traces for all parameters...
#>
#> Plotting traces for log likelihood and log posterior...
#>
#> Plotting histograms for log likelihood and log posterior...
#>
#> Calculating acceptance rates...
#> Average acceptance rate=  0.897
#> Update IDs with lowest acceptance rates   UPDATE_ID ACC_RATE
#> 1:         2  0.77528
#> 2:         1  0.81440
#>
#> Removing burnin in set to  5 % of chain, corresponding to the first iterations= 625
#> Calculating effective sample size for all parameters...
#>
#> Calculating posterior summaries for all parameters...
#> Summary of parameters with lowest effective samples
#>             VAR      MEAN         SD    MEDIAN      CI_L      CI_U           ID      NEFF
#> 1:         XI-2 0.4499586 0.01006440 0.4498079 0.4289564 0.4701081         XI-2  4769.454
#> 2:         XI-1 0.5998396 0.01059647 0.5999290 0.5796989 0.6196669         XI-1  5468.961
#> 3: LOG_LAMBDA-4 6.4520957 0.09947787 6.4531609 6.2548685 6.6448369 LOG_LAMBDA-4  6824.919
#> 4: LOG_LAMBDA-1 5.9531912 0.09225504 5.9535402 5.7711190 6.1320875 LOG_LAMBDA-1  8859.387
#> 5: LOG_LAMBDA-3 4.2844403 0.22661016 4.2919691 3.8214479 4.7072751 LOG_LAMBDA-3 11315.509
#> 6: LOG_LAMBDA-2 3.9930827 0.26076190 4.0023851 3.4517322 4.4765022 LOG_LAMBDA-2 11876.000
#>
#> Writing summary file to /Users/xx4515/phyloscanner/phyloflows/vignettes/twoGroupFlows1_mcmc_summary.
#> Plotting traces for worst parameters...
#>
#> Plotting marginal posterior densities for worst parameters...
#>
#> Plotting autocorrelations for worst parameters...
#> pdf
#>   2
```

That's it for now. Use your usual R wizadry to process the output further, and have a look at the other vignettes.