Measuring mixing for MCMC in phylogenetics
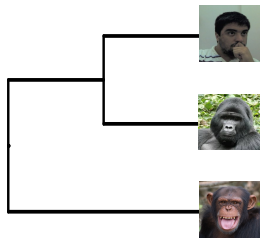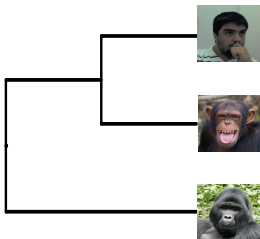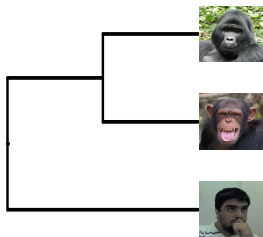
The (strange) case of clade indicators

Luiz Max Carvalho & Remco Bouckaert (Auckland)

lmax.fgv@gmail.com

FGV EMAp

Please do watch Andy Magee's excellent presentation first.

$$p(t, \boldsymbol{b}, \boldsymbol{\omega}|D) = \frac{f(D|t, \boldsymbol{b}, \boldsymbol{\omega})\pi(t, \boldsymbol{b}, \boldsymbol{\omega})}{\sum_{t_i \in \boldsymbol{T}_n} \int_{\boldsymbol{B}} \int_{\boldsymbol{\Omega}} f(D|t_i, \boldsymbol{b}_i, \boldsymbol{\omega})\pi(t_i, \boldsymbol{b}_i, \boldsymbol{\omega})d\boldsymbol{\omega}d\boldsymbol{b}_i}. \quad (1)$$

◎ D: observed sequence (DNA) data;

◎ $\boldsymbol{T}_n$: set of all binary ranked trees ($\mathbb{G}^{(2n-3)!!}$);

◎ $\boldsymbol{b}_k$: set of branch lengths of $t_k \in \boldsymbol{T}_n$ ($\mathbb{R}_+^{2n-2}$, kind of) ;

◎ $\boldsymbol{\omega}$: set of parameters of interest such as substitution model
   parameters, migration rates, heritability coefficients, etc.

A clade is a partition of the set of leaves and two clades
$A = A_1|A_2$ and $B = B_1|B_2$ are said to be compatible if at least
one of $A_i \cap B_j$, $i, j = 1, 2$ is empty. Here's a picture[1]:

_____
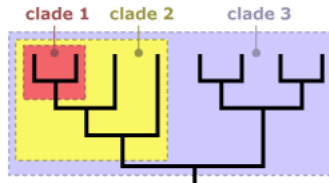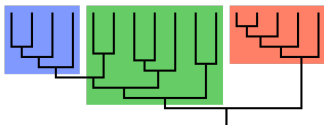
[1]Pictures taken from Wikipedia and from `https://evolution.berkeley.edu/evolibrary/news/080301_elephantshrew`

Why clades?

◎ **Dimension!** $|\mathbb{T}_n| = (2n - 3)!! \; vs \; |\mathbb{C}_n| = 2^{n-1} - 1$

◎ Interpretability;

◎ Under simplifying assumptions, clades are independent (Larget, 2013[2]);

◎ Clade distribution is known under popular prior distributions.

---

[2]but see Whidden & Matsen, 2015 and Zang & Matsen, 2018.

Setup

Let $X_j^{(i)} \in \{0, 1\}$ be the indicator of whether clade j in the tree sampled at the i-th iteration and $\hat{p}_j = M^{-1} \sum_{i=1}^{M} X_j^{(i)}$ be a simple MCMC estimator of its marginal success probability.

An attempt

Pretend for a second $\left(X_j^{(i)}\right)_{i \geq 0}$ is Markov on $\mathcal{X} = \{0, 1\}$ and reparametrise the usual two-state model as

$$\tilde{P}_x := \begin{bmatrix} 1 - \alpha & \alpha \\ \alpha \frac{1-p}{p} & \frac{p - \alpha(1-p)}{p} \end{bmatrix}, \tag{2}$$

where p is the marginal success probability and a α controls the "flipping rate" of the chain. It is straightforward to estimate $\hat{p}$ and then derive the maximum *a posteriori* estimate $\hat{\alpha}$ conditional on $\hat{p}$ in closed-form.

What an explicit model buys you

Under this model we can derive

- ◎ Distribution of occupation times;
- ◎ Distribution of state-transitions ($0 \rightarrow 1$ or $1 \rightarrow 0$);
- ◎ Effective sample size:

$$
\begin{aligned}
\text{ESS} &= \frac{M}{1 + 2\sum_{t=1}^{\infty} \rho_t}, \\
&= \frac{M}{1 + 2\frac{p-\alpha}{\alpha}}, \\
&= \frac{\alpha}{2p - \alpha} M. \tag{3}
\end{aligned}
$$

We can also compute a few measures of performance.

◎ Scaled average state-transtions:

$$\phi := \frac{\bar{y}}{2\alpha_{\max}(\hat{p})\hat{p}(1-\hat{p})}, \tag{4}$$

◎ Scaled switching rate:

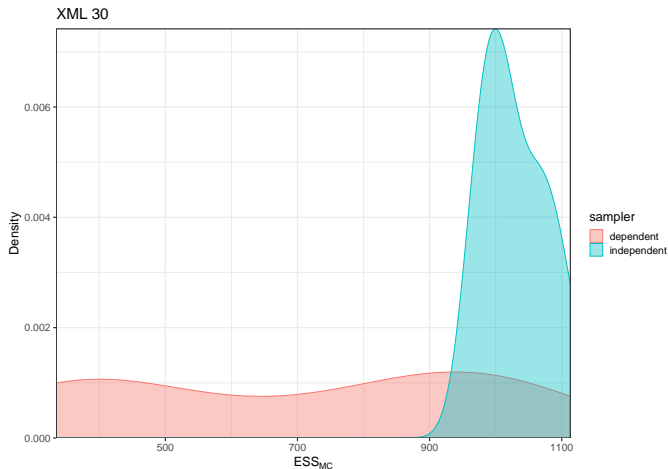$$\psi := \frac{\hat{\alpha}}{\alpha_{\max}(\hat{p})}, \tag{5}$$

◎ Switching ratio:

$$\delta = \frac{\Delta(\boldsymbol{X})}{\Delta_{\max}(\boldsymbol{X})}. \tag{6}$$
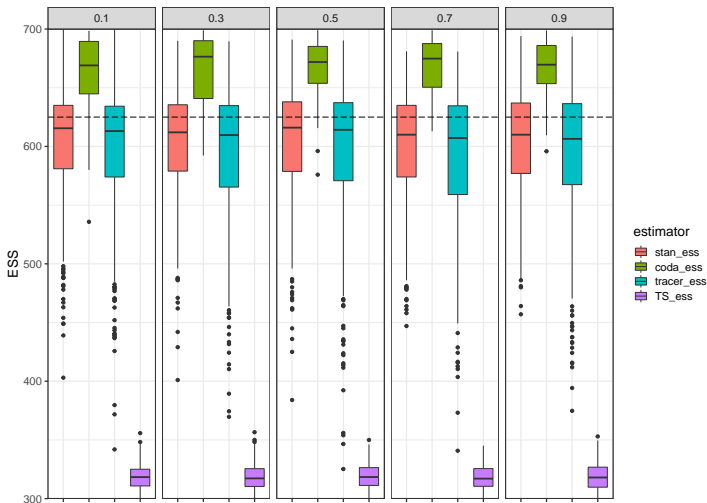
All closed-form and easy to compute.

Looking cool!

We can fake phylogenetic MCMC quite well. In particular we can sample from the posterior "exactly".
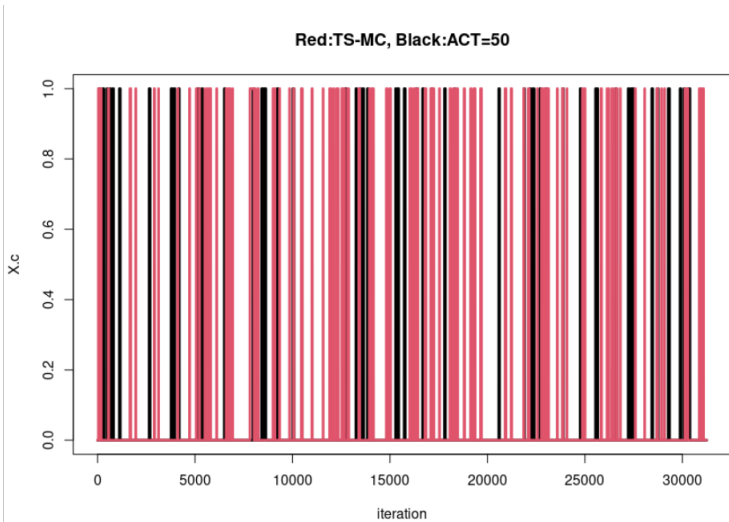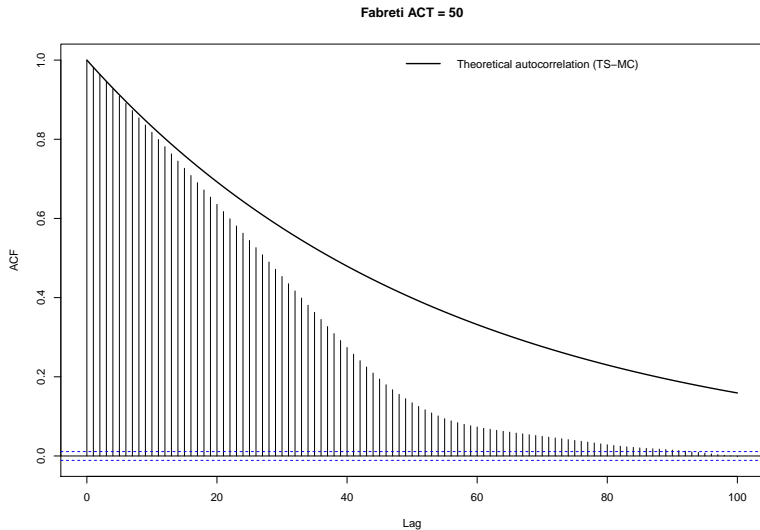
What could possibly go wrong?

Fabreti & Hönna, 2021 show how to create (non-Markov) binary processes with a pre-specified ESS.

A tale of two processes

Red:TS-MC, Black:ACT=50

Fabreti ACT = 50

A possible (expensive) solution

We could devise a hybrid estimator:

1. Estimate the autocorrelation time, $\hat{\tau}$;

2. **If** $\hat{\tau} > d$, return $\widehat{\text{ESS}} = M/\hat{\tau}$;
   Else
   2.1 Compute $(\hat{p}, \hat{\alpha})$;
   2.2 Return $\widehat{\text{ESS}} = \frac{\hat{\alpha}}{2\hat{p} - \hat{\alpha}} M$.

This should work OK for $d \in [1, 2]$.

- ◎ Advantage: Accommodates all sorts of binary processes (including super-efficient ones);

- ◎ Drawback: More expensive to compute.

Raftery, 1986. Under the null,

$$G^2 = 2 \sum_{i,j,k} n_{ijk} \log \left(\frac{n_{ijk}/n_{ij+}}{n_{+jk}/n_{+j+}} \right), \tag{7}$$

is chi-square-distributed with $\nu = m(m-1)^2 = 2$ degrees of freedom.

Bayes factors:

$$BF_{12} \approx \frac{1}{2}(G^2 - \nu \log(M)), \tag{8}$$

$$= \sum_{i,j,k} n_{ijk} \log \left(\frac{n_{ijk}/n_{ij+}}{n_{+jk}/n_{+j+}} \right) - \log(M). \tag{9}$$

Magee et al, 2021 point out that trees are fundamentally multivariate objects.

$$\text{mESS} = M \left(\frac{\det(\mathbf{\Lambda})}{\det(\mathbf{\Sigma})} \right)^{1/p}. \tag{10}$$

```
> ( evals.naive <- eigen(cov.dep, only.values = TRUE)$values )
 [1]  2.460008e-01  2.357391e-01  2.161817e-01  1.374673e-01  8.833706e-02  7.734214e-02
 [7]  5.809434e-02  3.283007e-02  1.535663e-02  8.976874e-03  3.982149e-03  2.242468e-03
[13]  1.437667e-03  6.836824e-04  4.688762e-04  3.356731e-04  1.117728e-17  4.321235e-18
[19]  1.419069e-18  5.143897e-20 -1.708911e-19 -1.086942e-18 -8.299469e-18 -3.081920e-17
> ( evals.robust <- eigen(robust.cov.dep, only.values = TRUE)$values )
 [1]  2.459980e-01  2.357382e-01  2.161232e-01  1.374668e-01  8.833950e-02  7.738005e-02
 [7]  5.809705e-02  3.281389e-02  1.535756e-02  8.976479e-03  3.981357e-03  2.244039e-03
[13]  1.442280e-03  6.864393e-04  4.714446e-04  3.383832e-04  4.970055e-06  4.970055e-06
[19]  4.970055e-06  2.988021e-06  9.980030e-07  9.980030e-07  9.980030e-07  9.980030e-07
```

Figure: Eigenvalues can be numerically unstable.

Take home

Assumptions are useful!

Assuming Markov we can do a bunch of closed-form calculations that might be useful in practice.

Assumptions matter, though

Do not assume Markov when you can do without.

Accommodating antithetic chains

Assuming you encounter it in the wild, this framework will help detect super efficiency;

Multivariate binary ESS

can provide a nice and cheap diagnostic. To be seen.

THE
END