Conjugate Gibbs Sampling for Bayesian Phylogenetic Models

cEvo journal club – Basel

Marc Manceau

August 31, 2020

Some reasons to speak about this paper

1. It's a follow up on a previous discussion on "exponential families",
2. It uses this concept within a Gibbs Sampler, in a phylogenetic setting,
3. It seems quite efficient, and it might give us ideas to speed up future projects,
4. I usually like Nicolas Lartillot's papers.

Some reasons to speak about this paper

1. It's a follow up on a previous discussion on "exponential families",
2. It uses this concept within a Gibbs Sampler, in a phylogenetic setting,
3. It seems quite efficient, and it might give us ideas to speed up future projects,
4. I usually like Nicolas Lartillot's papers.

Some reasons to speak about this paper

1. It's a follow up on a previous discussion on "exponential families",
2. It uses this concept within a Gibbs Sampler, in a phylogenetic setting,
3. It seems quite efficient, and it might give us ideas to speed up future projects,
4. I usually like Nicolas Lartillot's papers.

Some reasons to speak about this paper

1. It's a follow up on a previous discussion on "exponential families",
2. It uses this concept within a Gibbs Sampler, in a phylogenetic setting,
3. It seems quite efficient, and it might give us ideas to speed up future projects,
4. I usually like Nicolas Lartillot's papers.

Some reasons to speak about this paper

1. It's a follow up on a previous discussion on “exponential families”,
2. It uses this concept within a Gibbs Sampler, in a phylogenetic setting,
3. It seems quite efficient, and it might give us ideas to speed up future projects,
4. I usually like Nicolas Lartillot's papers.

Introduction

Aim of the paper

- They want to model an amino-acid substitution process,
- evolving along a fixed tree topology,
- while allowing for a lot of heterogeneities at the site level.

Standard Metropolis-Hastings MCMC methods take very long to do so,

Could it be improved by considering a clever Gibbs sampler ?

Introduction

Aim of the paper

- They want to model an amino-acid substitution process,
- evolving along a fixed tree topology,
- while allowing for a lot of heterogeneities at the site level.

Standard Metropolis-Hastings MCMC methods take very long to do so,

Could it be improved by considering a clever Gibbs sampler ?

Introduction

Aim of the paper

- They want to model an amino-acid substitution process,
- evolving along a fixed tree topology,
- while allowing for a lot of heterogeneities at the site level.

Standard Metropolis-Hastings MCMC methods take very long to do so,

Could it be improved by considering a clever Gibbs sampler ?

Introduction

Aim of the paper

- They want to model an amino-acid substitution process,
- evolving along a fixed tree topology,
- while allowing for a lot of heterogeneities at the site level.

Standard Metropolis-Hastings MCMC methods take very long to do so,

Could it be improved by considering a clever Gibbs sampler ?

Introduction

Aim of the paper

- They want to model an amino-acid substitution process,
- evolving along a fixed tree topology,
- while allowing for a lot of heterogeneities at the site level.

Standard Metropolis-Hastings MCMC methods take very long to do so,

Could it be improved by considering a clever Gibbs sampler ?

Introduction

Metropolis-Hastings or Gibbs sampling ?

Metropolis-Hastings

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:
$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$
3. Draw $u \sim \mathcal{U}(0,1)$.
If $u \leq r$, set $x_{i+1} := y_{i+1}$.
otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x,y)) = \nu_y q_{yx} \min(1, r(y,x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

Gibbs sampler

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \dots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law
$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j \neq i} = \left(x_n^{(j)}\right)_{j \neq i}\right).$$
3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \dots, x_{n+1}^{(i)}, \dots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Metropolis-Hastings or Gibbs sampling ?

METROPOLIS-HASTINGS

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:
$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$
3. Draw $u \sim \mathcal{U}(0,1)$.
If $u \leq r$, set $x_{i+1} := y_{i+1}$.
otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x,y)) = \nu_y q_{yx} \min(1, r(y,x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

GIBBS SAMPLER

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \dots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law
$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j \neq i} = \left(x_n^{(j)}\right)_{j \neq i}\right).$$
3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \dots, x_{n+1}^{(i)}, \dots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Metropolis-Hastings or Gibbs sampling ?

METROPOLIS-HASTINGS

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:
$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$
3. Draw $u \sim \mathcal{U}(0,1)$.
If $u \leq r$, set $x_{i+1} := y_{i+1}$.
otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x,y)) = \nu_y q_{yx} \min(1, r(y,x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

GIBBS SAMPLER

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \ldots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \ldots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law
$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j \neq i} = \left(x_n^{(j)}\right)_{j \neq i}\right).$$
3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \ldots, x_{n+1}^{(i)}, \ldots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Metropolis-Hastings or Gibbs sampling ?

METROPOLIS-HASTINGS

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:

$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$

3. Draw $u \sim \mathcal{U}(0,1)$.
   If $u \leq r$, set $x_{i+1} := y_{i+1}$.
   otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x,y)) = \nu_y q_{yx} \min(1, r(y,x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

GIBBS SAMPLER

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \dots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law

$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j\neq i} = \left(x_n^{(j)}\right)_{j\neq i}\right).$$

3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \dots, x_{n+1}^{(i)}, \dots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Metropolis-Hastings or Gibbs sampling ?

METROPOLIS-HASTINGS

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:
$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$
3. Draw $u \sim \mathcal{U}(0,1)$.
If $u \leq r$, set $x_{i+1} := y_{i+1}$.
otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x,y)) = \nu_y q_{yx} \min(1, r(y,x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

GIBBS SAMPLER

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \dots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law
$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j \neq i} = \left(x_n^{(j)}\right)_{j \neq i}\right).$$
3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \dots, x_{n+1}^{(i)}, \dots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Metropolis-Hastings or Gibbs sampling ?

Metropolis-Hastings

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:
$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$
3. Draw $u \sim \mathcal{U}(0, 1)$.
If $u \leq r$, set $x_{i+1} := y_{i+1}$.
otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x, y)) = \nu_y q_{yx} \min(1, r(y, x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

Gibbs sampler

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \dots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law
$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j \neq i} = \left(x_n^{(j)}\right)_{j \neq i}\right).$$
3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \dots, x_{n+1}^{(i)}, \dots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Metropolis-Hastings or Gibbs sampling ?

METROPOLIS-HASTINGS

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:
$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$
3. Draw $u \sim \mathcal{U}(0, 1)$.
If $u \leq r$, set $x_{i+1} := y_{i+1}$.
otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x, y)) = \nu_y q_{yx} \min(1, r(y, x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

GIBBS SAMPLER

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \dots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law
$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j\neq i} = \left(x_n^{(j)}\right)_{j\neq i}\right).$$
3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \dots, x_{n+1}^{(i)}, \dots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Metropolis-Hastings or Gibbs sampling ?

METROPOLIS-HASTINGS

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:
$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$
3. Draw $u \sim \mathcal{U}(0, 1)$.
If $u \leq r$, set $x_{i+1} := y_{i+1}$.
otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x, y)) = \nu_y q_{yx} \min(1, r(y, x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

GIBBS SAMPLER

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \dots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law
$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j \neq i} = \left(x_n^{(j)}\right)_{j \neq i}\right).$$
3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \dots, x_{n+1}^{(i)}, \dots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Metropolis-Hastings or Gibbs sampling ?

METROPOLIS-HASTINGS

Algorithm Initialize a first state x_0.
At step i, the chain being in state x_i,

1. Propose a next state y_{i+1} by drawing a realisation in distribution $q(x_i, \cdot)$.
2. Compute the ratio:
$$r(x_i, y_{i+1}) := \frac{\nu(y_{i+1})q(y_{i+1}, x_i)}{\nu(x_i)q(x_i, y_{i+1})}$$
3. Draw $u \sim \mathcal{U}(0,1)$.
If $u \leq r$, set $x_{i+1} := y_{i+1}$.
otherwise, keep $x_{i+1} := x_i$.

Reversibility One can check that $\nu_x q_{xy} \min(1, r(x,y)) = \nu_y q_{yx} \min(1, r(y,x))$. Hence, it converges to the stationary distribution ν.

Advantage One can use (almost) any proposal distribution q.

Drawback One needs to carefully tune q to ensure fast convergence.

GIBBS SAMPLER

Algorithm First, initialize the chain in state x_0.
At step n, $x_n = (x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)})$,

1. Draw $i \sim \mathcal{U}\{1, 2, \dots, k\}$.
2. Draw $x_{n+1}^{(i)}$ in the conditional law
$$p\left(X^{(i)} \mid \left(X^{(j)}\right)_{j \neq i} = \left(x_n^{(j)}\right)_{j \neq i}\right).$$
3. Fix $x_{n+1} := \left(x_n^{(1)}, x_n^{(2)}, \dots, x_{n+1}^{(i)}, \dots, x_n^{(k)}\right)$.

Reversibility One can check that $\nu_x p_{xy} = \nu_y p_{yx}$. Hence, it converges to the stationary distribution ν.

Advantage It is generally assumed that it converges faster.

Drawback One needs to know how to sample step 2.

Introduction

Exponential families and prior conjugacy

Definition 1

A family of probability distributions parametrized by a parameter θ is called an exponential family if its probability mass function, or density, can be expressed as

$$f(x|\theta) = h(x)e^{\eta(\theta)^t T(x) - A(\eta(\theta))}$$

Property 1

These exponential families admit conjugate priors that belong to another exponential family. I.e. if

$$X|\eta \sim f_\eta, \quad \text{where } f_\eta \text{ belongs to an exponential family } \mathcal{F}(\eta)$$

then there exists another exponential family $\mathcal{H}$ such that if $g \in \mathcal{H}$ and if

$$\eta \sim g, \quad \text{the posterior is given by}$$

$$\eta|X \sim h, \quad \text{where } h \text{ belongs to the same family } \mathcal{H}.$$

Introduction

Exponential families and prior conjugacy

Definition 1

A family of probability distributions parametrized by a parameter θ is called an exponential family if its probability mass function, or density, can be expressed as

$$f(x|\theta) = h(x)e^{\eta(\theta)^t T(x) - A(\eta(\theta))}$$

Property 1

These exponential families admit conjugate priors that belong to another exponential family. I.e. if

$$X|\eta \sim f_\eta, \quad \text{where } f_\eta \text{ belongs to an exponential family } \mathcal{F}(\eta)$$

then there exists another exponential family $\mathcal{H}$ such that if $g \in \mathcal{H}$ and if

$$\eta \sim g, \quad \text{the posterior is given by}$$

$$\eta|X \sim h, \quad \text{where } h \text{ belongs to the same family } \mathcal{H}.$$

Introduction

Exponential families and prior conjugacy

Definition 1

A family of probability distributions parametrized by a parameter θ is called an exponential family if its probability mass function, or density, can be expressed as

$$f(x|\theta) = h(x)e^{\eta(\theta)^t T(x) - A(\eta(\theta))}$$

Property 1

These exponential families admit conjugate priors that belong to another exponential family. I.e. if

$$X|\eta \sim f_\eta, \quad \text{where } f_\eta \text{ belongs to an exponential family } \mathcal{F}(\eta)$$

then there exists another exponential family $\mathcal{H}$ such that if $g \in \mathcal{H}$ and if

$$\eta \sim g, \quad \text{the posterior is given by}$$

$$\eta|X \sim h, \quad \text{where } h \text{ belongs to the same family } \mathcal{H}.$$

Introduction

Exponential families and prior conjugacy

Definition 1

A family of probability distributions parametrized by a parameter θ is called an exponential family if its probability mass function, or density, can be expressed as

$$f(x|\theta) = h(x)e^{\eta(\theta)^t T(x) - A(\eta(\theta))}$$

Property 1

These exponential families admit conjugate priors that belong to another exponential family. I.e. if

$$X|\eta \sim f_\eta, \quad \textit{where } f_\eta \textit{ belongs to an exponential family } \mathcal{F}(\eta)$$

then there exists another exponential family $\mathcal{H}$ such that if $g \in \mathcal{H}$ and if

$$\eta \sim g, \quad \textit{the posterior is given by}$$

$$\eta|X \sim h, \quad \textit{where } h \textit{ belongs to the same family } \mathcal{H}.$$

Example 1

The exponential distribution is conjugate to a Gamma distribution.

$$\begin{aligned} \textit{Assume that} \quad & X|\lambda \sim \mathcal{E}(\lambda) \\ \textit{and} \quad & \lambda|\alpha,\beta \sim \Gamma(\alpha,\beta) \\ \implies \quad & \lambda|X,\alpha,\beta \sim \Gamma(\alpha+1,\beta+X) \quad . \end{aligned}$$

Introduction

Exponential families and prior conjugacy

Definition 1

A family of probability distributions parametrized by a parameter θ is called an exponential family if its probability mass function, or density, can be expressed as

$$f(x|\theta) = h(x)e^{\eta(\theta)^t T(x) - A(\eta(\theta))}$$

Property 1

These exponential families admit conjugate priors that belong to another exponential family. I.e. if

$$X|\eta \sim f_\eta, \quad \textit{where } f_\eta \textit{ belongs to an exponential family } \mathcal{F}(\eta)$$

then there exists another exponential family $\mathcal{H}$ such that if $g \in \mathcal{H}$ and if

$$\eta \sim g, \quad \textit{the posterior is given by}$$

$$\eta|X \sim h, \quad \textit{where } h \textit{ belongs to the same family } \mathcal{H}.$$

Example 1

The family of Poisson distributions $(\mathcal{P}(\lambda))_{\lambda \in \mathbb{R}^+}$ is an exponential family.
The Poisson distribution is conjugate to a Gamma distribution.

$$\begin{aligned} & X|\lambda \sim \mathcal{P}(\lambda) \\ & \lambda|\alpha, \beta \ \sim \Gamma(\alpha, \beta) \\ \implies & \lambda|\alpha, \beta, x \ \sim \Gamma(\alpha + x, \beta + 1) \ . \end{aligned}$$

Models

Models of substitution

They model aa substitutions ($S = 20$), but for simplicity, it would give the following for nt ($S = 4$):

$$\text{choose } \rho = \begin{pmatrix} \cdot & \rho_{12} & \rho_{13} & \rho_{14} \\ \vdots & \ddots & \rho_{23} & \rho_{24} \\ \vdots & & \ddots & \rho_{34} \\ \cdot & \dots & \dots & \cdot \end{pmatrix} \text{ the relative exchangeabilities}$$

$$\text{and} \quad \pi = (\pi_1, \pi_2, \pi_3, .) \text{ the equilibrium frequencies, or ``profile''}$$

$$\Longrightarrow \; Q = \begin{pmatrix} \cdot & \rho_{12}\pi_2 & \rho_{13}\pi_3 & \rho_{14}\pi_4 \\ \rho_{12}\pi_1 & \cdot & \rho_{23}\pi_3 & \rho_{24}\pi_4 \\ \rho_{13}\pi_1 & \rho_{23}\pi_2 & \cdot & \rho_{34}\pi_4 \\ \rho_{14}\pi_1 & \rho_{24}\pi_2 & \rho_{34}\pi_3 & \cdot \end{pmatrix}$$

- Two model configurations:
  - SUB all sites in the alignment are evolving under the same substitution process (i.e. share π and ρ).
  - MAX each site is evolving under its substitution process.
- Three possibilities for ρ:
  - WAG all ρ_{ab} values are fixed to a known value from the literature.
  - Poisson all ρ_{ab} values are fixed equal to 1.
  - GTR all ρ_{ab} values are free parameters.
- Leading to 6 possibilities.

Models

Models of substitution

They model aa substitutions ($S = 20$), but for simplicity, it would give the following for nt ($S = 4$):

$$\text{choose } \rho = \begin{pmatrix} \cdot & \rho_{12} & \rho_{13} & \rho_{14} \\ \vdots & \ddots & \rho_{23} & \rho_{24} \\ \vdots & & \ddots & \rho_{34} \\ \cdot & \cdots & \cdots & \cdot \end{pmatrix} \text{ the relative exchangeabilities}$$

$$\text{and} \quad \pi = (\pi_1, \pi_2, \pi_3, .) \text{ the equilibrium frequencies, or ``profile''}$$

$$\Longrightarrow \quad Q = \begin{pmatrix} \cdot & \rho_{12}\pi_2 & \rho_{13}\pi_3 & \rho_{14}\pi_4 \\ \rho_{12}\pi_1 & \cdot & \rho_{23}\pi_3 & \rho_{24}\pi_4 \\ \rho_{13}\pi_1 & \rho_{23}\pi_2 & \cdot & \rho_{34}\pi_4 \\ \rho_{14}\pi_1 & \rho_{24}\pi_2 & \rho_{34}\pi_3 & \cdot \end{pmatrix}$$

- Two model configurations:

  SUB all sites in the alignment are evolving under the same substitution process (i.e. share π and ρ).
  MAX each site is evolving under its substitution process.

- Three possibilities for ρ:

  WAG all ρ_{ab} values are fixed to a known value from the literature.
  Poisson all ρ_{ab} values are fixed equal to 1.
  GTR all ρ_{ab} values are free parameters.

- Leading to 6 possibilities.

Models

Models of substitution

They model aa substitutions ($S = 20$), but for simplicity, it would give the following for nt ($S = 4$):

$$\text{choose } \rho = \begin{pmatrix} \cdot & \rho_{12} & \rho_{13} & \rho_{14} \\ \vdots & \ddots & \rho_{23} & \rho_{24} \\ \vdots & & \ddots & \rho_{34} \\ \cdot & \dots & \dots & \cdot \end{pmatrix} \text{ the relative exchangeabilities}$$

$$\text{and} \quad \pi = (\pi_1, \pi_2, \pi_3, .) \text{ the equilibrium frequencies, or ``profile''}$$

$$\implies Q = \begin{pmatrix} \cdot & \rho_{12}\pi_2 & \rho_{13}\pi_3 & \rho_{14}\pi_4 \\ \rho_{12}\pi_1 & \cdot & \rho_{23}\pi_3 & \rho_{24}\pi_4 \\ \rho_{13}\pi_1 & \rho_{23}\pi_2 & \cdot & \rho_{34}\pi_4 \\ \rho_{14}\pi_1 & \rho_{24}\pi_2 & \rho_{34}\pi_3 & \cdot \end{pmatrix}$$

- Two model configurations:

  SUB all sites in the alignment are evolving under the same substitution process (i.e. share π and ρ).
  MAX each site is evolving under its substitution process.

- Three possibilities for ρ:

  WAG all ρ_{ab} values are fixed to a known value from the literature.
  Poisson all ρ_{ab} values are fixed equal to 1.
  GTR all ρ_{ab} values are free parameters.

- Leading to 6 possibilities.

Models

Models of substitution

They model aa substitutions ($S = 20$), but for simplicity, it would give the following for nt ($S = 4$):

$$\text{choose } \rho = \begin{pmatrix} \cdot & \rho_{12} & \rho_{13} & \rho_{14} \\ \vdots & \ddots & \rho_{23} & \rho_{24} \\ \vdots & & \ddots & \rho_{34} \\ \cdot & \cdots & \cdots & \cdot \end{pmatrix} \text{ the relative exchangeabilities}$$

$$\text{and} \quad \pi = (\pi_1, \pi_2, \pi_3, .) \text{ the equilibrium frequencies, or ``profile''}$$

$$\Longrightarrow \quad Q = \begin{pmatrix} \cdot & \rho_{12}\pi_2 & \rho_{13}\pi_3 & \rho_{14}\pi_4 \\ \rho_{12}\pi_1 & \cdot & \rho_{23}\pi_3 & \rho_{24}\pi_4 \\ \rho_{13}\pi_1 & \rho_{23}\pi_2 & \cdot & \rho_{34}\pi_4 \\ \rho_{14}\pi_1 & \rho_{24}\pi_2 & \rho_{34}\pi_3 & \cdot \end{pmatrix}$$

- Two model configurations:

  SUB all sites in the alignment are evolving under the same substitution process (i.e. share π and ρ).
  MAX each site is evolving under its substitution process.

- Three possibilities for ρ:

  WAG all ρ_{ab} values are fixed to a known value from the literature.
  Poisson all ρ_{ab} values are fixed equal to 1.
  GTR all ρ_{ab} values are free parameters.

- Leading to 6 possibilities.

Models

Priors

branch lengths l independent $\mathcal{E}(\beta)$ prior distribution.

$$p(l) = \prod_j \beta e^{-\beta l_j}$$

site-specific rates r $\Gamma(\alpha, \alpha)$ prior distribution.

$$p(r) \propto \prod_i r_i^{\alpha - 1} e^{-\alpha r_i}$$

site-specific profiles π with flat Dirichlet prior distribution.

relative exchangeabilities when not fixed, $\mathcal{E}(1)$ prior distribution.

These are chosen because they are conjugate to the likelihood of the augmented data.

Models

Priors

branch lengths l independent $\mathcal{E}(\beta)$ prior distribution.

$$p(l) = \prod_j \beta e^{-\beta l_j}$$

site-specific rates r $\Gamma(\alpha, \alpha)$ prior distribution.

$$p(r) \propto \prod_i r_i^{\alpha - 1} e^{-\alpha r_i}$$

site-specific profiles π with flat Dirichlet prior distribution.

relative exchangeabilities when not fixed, $\mathcal{E}(1)$ prior distribution.

These are chosen because they are conjugate to the likelihood of the augmented data.

Models

Priors

branch lengths l independent $\mathcal{E}(\beta)$ prior distribution.

$$p(l) = \prod_j \beta e^{-\beta l_j}$$

site-specific rates r $\Gamma(\alpha, \alpha)$ prior distribution.

$$p(r) \propto \prod_i r_i^{\alpha - 1} e^{-\alpha r_i}$$

site-specific profiles π with flat Dirichlet prior distribution.

relative exchangeabilities when not fixed, $\mathcal{E}(1)$ prior distribution.

These are chosen because they are conjugate to the likelihood of the augmented data.

Models

Priors

branch lengths l independent $\mathcal{E}(\beta)$ prior distribution.

$$p(l) = \prod_j \beta e^{-\beta l_j}$$

site-specific rates r $\Gamma(\alpha, \alpha)$ prior distribution.

$$p(r) \propto \prod_i r_i^{\alpha - 1} e^{-\alpha r_i}$$

site-specific profiles π with flat Dirichlet prior distribution.

relative exchangeabilities when not fixed, $\mathcal{E}(1)$ prior distribution.

These are chosen because they are conjugate to the likelihood of the augmented data.

Models

Priors

branch lengths l independent $\mathcal{E}(\beta)$ prior distribution.

$$p(l) = \prod_j \beta e^{-\beta l_j}$$

site-specific rates r $\Gamma(\alpha, \alpha)$ prior distribution.

$$p(r) \propto \prod_i r_i^{\alpha - 1} e^{-\alpha r_i}$$

site-specific profiles π with flat Dirichlet prior distribution.

relative exchangeabilities when not fixed, $\mathcal{E}(1)$ prior distribution.

These are chosen because they are conjugate to the likelihood of the augmented data.

MCMC sampling

Simulation of a “mapping” in the Poisson case

1. sample the internal states in $p(X_{ij}|X_{\mathrm{leaves}})$ using two depth-first traversals of the tree before sampling the substitution process along each branch, conditioned on both extremities, i.e.
2. number of substitutions n_{ij} is Poisson distributed with parameter $r_i l_j$ if the state is the same at both extremities (otherwise, Poisson excluding $n_{ij} = 0$).
3. successive states (σ_{ij}^k) are drawn from π.

The augmented likelihood is:

$$
\begin{aligned}
p(\Xi|r, l, \pi) &= \prod_i \pi\left(\sigma_{i0}^0\right) \prod_j e^{-r_i l_j} \frac{(r_i l_j)^{n_{ij}}}{n_{ij}!} \prod_{k=1}^{n_{ij}} \pi\left(\sigma_{ij}^k\right) \\
&= \left(\prod_{a=1}^{20} \pi(a)^{w_a}\right) \left(\prod_i r_i^{u_i}\right) \left(\prod_j l_j^{v_j}\right) \left(\prod_{ij} \frac{e^{-r_i l_j}}{n_{ij}!}\right)
\end{aligned}
$$

MCMC sampling

Simulation of a “mapping” in the Poisson case

1. sample the internal states in $p(X_{ij}|X_{\text{leaves}})$ using two depth-first traversals of the tree
   before sampling the substitution process along each branch, conditioned on both extremities, i.e.
2. number of substitutions n_{ij} is Poisson distributed with parameter $r_i l_j$ if the state is the same at both extremities (otherwise, Poisson excluding $n_{ij} = 0$).
3. successive states (σ_{ij}^k) are drawn from π.

The augmented likelihood is:

$$
\begin{aligned}
p(\Xi|r, l, \pi) &= \prod_i \pi\left(\sigma_{i0}^0\right) \prod_j e^{-r_i l_j} \frac{(r_i l_j)^{n_{ij}}}{n_{ij}!} \prod_{k=1}^{n_{ij}} \pi\left(\sigma_{ij}^k\right) \\
&= \left(\prod_{a=1}^{20} \pi(a)^{w_a}\right) \left(\prod_i r_i^{u_i}\right) \left(\prod_j l_j^{v_j}\right) \left(\prod_{ij} \frac{e^{-r_i l_j}}{n_{ij}!}\right)
\end{aligned}
$$

MCMC sampling

Simulation of a “mapping” in the Poisson case

1. sample the internal states in $p(X_{ij}|X_{\text{leaves}})$ using two depth-first traversals of the tree before sampling the substitution process along each branch, conditioned on both extremities, i.e.
2. number of substitutions n_{ij} is Poisson distributed with parameter $r_i l_j$ if the state is the same at both extremities (otherwise, Poisson excluding $n_{ij} = 0$).
3. successive states (σ_{ij}^k) are drawn from π.

The augmented likelihood is:

$$
\begin{aligned}
p(\Xi|r, l, \pi) &= \prod_i \pi\left(\sigma_{i0}^0\right) \prod_j e^{-r_i l_j} \frac{(r_i l_j)^{n_{ij}}}{n_{ij}!} \prod_{k=1}^{n_{ij}} \pi\left(\sigma_{ij}^k\right) \\
&= \left(\prod_{a=1}^{20} \pi(a)^{w_a}\right) \left(\prod_i r_i^{u_i}\right) \left(\prod_j l_j^{v_j}\right) \left(\prod_{ij} \frac{e^{-r_i l_j}}{n_{ij}!}\right)
\end{aligned}
$$

MCMC sampling

Simulation of a “mapping” in the Poisson case

1. sample the internal states in $p(X_{ij}|X_{\text{leaves}})$ using two depth-first traversals of the tree before sampling the substitution process along each branch, conditioned on both extremities, i.e.
2. number of substitutions n_{ij} is Poisson distributed with parameter $r_i l_j$ if the state is the same at both extremities (otherwise, Poisson excluding $n_{ij} = 0$).
3. successive states (σ_{ij}^k) are drawn from π.

The augmented likelihood is:

$$p(\Xi|r, l, \pi) = \prod_i \pi\left(\sigma_{i0}^0\right) \prod_j e^{-r_i l_j} \frac{(r_i l_j)^{n_{ij}}}{n_{ij}!} \prod_{k=1}^{n_{ij}} \pi\left(\sigma_{ij}^k\right)$$

$$= \left(\prod_{a=1}^{20} \pi(a)^{w_a}\right) \left(\prod_i r_i^{u_i}\right) \left(\prod_j l_j^{v_j}\right) \left(\prod_{ij} \frac{e^{-r_i l_j}}{n_{ij}!}\right)$$

MCMC sampling

Conjugacy: an illustration with l

- Recall that the prior is exponential (special case of Gamma),

$$p(l) \propto \prod_j e^{-\beta l_j}$$

- And observe that the (part we are interested in in the) augmented likelihood is Poisson,

$$\begin{aligned} p(\Xi | r, l, \pi) &= \left(\prod_{a=1}^{20} \pi(a)^{w_a} \right) \left(\prod_i r_i^{u_i} \right) \left(\prod_j l_j^{v_j} \right) \left(\prod_{ij} \frac{e^{-r_i l_j}}{n_{ij}!} \right) \\ &\propto \prod_j l_j^{v_j} e^{-R l_j} \end{aligned}$$

- Resulting posterior is Gamma again,

$$p(l | \Xi, r, \pi) \propto \prod_j l_j^{v_j} e^{-(\beta + R) l_j}$$

MCMC sampling

Conjugacy: an illustration with l

- Recall that the prior is exponential (special case of Gamma),

$$p(l) \propto \prod_j e^{-\beta l_j}$$

- And observe that the (part we are interested in in the) augmented likelihood is Poisson,

$$\begin{aligned} p(\Xi | r, l, \pi) &= \left(\prod_{a=1}^{20} \pi(a)^{w_a} \right) \left(\prod_i r_i^{u_i} \right) \left(\prod_j l_j^{v_j} \right) \left(\prod_{ij} \frac{e^{-r_i l_j}}{n_{ij}!} \right) \\ &\propto \prod_j l_j^{v_j} e^{-R l_j} \end{aligned}$$

- Resulting posterior is Gamma again,

$$p(l | \Xi, r, \pi) \propto \prod_j l_j^{v_j} e^{-(\beta + R) l_j}$$

MCMC sampling

Conjugacy: an illustration with l

- Recall that the prior is exponential (special case of Gamma),

$$p(l) \propto \prod_j e^{-\beta l_j}$$

- And observe that the (part we are interested in in the) augmented likelihood is Poisson,

$$\begin{aligned} p(\Xi | r, l, \pi) &= \left(\prod_{a=1}^{20} \pi(a)^{w_a} \right) \left(\prod_i r_i^{u_i} \right) \left(\prod_j l_j^{v_j} \right) \left(\prod_{ij} \frac{e^{-r_i l_j}}{n_{ij}!} \right) \\ &\propto \prod_j l_j^{v_j} e^{-R l_j} \end{aligned}$$

- Resulting posterior is Gamma again,

$$p(l | \Xi, r, \pi) \propto \prod_j l_j^{v_j} e^{-(\beta + R) l_j}$$

MCMC sampling

Conjugacy: an illustration with l

- Recall that the prior is exponential (special case of Gamma),

$$p(l) \propto \prod_j e^{-\beta l_j}$$

- And observe that the (part we are interested in in the) augmented likelihood is Poisson,

$$p(\Xi|r, l, \pi) = \left(\prod_{a=1}^{20} \pi(a)^{w_a}\right) \left(\prod_i r_i^{u_i}\right) \left(\prod_j l_j^{v_j}\right) \left(\prod_{ij} \frac{e^{-r_i l_j}}{n_{ij}!}\right)$$
$$\propto \prod_j l_j^{v_j} e^{-R l_j}$$

- Resulting posterior is Gamma again,

$$p(l|\Xi, r, \pi) \propto \prod_j l_j^{v_j} e^{-(\beta+R) l_j}$$

MCMC sampling

Overall strategy

1. along each branch j and at any site i, sample n_{ij} the total number of substitutions, (t_{ij}^k) the times at which substitutions occur, and (σ_{ij}^k) the successive states.

2. sampling the branch-length l given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is a Gamma again with known parameters.

3. sampling the site-specific rates r given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is Gamma again with known parameters.

4. sampling the stationary profile π given everything else.
   The prior is a Dirichlet, conjugate to a multinomial distribution.
   Posterior is Dirichlet again with known parameters.

5. update the hyperparameters with a MH step.

MCMC sampling

Overall strategy

1. along each branch j and at any site i, sample n_{ij} the total number of substitutions, (t_{ij}^{k}) the times at which substitutions occur, and (σ_{ij}^{k}) the successive states.
2. sampling the branch-length l given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is a Gamma again with known parameters.
3. sampling the site-specific rates r given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is Gamma again with known parameters.
4. sampling the stationary profile π given everything else.
   The prior is a Dirichlet, conjugate to a multinomial distribution.
   Posterior is Dirichlet again with known parameters.
5. update the hyperparameters with a MH step.

MCMC sampling

Overall strategy

1. along each branch j and at any site i, sample n_{ij} the total number of substitutions, (t_{ij}^k) the times at which substitutions occur, and (σ_{ij}^k) the successive states.
2. sampling the branch-length l given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is a Gamma again with known parameters.
3. sampling the site-specific rates r given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is Gamma again with known parameters.
4. sampling the stationary profile π given everything else.
   The prior is a Dirichlet, conjugate to a multinomial distribution.
   Posterior is Dirichlet again with known parameters.
5. update the hyperparameters with a MH step.

MCMC sampling

Overall strategy

1. along each branch j and at any site i, sample n_{ij} the total number of substitutions, (t_{ij}^k) the times at which substitutions occur, and (σ_{ij}^k) the successive states.

2. sampling the branch-length l given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is a Gamma again with known parameters.

3. sampling the site-specific rates r given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is Gamma again with known parameters.

4. sampling the stationary profile π given everything else.
   The prior is a Dirichlet, conjugate to a multinomial distribution.
   Posterior is Dirichlet again with known parameters.

5. update the hyperparameters with a MH step.

MCMC sampling

Overall strategy

1. along each branch j and at any site i, sample n_{ij} the total number of substitutions, (t_{ij}^k) the times at which substitutions occur, and (σ_{ij}^k) the successive states.
2. sampling the branch-length l given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is a Gamma again with known parameters.
3. sampling the site-specific rates r given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is Gamma again with known parameters.
4. sampling the stationary profile π given everything else.
   The prior is a Dirichlet, conjugate to a multinomial distribution.
   Posterior is Dirichlet again with known parameters.
5. update the hyperparameters with a MH step.

MCMC sampling

Overall strategy

1. along each branch j and at any site i, sample n_{ij} the total number of substitutions, (t_{ij}^k) the times at which substitutions occur, and (σ_{ij}^k) the successive states.
2. sampling the branch-length l given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is a Gamma again with known parameters.
3. sampling the site-specific rates r given everything else.
   The prior is Gamma, conjugate to a Poisson variable.
   Posterior is Gamma again with known parameters.
4. sampling the stationary profile π given everything else.
   The prior is a Dirichlet, conjugate to a multinomial distribution.
   Posterior is Dirichlet again with known parameters.
5. update the hyperparameters with a MH step.

Evaluation and Results

Estimating the decorrelation time

1. Run M independent chains, discard Q steps an record K points: (θ_{mk}). Apply any test function to this output, and record $(x_{mk} = x(\theta_{mk}))$.
2. The last points of the M chains (x_{mK}) have unknown mean μ and variance v.
3. The within chains means $(\bar{x}_m = 1/K \sum_{k=1}^{K} x_{mk})$ have unknown mean μ an variance v', where
$$v' = \frac{v}{K_{\text{eff}}} \quad \Longrightarrow \quad K_{\text{eff}} \approx \frac{\text{Empirical variance of } (x_{mK})}{\text{Empirical variance of } (\bar{x}_m)}$$
4. Measure the time spent between two successive recorded points, τ, and compute the decorrelation time,
$$\tau_d = \frac{\tau K}{K_{\text{eff}}}$$

Evaluation and Results

Estimating the decorrelation time

1. Run M independent chains, discard Q steps an record K points: (θ_{mk}). Apply any test function to this output, and record $(x_{mk} = x(\theta_{mk}))$.

2. The last points of the M chains (x_{mK}) have unknown mean μ and variance v.

3. The within chains means $(\bar{x}_m = 1/K \sum_{k=1}^{K} x_{mk})$ have unknown mean μ an variance v', where

$$v' = \frac{v}{K_{\text{eff}}} \quad \Longrightarrow \quad K_{\text{eff}} \approx \frac{\text{Empirical variance of } (x_{mK})}{\text{Empirical variance of } (\bar{x}_m)}$$

4. Measure the time spent between two successive recorded points, τ, and compute the decorrelation time,

$$\tau_d = \frac{\tau K}{K_{\text{eff}}}$$

Evaluation and Results

Estimating the decorrelation time

1. Run M independent chains, discard Q steps an record K points: (θ_{mk}). Apply any test function to this output, and record $(x_{mk} = x(\theta_{mk}))$.
2. The last points of the M chains (x_{mK}) have unknown mean μ and variance v.
3. The within chains means $(\bar{x}_m = 1/K \sum_{k=1}^{K} x_{mk})$ have unknown mean μ an variance v', where
$$v' = \frac{v}{K_{\text{eff}}} \quad \Longrightarrow \quad K_{\text{eff}} \approx \frac{\text{Empirical variance of } (x_{mK})}{\text{Empirical variance of } (\bar{x}_m)}$$
4. Measure the time spent between two successive recorded points, τ, and compute the decorrelation time,
$$\tau_d = \frac{\tau K}{K_{\text{eff}}}$$

Evaluation and Results

Estimating the decorrelation time

1. Run M independent chains, discard Q steps an record K points: (θ_{mk}). Apply any test function to this output, and record $(x_{mk} = x(\theta_{mk}))$.

2. The last points of the M chains (x_{mK}) have unknown mean μ and variance v.

3. The within chains means $(\bar{x}_m = 1/K \sum_{k=1}^{K} x_{mk})$ have unknown mean μ an variance v', where

$$v' = \frac{v}{K_{\text{eff}}} \quad \Longrightarrow \quad K_{\text{eff}} \approx \frac{\text{Empirical variance of } (x_{mK})}{\text{Empirical variance of } (\bar{x}_m)}$$

4. Measure the time spent between two successive recorded points, τ, and compute the decorrelation time,

$$\tau_d = \frac{\tau K}{K_{\text{eff}}}$$

Evaluation and Results

Results

- Using Gibbs sampling leads to much smaller decorrelation times,
- At least one order of magnitude smaller,
- and 2-3 orders of magnitude when the substitution process is very site-specific, with lots of parameters (MAX-GTR).

Evaluation and Results

Results

- Using Gibbs sampling leads to much smaller decorrelation times,
- At least one order of magnitude smaller,
- and 2-3 orders of magnitude when the substitution process is very site-specific, with lots of parameters (MAX-GTR).

Evaluation and Results

Results

- Using Gibbs sampling leads to much smaller decorrelation times,
- At least one order of magnitude smaller,
- and 2-3 orders of magnitude when the substitution process is very site-specific, with lots of parameters (MAX-GTR).