

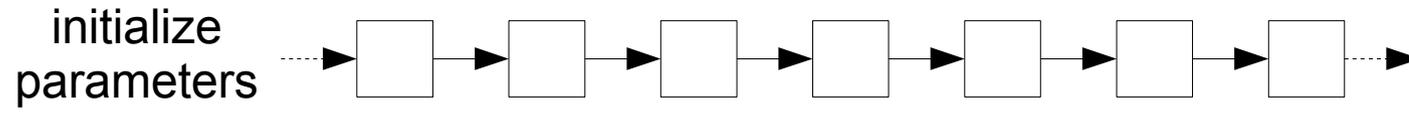
ベイジアン MCMC による 分子系統解析法

講義編

田辺晶史

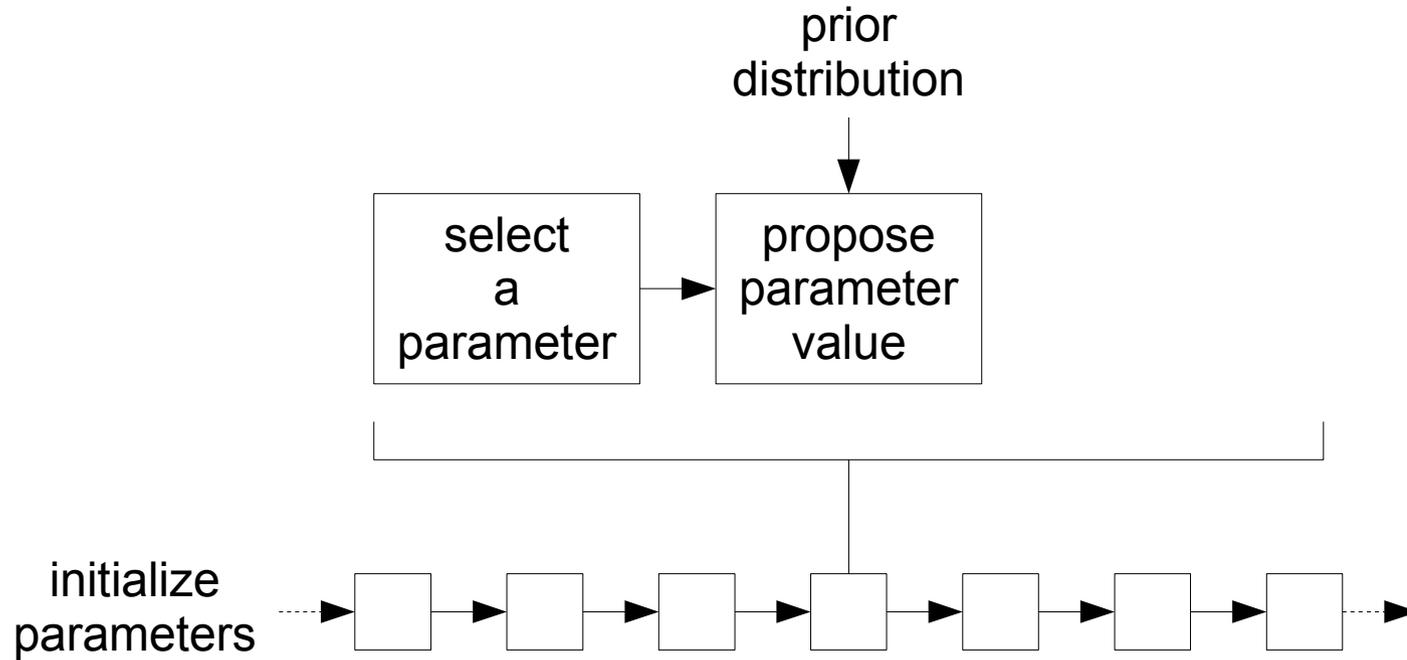
ベイジアン MCMC による系統樹推定 – M-H アルゴリズム

ベイジアン MCMC による系統樹推定 – M-H アルゴリズム

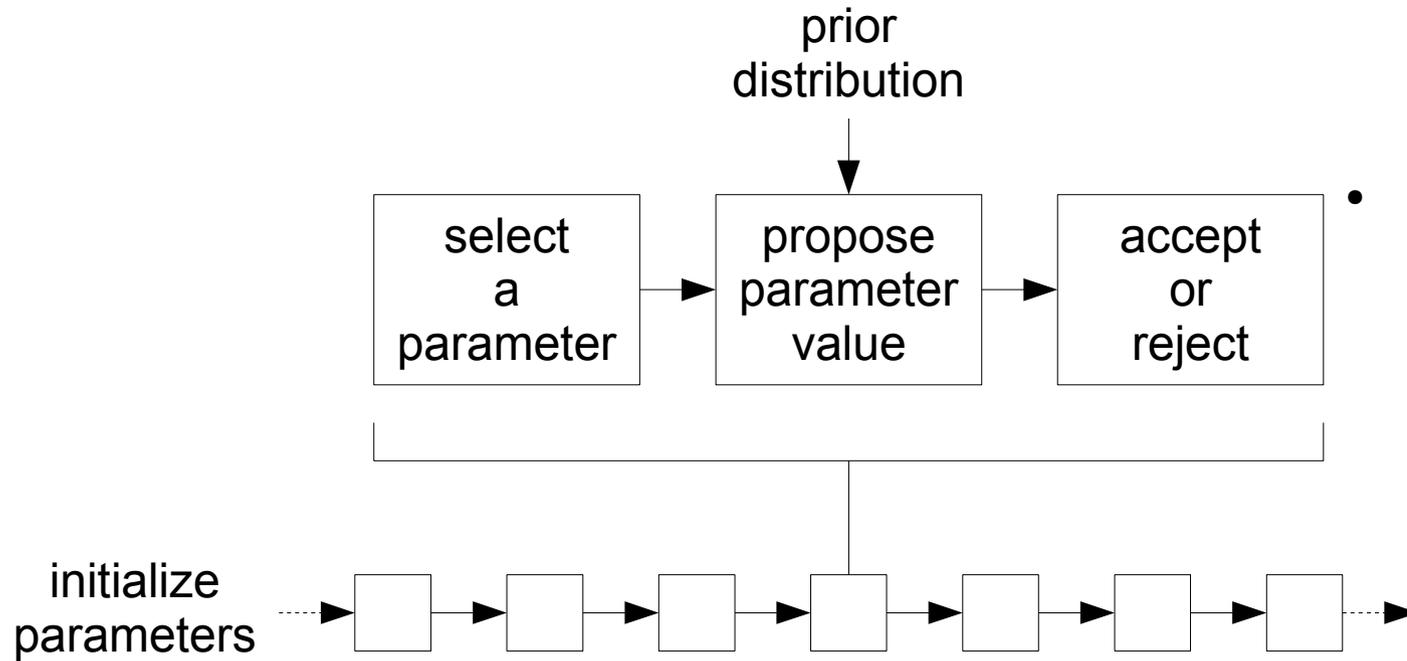


ベイジアン MCMC による系統樹推定 – M-H アルゴリズム

- 無作為にパラメータを選んで事前分布から候補を提案

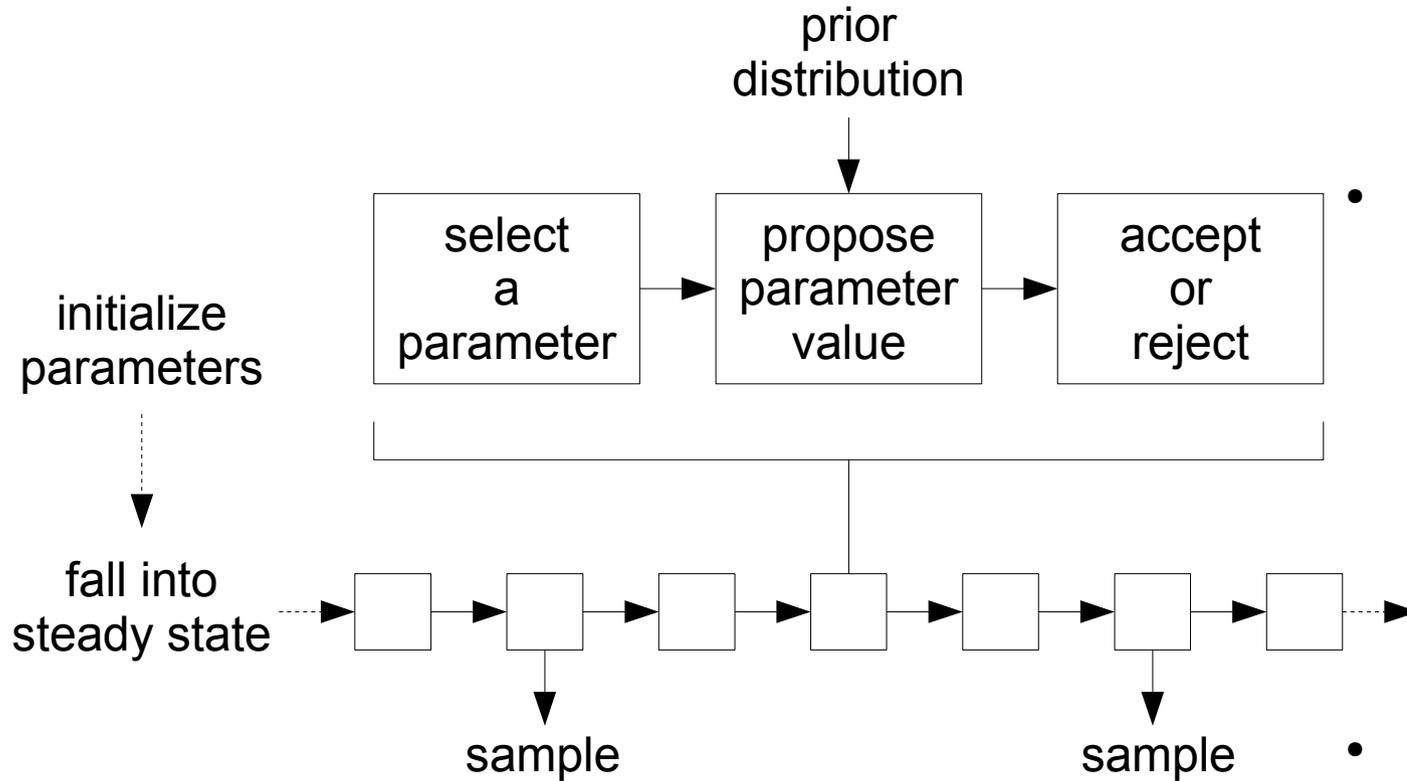


ベイジアン MCMC による系統樹推定 – M-H アルゴリズム



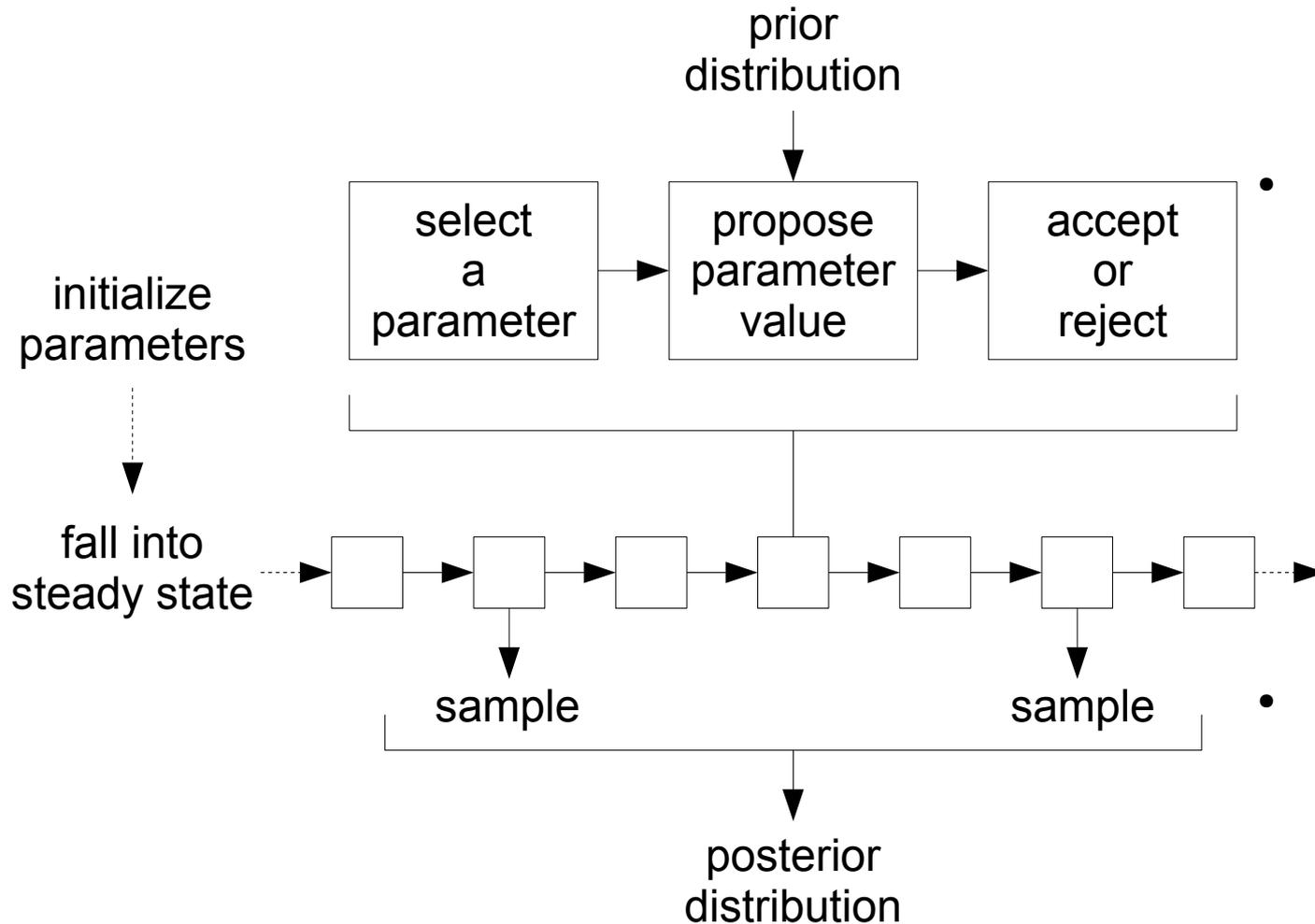
- 無作為にパラメータを選んで事前分布から候補を提案
- 候補の尤度が良ければ受理、悪くても尤度比に応じて受理

ベイジアン MCMC による系統樹推定 – M-H アルゴリズム



- 無作為にパラメータを選んで事前分布から候補を提案
- 候補の尤度が良ければ受理、悪くても尤度比に応じて受理
- 定常状態に収束後、十分に間隔を空けてサンプリングする

ベイジアン MCMC による系統樹推定 – M-H アルゴリズム

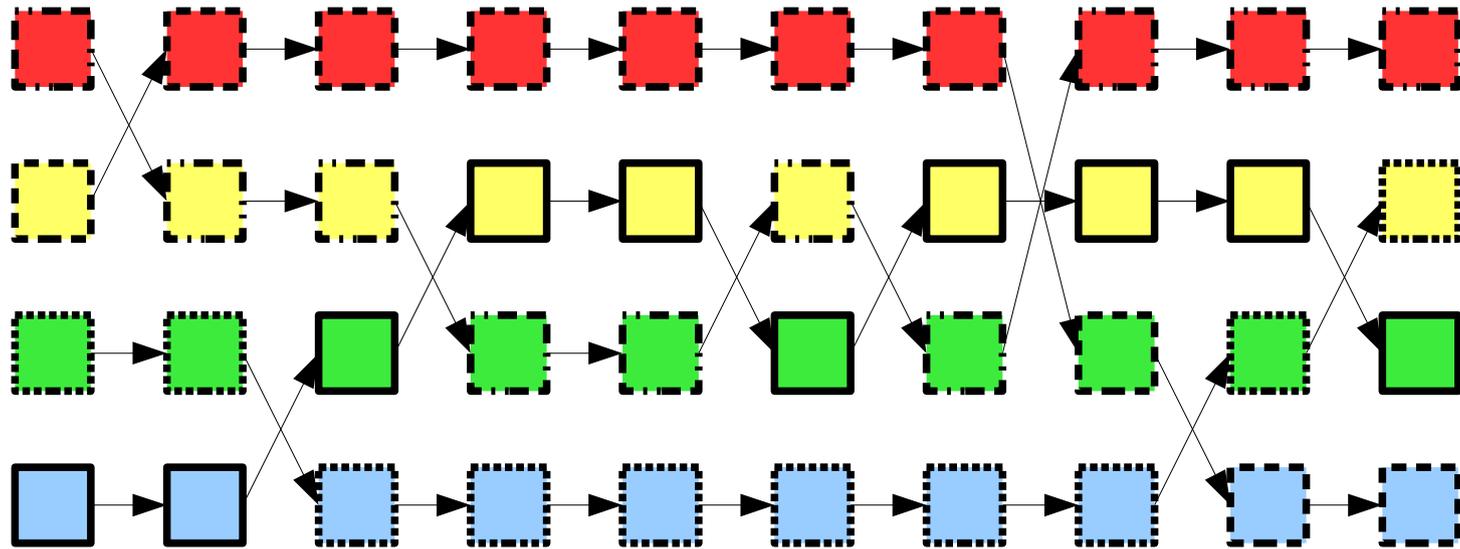


- 無作為にパラメータを選んで事前分布から候補を提案
- 候補の尤度が良ければ受理、悪くても尤度比に応じて受理

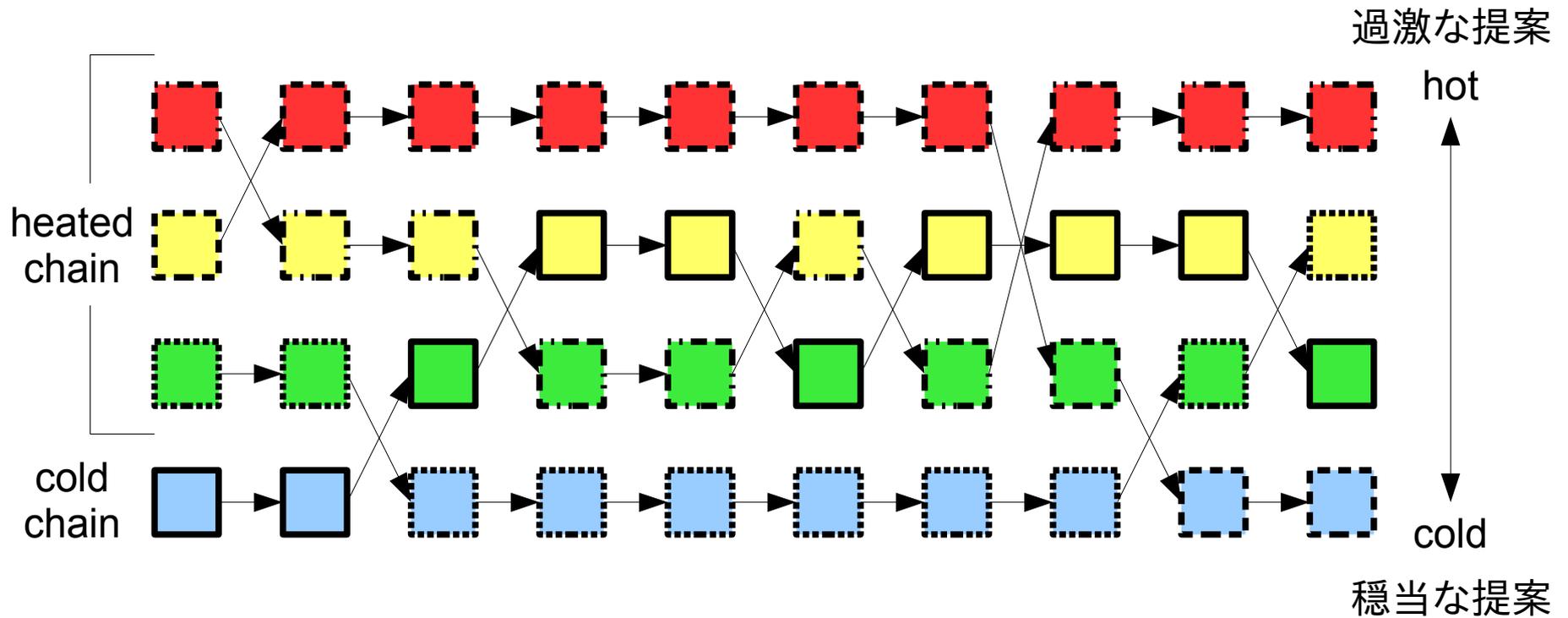
- 定常状態に収束後、十分に間隔を空けてサンプリングする
- サンプル群から事後分布を得る

ベイズアン MCMC による系統樹推定 – MC³

ベイズ MCMC による系統樹推定 - MC³

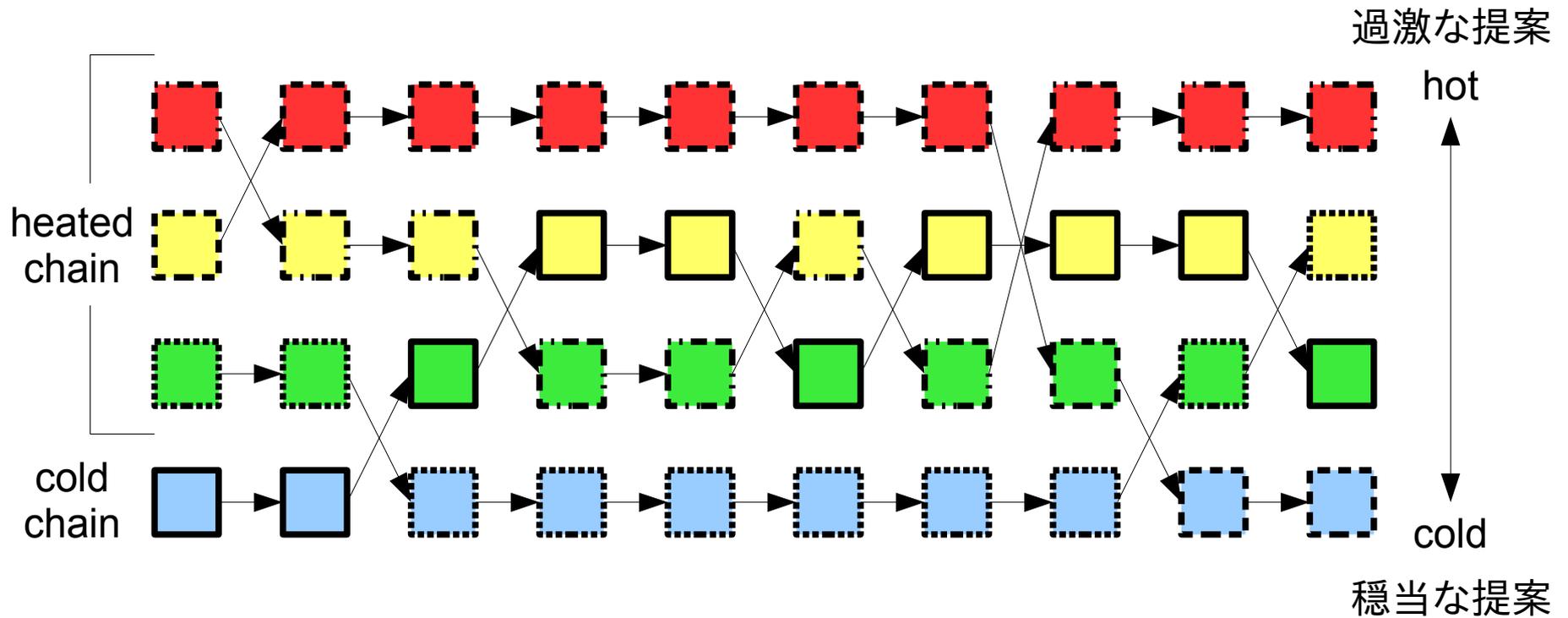


ベイジアン MCMC による系統樹推定 - MC³



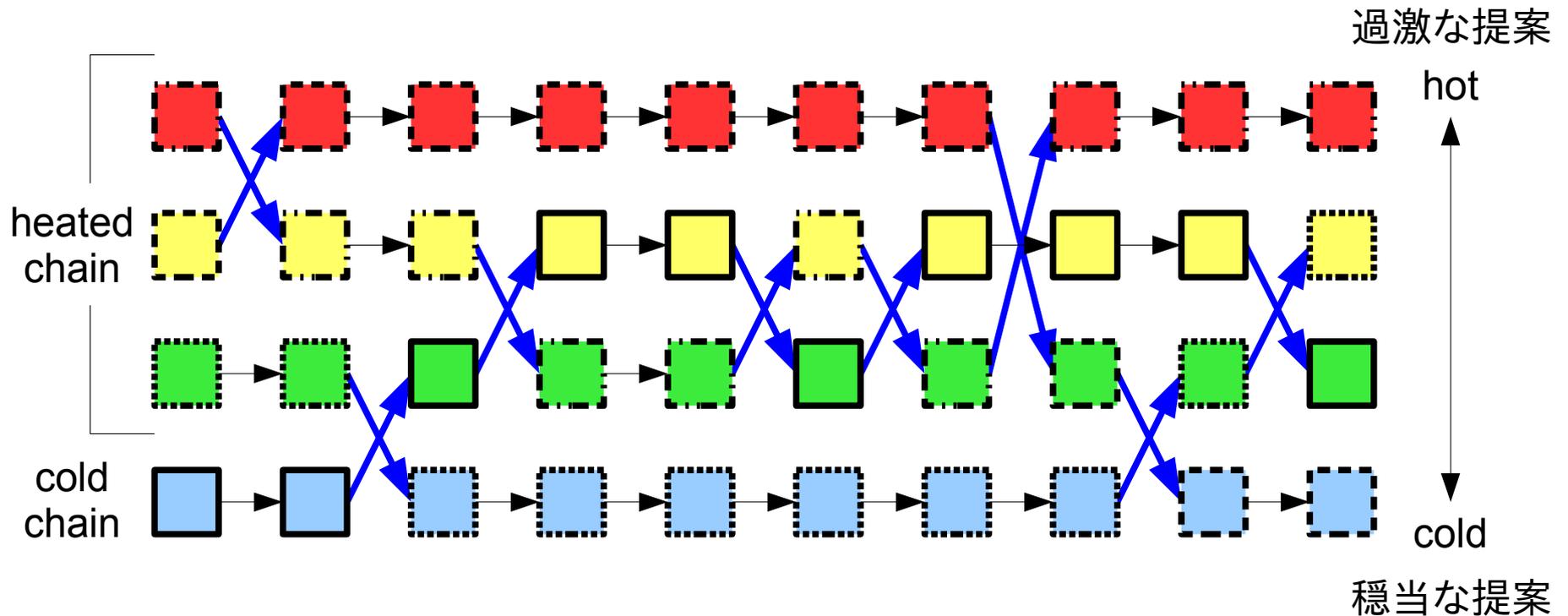
- 制約の緩い（温度の高い）MCMC を同時に走らせる

ベイジアン MCMC による系統樹推定 - MC³



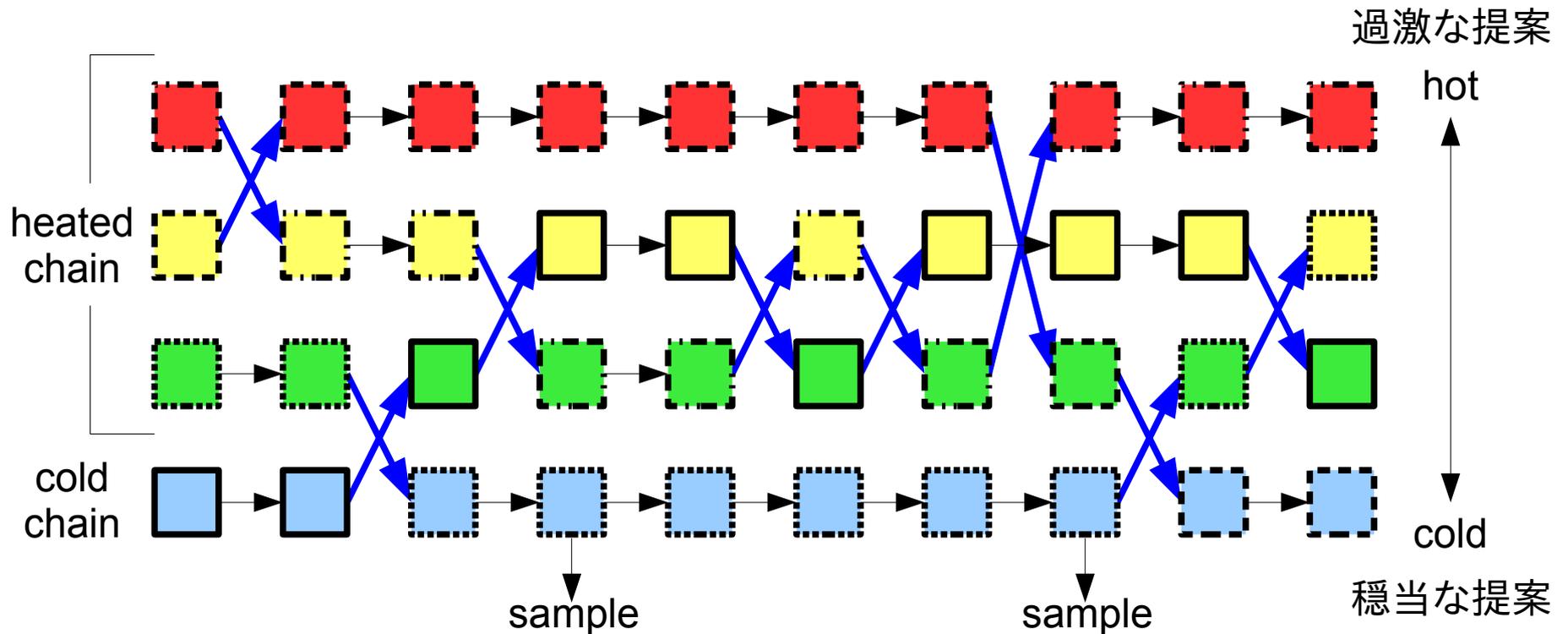
- 制約の緩い（温度の高い）MCMC を同時に走らせる
- 無作為に選んだ MCMC の組み合わせで状態交換を試行する

ベイジアン MCMC による系統樹推定 - MC³



- 制約の緩い（温度の高い）MCMC を同時に走らせる
- 無作為に選んだ MCMC の組み合わせで状態交換を試行する
- 高温の方が尤度が高ければ受理、尤度が低くても尤度比に応じて受理する

ベイジアン MCMC による系統樹推定 - MC³



- 制約の緩い（温度の高い）MCMC を同時に走らせる
- 無作為に選んだ MCMC の組み合わせで状態交換を試行する
- 高温の方が尤度が高ければ受理、尤度が低くても尤度比に応じて受理する
- 低温の MCMC からのみサンプリングする

ベイズアン MCMC による系統樹推定 – 事前分布

ベイジアンMCMCによる系統樹推定 – 事前分布

- 共役事前分布を使うもの
 - トランジション / トランスバージョン比 – ベータ分布
 - 塩基・アミノ酸頻度組成 – ディリクレ分布
 - 相対置換速度 – ディリクレ分布

ベイジアンMCMCによる系統樹推定 – 事前分布

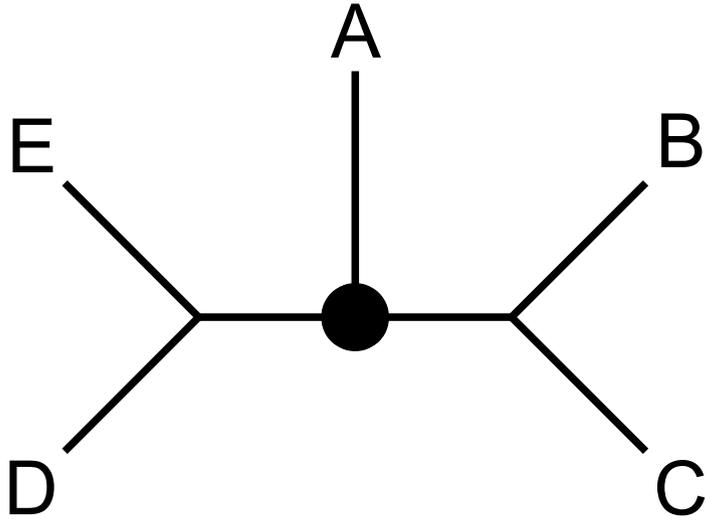
- 共役事前分布を使うもの
 - トランジション / トランスバージョン比 – ベータ分布
 - 塩基・アミノ酸頻度組成 – ディリクレ分布
 - 相対置換速度 – ディリクレ分布
- 無情報事前分布を使うもの
 - 樹形 – 一様分布
 - ガンマシェイプ – 一様分布
 - 不変座位率 – 一様分布 (ベータ分布だったかも?)

ベイジアンMCMCによる系統樹推定 – 事前分布

- 共役事前分布を使うもの
 - トランジション / トランスバージョン比 – ベータ分布
 - 塩基・アミノ酸頻度組成 – ディリクレ分布
 - 相対置換速度 – ディリクレ分布
- 無情報事前分布を使うもの
 - 樹形 – 一様分布
 - ガンマシェイプ – 一様分布
 - 不変座位率 – 一様分布 (ベータ分布だったかも?)
- その他
 - 枝長 – 指数分布 (短い枝ほど尤度への影響が大きいいため)

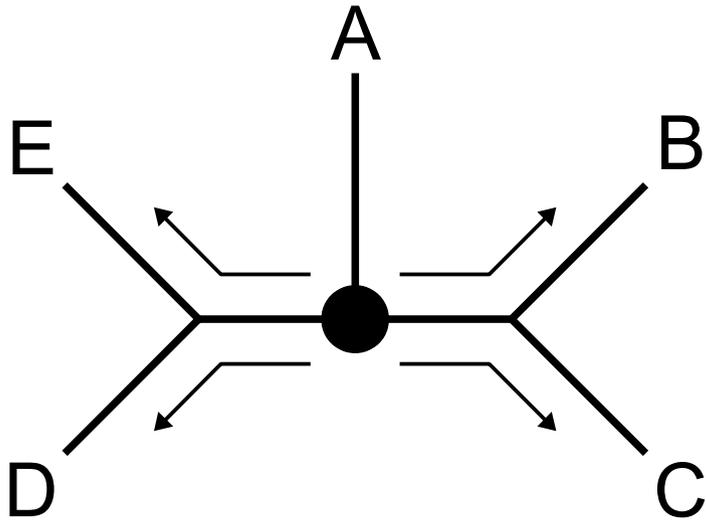
ベイズアン MCMC による系統樹推定 – 樹形の提案

ベイジアン MCMC による系統樹推定 – 樹形の提案



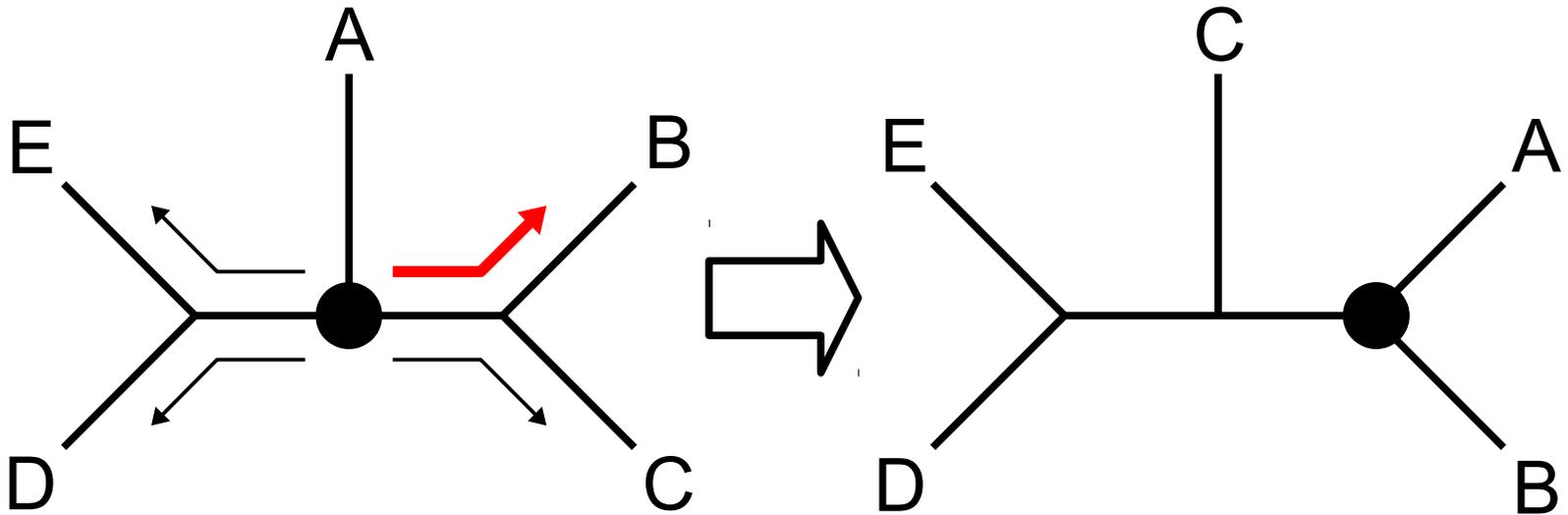
- 無作為に系統樹中のノード (●) を 1 つ選ぶ
- そのノードに接続している枝から無作為に 1 つ選ぶ (A の枝)

ベイジアン MCMC による系統樹推定 – 樹形の提案



- 無作為に系統樹中のノード (●) を 1 つ選ぶ
- そのノードに接続している枝から無作為に 1 つ選ぶ (A の枝)
- 選ばれた枝以下の部分木を選ばれなかった枝の部分木へ無作為に移動させる

ベイジアン MCMC による系統樹推定 – 樹形の提案



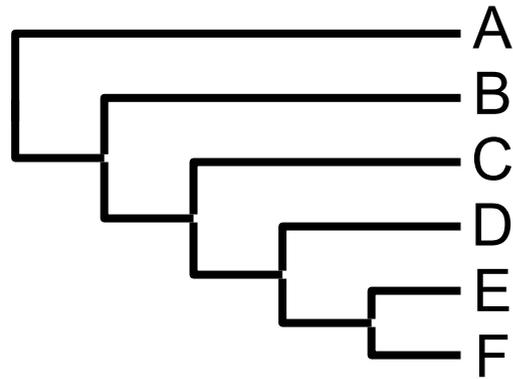
- 無作為に系統樹中のノード (●) を 1 つ選ぶ
- そのノードに接続している枝から無作為に 1 つ選ぶ (A の枝)
- 選ばれた枝以下の部分木を選ばれなかった枝の部分木へ無作為に移動させる
- 赤矢印方向へ移動させられると右図の樹形が提案される

ベイズアン MCMC による系統樹推定 – 収束の判定方法

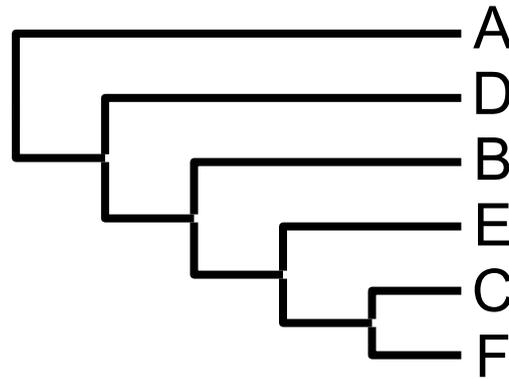
ベイズアン MCMC による系統樹推定 – 収束の判定方法

- データとモデルと事前分布が同じなら事後分布も同じはず

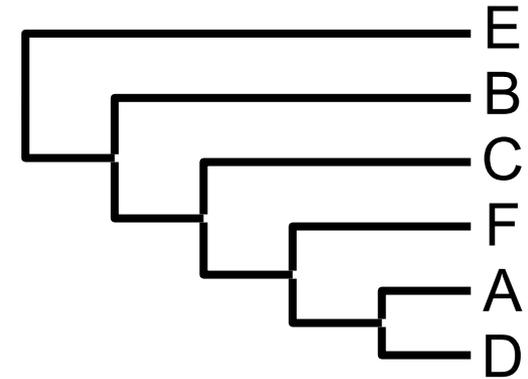
ベイジアン MCMC による系統樹推定 – 収束の判定方法



A:0.25, B:0.24



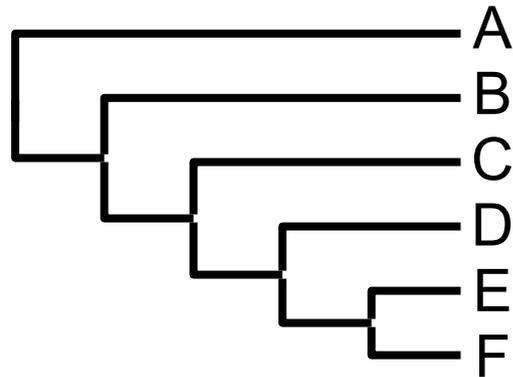
A:0.51, B:0.52



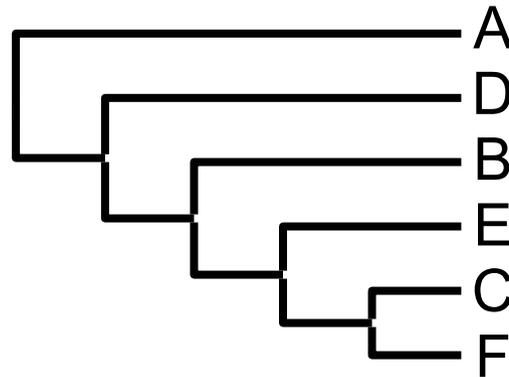
A:0.10, B:0.09

- データとモデルと事前分布が同じなら事後分布も同じはず
- 収束していれば各樹形の出現頻度が複数の MC³ で一致するはず

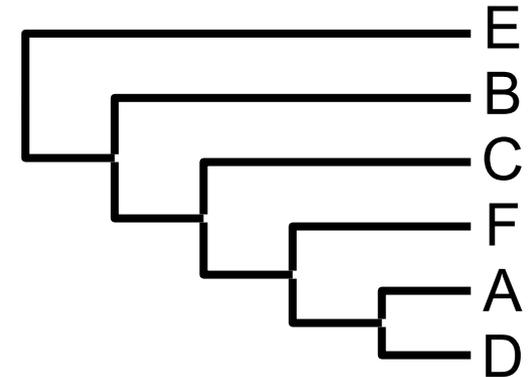
ベイジアン MCMC による系統樹推定 – 収束の判定方法



A:0.25, B:0.24



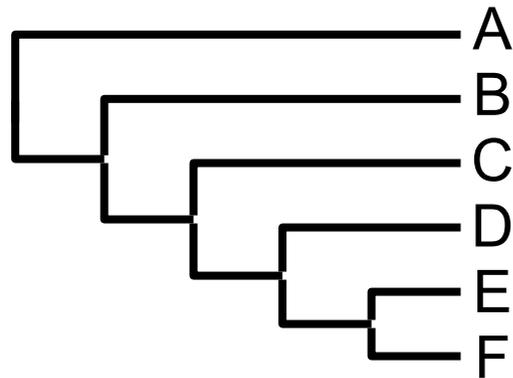
A:0.51, B:0.52



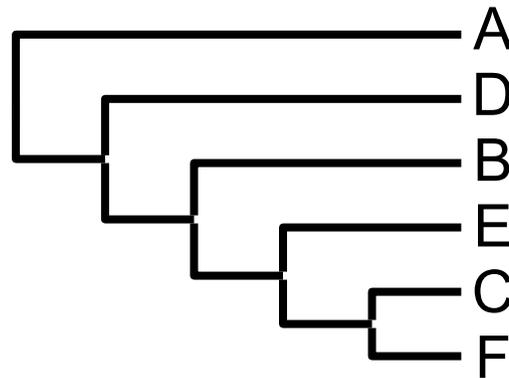
A:0.10, B:0.09

- データとモデルと事前分布が同じなら事後分布も同じはず
- 収束していれば各樹形の出現頻度が複数の MC³ で一致するはず
- 樹形の不一致度を表す指標が ASDSF

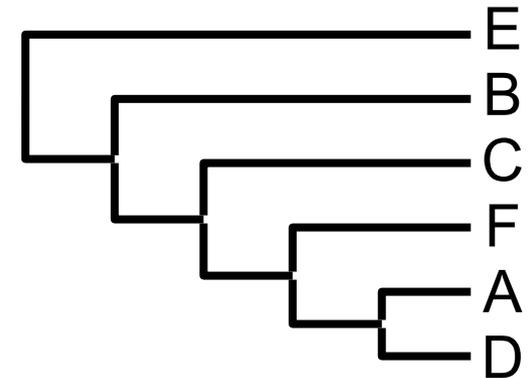
ベイジアン MCMC による系統樹推定 – 収束の判定方法



A:0.25, B:0.24



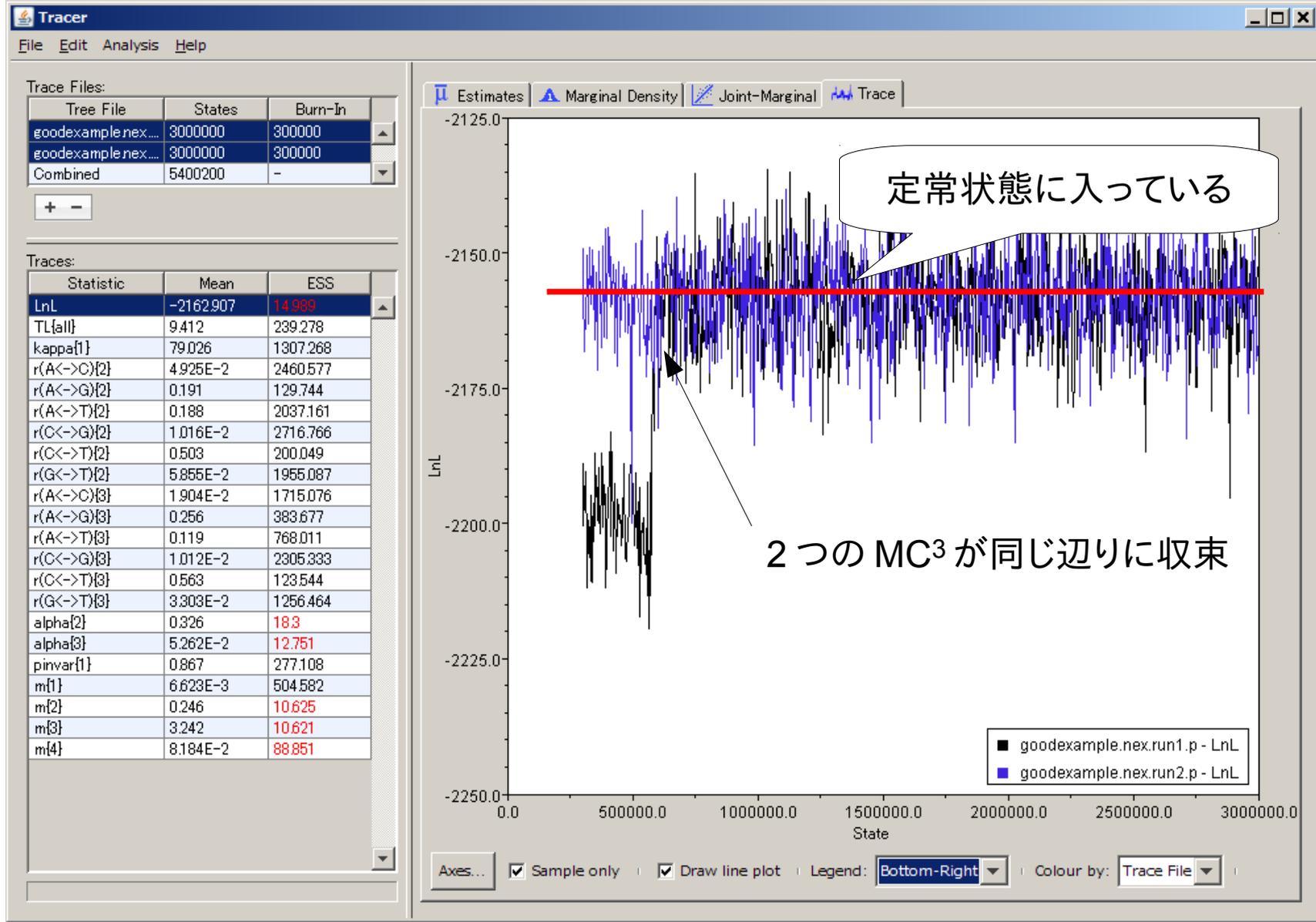
A:0.51, B:0.52



A:0.10, B:0.09

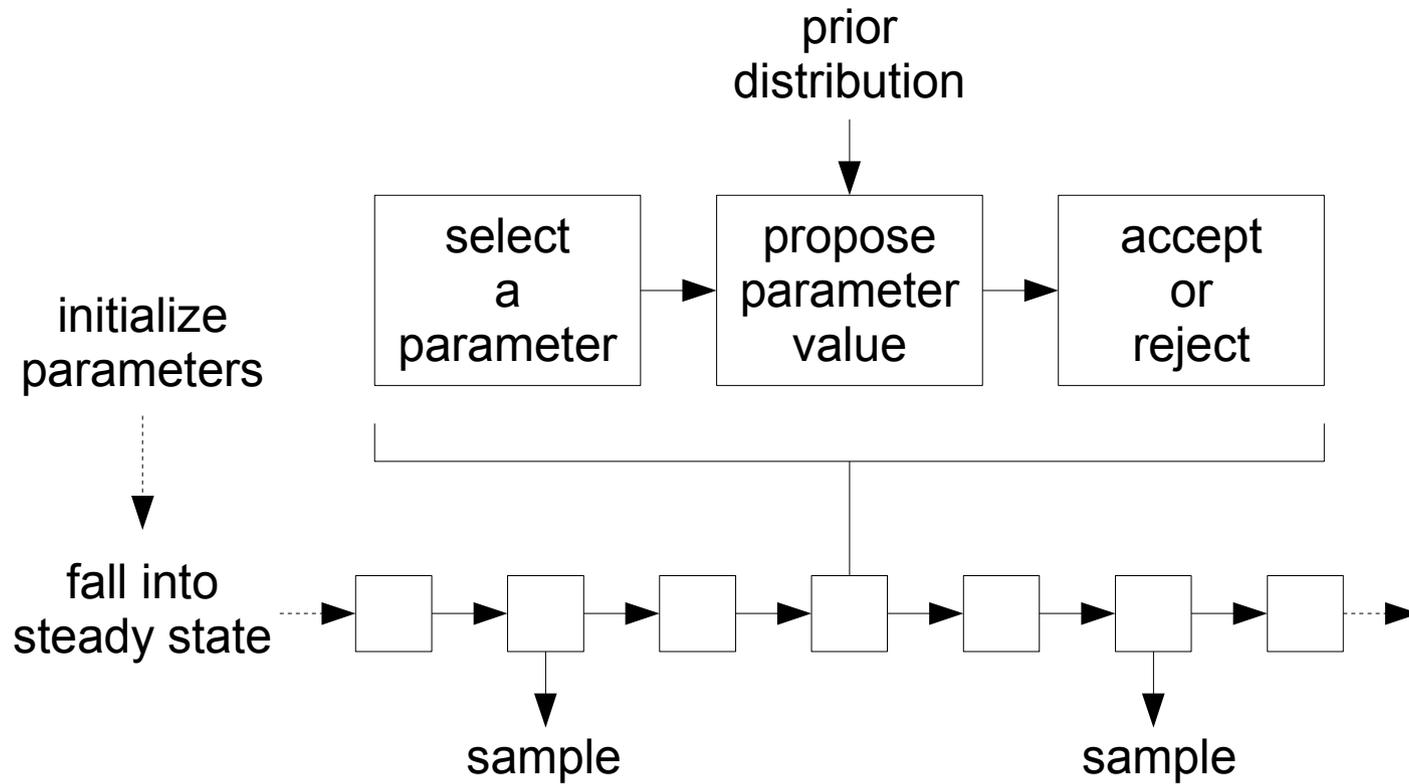
- データとモデルと事前分布が同じなら事後分布も同じはず
- 収束していれば各樹形の出現頻度が複数の MC³ で一致するはず
- 樹形の不一致度を表す指標が ASDSF
- $ASDSF < 0.01$ で定常状態に入ったとみなす

ベイジアン MCMC による系統樹推定 – 収束の判定方法

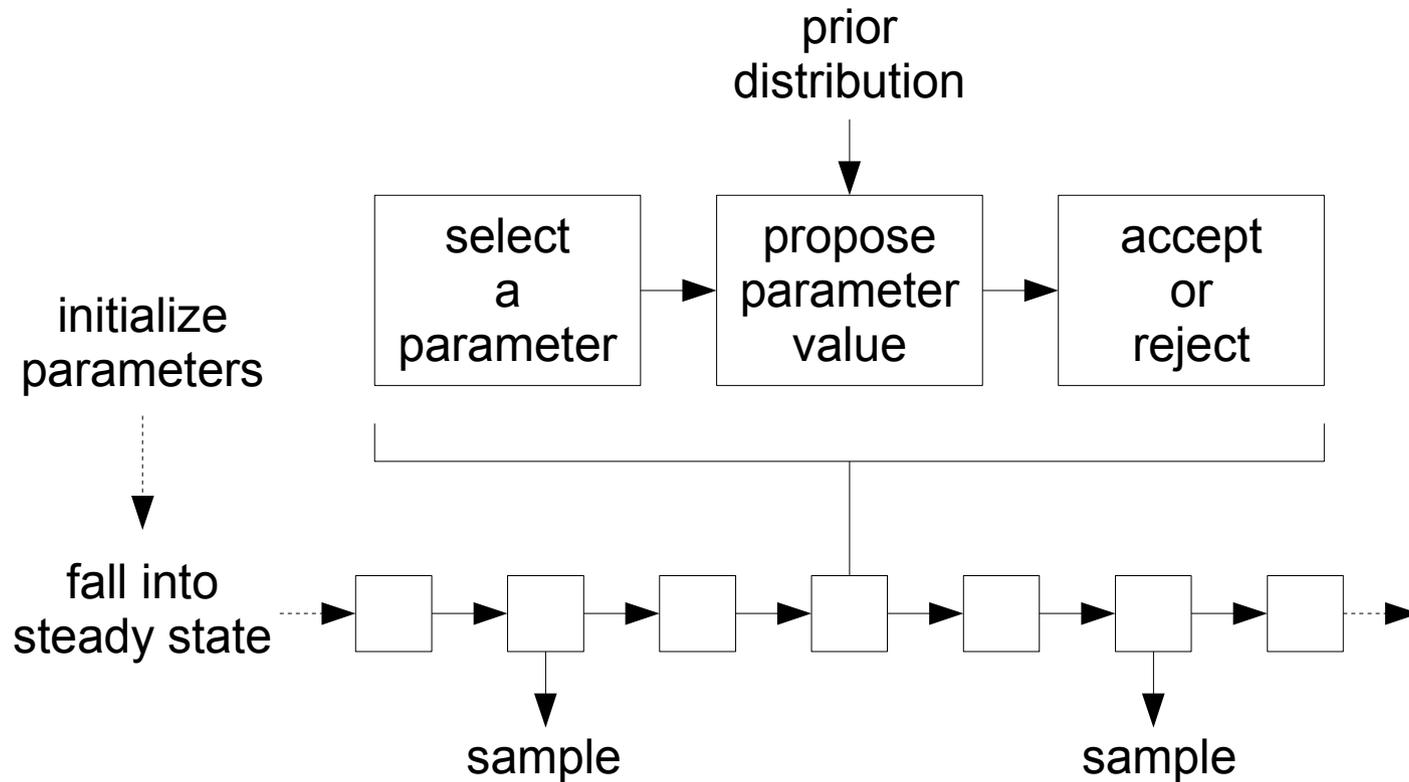


ベイジアン MCMC による系統樹推定 – ESS の推定

ベイズ MCMC による系統樹推定 – ESS の推定

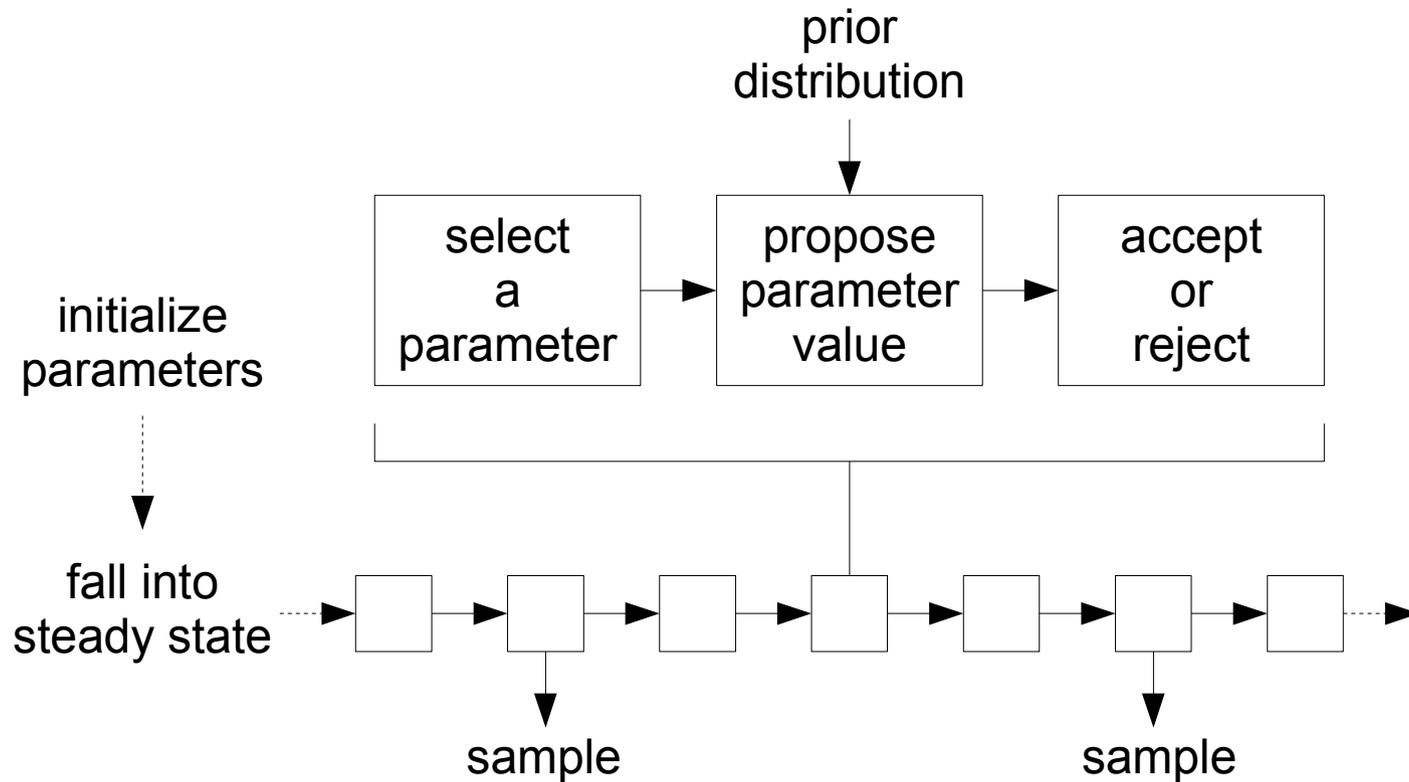


ベイズアン MCMC による系統樹推定 – ESS の推定



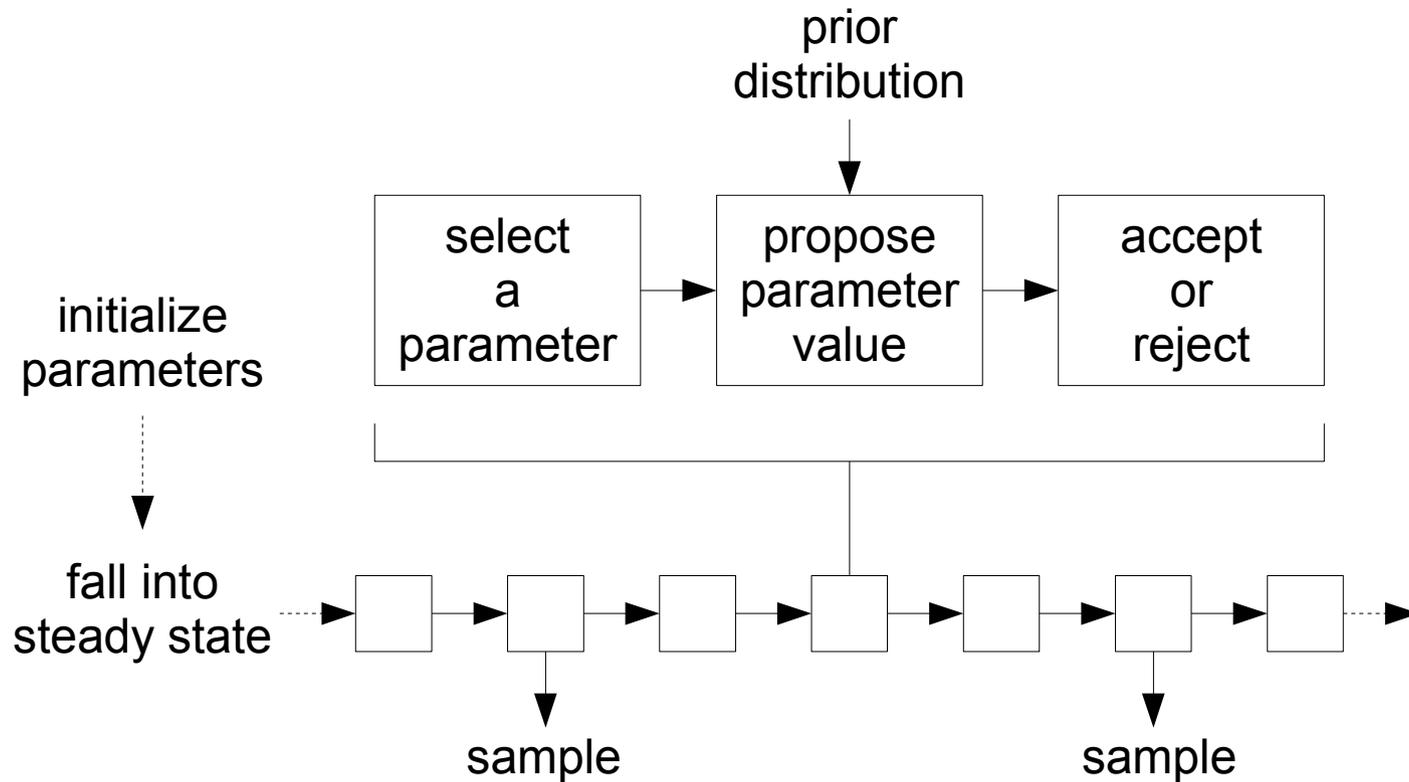
- MCMC からのサンプルは独立とみなしているが、必ずしもそうではないことがある

ベイズアン MCMC による系統樹推定 – ESS の推定



- MCMC からのサンプルは独立とみなしているが、必ずしもそうではないことがある
- サンプル間の独立性に基づいて有効なサンプルサイズを推定

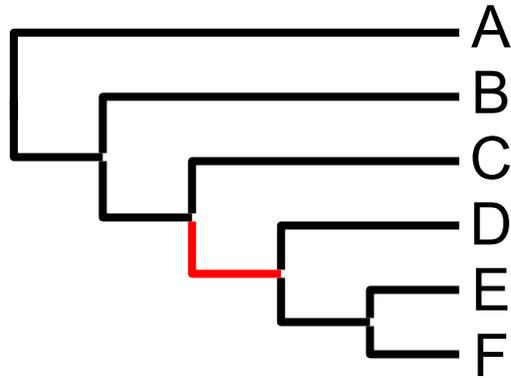
ベイズアン MCMC による系統樹推定 – ESS の推定



- MCMC からのサンプルは独立とみなしているが、必ずしもそうではないことがある
- サンプル間の独立性に基づいて有効なサンプルサイズを推定
- $ESS \geq 200$ を十分量とみなす

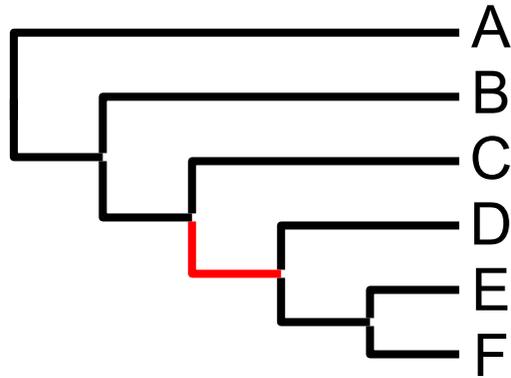
制約付き系統推定と Bayes factor による系統仮説比較

制約付き系統推定と Bayes factor による系統仮説比較



赤い枝の仮説
((A, B, C), (D, E, F))
を検証する

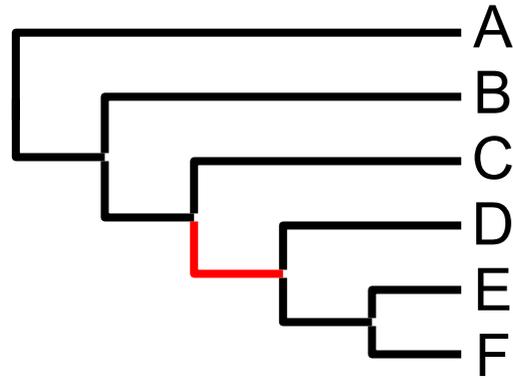
制約付き系統推定と Bayes factor による系統仮説比較



赤い枝の仮説
((A, B, C), (D, E, F))
を検証する

1. 制約無し MCMC のサンプル群から、赤い枝（仮説 1）に対立する仮説の中で、最も多く出現する仮説（仮説 0）を探す

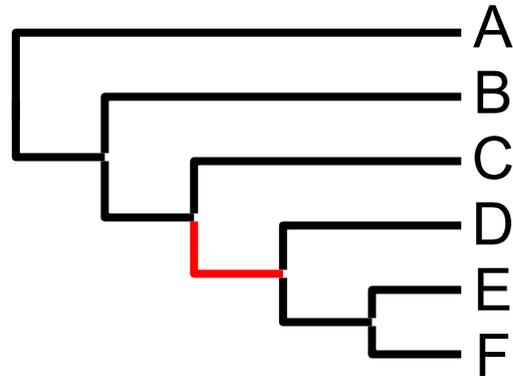
制約付き系統推定と Bayes factor による系統仮説比較



赤い枝の仮説
((A,B,C),(D,E,F))
を検証する

1. 制約無し MCMC のサンプル群から、赤い枝（仮説 1）に対立する仮説の中で、最も多く出現する仮説（仮説 0）を探す
2. 赤い枝の仮説を制約とする MCMC を実行

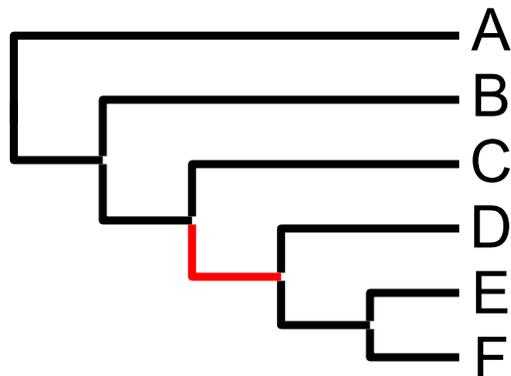
制約付き系統推定と Bayes factor による系統仮説比較



赤い枝の仮説
((A,B,C),(D,E,F))
を検証する

1. 制約無し MCMC のサンプル群から、赤い枝（仮説 1）に対立する仮説の中で、最も多く出現する仮説（仮説 0）を探す
2. 赤い枝の仮説を制約とする MCMC を実行
3. 対立仮説を制約とする MCMC を実行

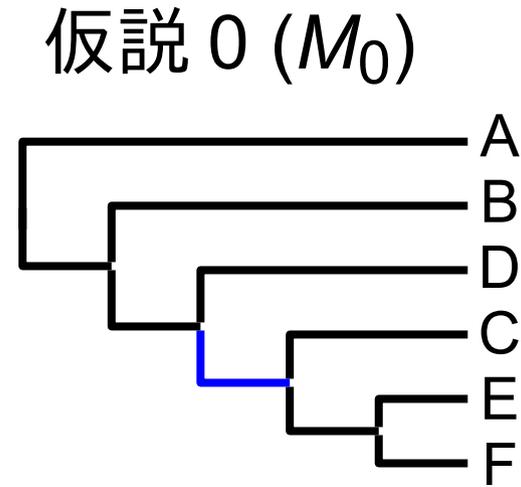
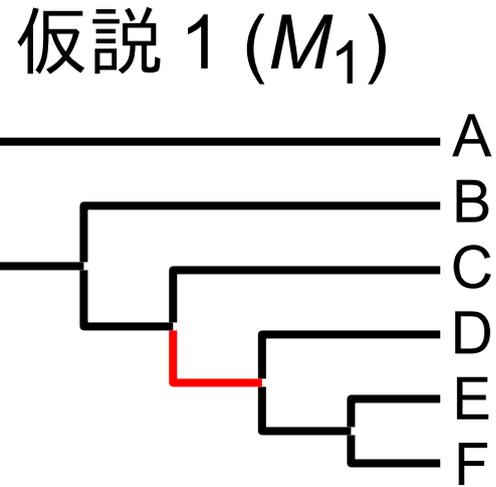
制約付き系統推定と Bayes factor による系統仮説比較



赤い枝の仮説
((A,B,C),(D,E,F))
を検証する

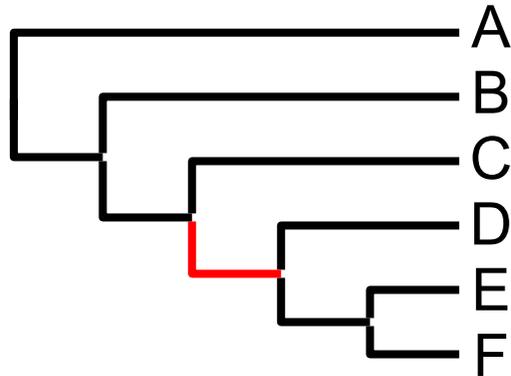
1. 制約無し MCMC のサンプル群から、赤い枝（仮説 1）に対立する仮説の中で、最も多く出現する仮説（仮説 0）を探す
2. 赤い枝の仮説を制約とする MCMC を実行
3. 対立仮説を制約とする MCMC を実行
4. Bayes factor の値に基づいて仮説の優劣を判断する

制約付き系統推定と Bayes factor による系統仮説比較

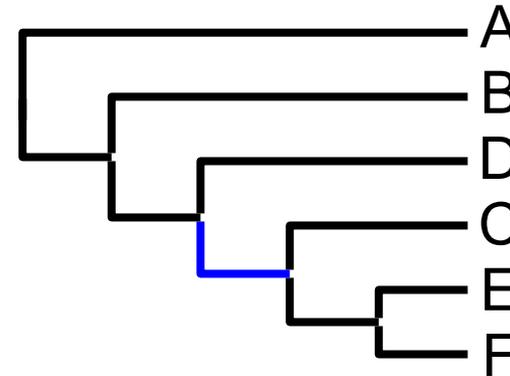


制約付き系統推定と Bayes factor による系統仮説比較

仮説 1 (M_1)



仮説 0 (M_0)

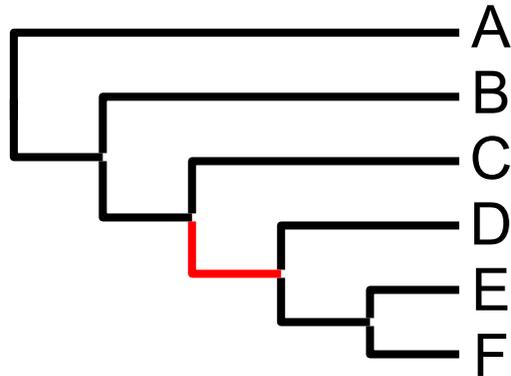


$$B_{10} = \frac{Pr(D|M_1)}{Pr(D|M_0)}$$

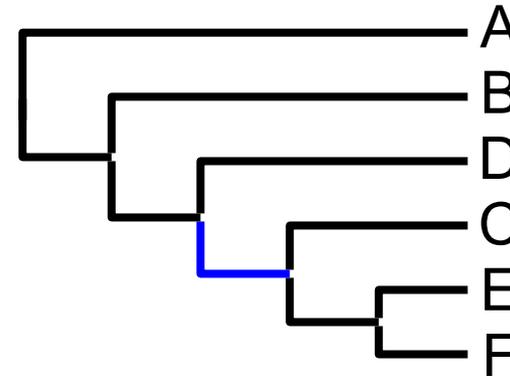
- 仮説 1 の仮説 0 に対する Bayes factor (B_{10}) は左の式で算出

制約付き系統推定と Bayes factor による系統仮説比較

仮説 1 (M_1)



仮説 0 (M_0)



$$B_{10} = \frac{Pr(D|M_1)}{Pr(D|M_0)}$$

B_{10}	仮説1が
3~20	より優れている
20~150	強く支持されている
150~	非常に強く支持されている

- 仮説 1 の仮説 0 に対する Bayes factor (B_{10}) は左の式で算出
- この値を左下表に基づいて判断
- $Pr(D|M)$ の値は MCMC サンプルの対数尤度の調和平均から求める

最尤系統推定に劣る点

最尤系統推定に劣る点

- M-H 法の各種パラメータ設定に職人技を要する

最尤系統推定に劣る点

- M-H 法の各種パラメータ設定に職人技を要する
- 樹形の信頼性を過大評価してしまう

最尤系統推定に劣る点

- M-H 法の各種パラメータ設定に職人技を要する
- 樹形の信頼性を過大評価してしまう
 - 複数パラメータを変化させて樹形の提案を行うため、受理確率が著しく低下する

最尤系統推定に劣る点

- M-H 法の各種パラメータ設定に職人技を要する
- 樹形の信頼性を過大評価してしまう
 - 複数パラメータを変化させて樹形の提案を行うため、受理確率が著しく低下する
 - 枝長の事前分布による問題
 - ダメな系統樹は著しく短い枝を持つことが多く、事前分布が適していない

最尤系統推定に劣る点

- M-H 法の各種パラメータ設定に職人技を要する
- 樹形の信頼性を過大評価してしまう
 - 複数パラメータを変化させて樹形の提案を行うため、受理確率が著しく低下する
 - 枝長の事前分布による問題
 - ダメな系統樹は著しく短い枝を持つことが多く、事前分布が適していない
- 系統仮説の多重比較時に補正方法が無い
 - 最尤法では SH 検定・ AU 検定という方法がある

最尤系統推定に劣る点

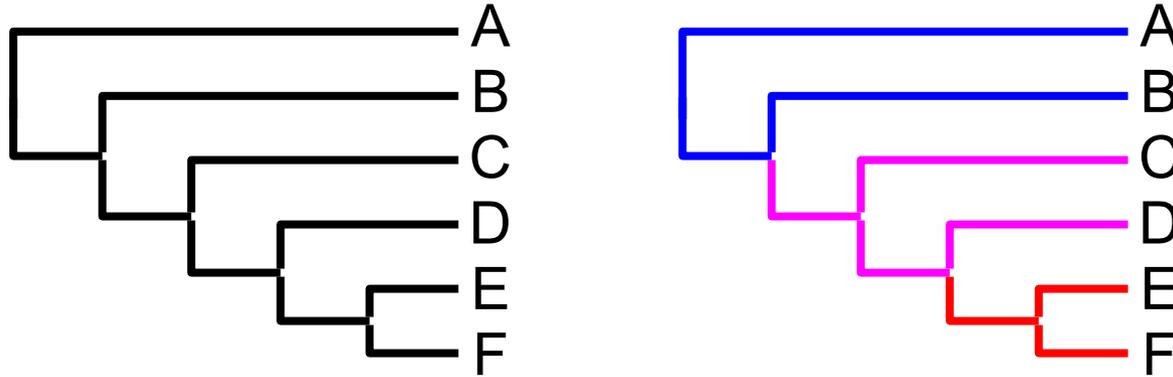
- M-H 法の各種パラメータ設定に職人技を要する
- 樹形の信頼性を過大評価してしまう
 - 複数パラメータを変化させて樹形の提案を行うため、受理確率が著しく低下する
 - 枝長の事前分布による問題
 - ダメな系統樹は著しく短い枝を持つことが多く、事前分布が適していない
- 系統仮説の多重比較時に補正方法が無い
 - 最尤法では SH 検定・ AU 検定という方法がある
- 核心部が逐次処理のため並列化に不向き

最尤系統推定に優る点

最尤系統推定に優る点

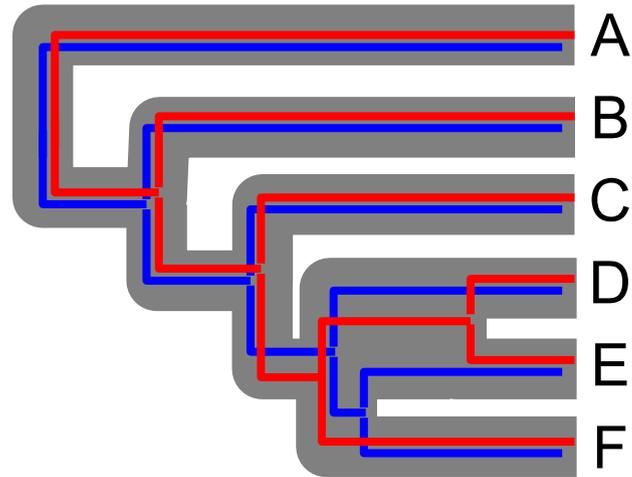
- 柔軟なモデリングが可能

最尤系統推定に優る点



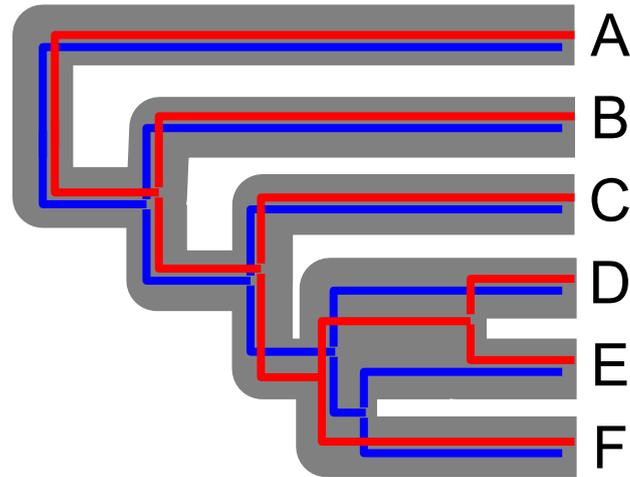
- 柔軟なモデリングが可能
 - 現在のほとんどの方法は、系統樹上の枝間で置換モデルは共通と仮定しているが、MCMC なら仮定しないモデルも計算しやすい

最尤系統推定に優る点



- 柔軟なモデリングが可能
 - 現在のほとんどの方法は、系統樹上の枝間で置換モデルは共通と仮定しているが、MCMC なら仮定しないモデルも計算しやすい
 - 現在のほとんどの方法は、全ての形質が共通の系統樹に由来すると仮定しているが、MCMC なら以下略

最尤系統推定に優る点



- 柔軟なモデリングが可能
 - 現在のほとんどの方法は、系統樹上の枝間で置換モデルは共通と仮定しているが、MCMC なら仮定しないモデルも計算しやすい
 - 現在のほとんどの方法は、全ての形質が共通の系統樹に由来すると仮定しているが、MCMC なら以下略
- 同時推定化による解析ステップの統合が可能
 - 不確実性の考慮の漏れを減らし、信頼性の過大評価・信頼区間の過小評価を抑制できる

現在の系統樹推定の手順と問題点

現在の系統樹推定の手順と問題点

1. データ配列の整列
2. データ配列の区分け
3. 区分けされた領域ごとに最適置換モデルを選択
4. 選択された置換モデルを当てはめての最適樹形を選択
5. ブートストラップ法を用いた樹形の信頼性の推定
6. 系統樹を用いた解析
 - a. 分岐年代推定
 - b. 祖先形質復元
 - c. 系統的独立比較
 - d. 系統的多様性の算出

現在の系統樹推定の手順と問題点

1. データ配列の整列 ← 仮の系統樹に固定して行う
2. データ配列の区分け
3. 区分けされた領域ごとに最適置換モデルを選択
4. 選択された置換モデルを当てはめての最適樹形の選択
5. ブートストラップ法を用いた樹形の信頼性の推定
6. 系統樹を用いた解析
 - a. 分岐年代推定
 - b. 祖先形質復元
 - c. 系統的独立比較
 - d. 系統的多様性の算出

現在の系統樹推定の手順と問題点

1. データ配列の整列 ← 仮の系統樹に固定して行う

2. データ配列の区分け

3. 区分けされた領域ごとに最適置換モデルを選択

4. 選択された置換モデルを当てはめての最適樹形の選択

5. ブートストラップ法を用いた樹形の信頼性の推定

6. 系統樹を用いた解析

a. 分岐年代推定

b. 祖先形質復元

c. 系統的独立比較

d. 系統的多様性の算出

整列・区分状態と置換モデルは固定され、
その不確実性は考慮されない

現在の系統樹推定の手順と問題点

1. データ配列の整列 ← 仮の系統樹に固定して行う

2. データ配列の区分け

3. 区分けされた領域ごとに最適置換モデルを選択

4. 選択された置換モデルを当てはめての最適樹形の選択

5. ブートストラップ法を用いた樹形の信頼性の推定

6. 系統樹を用いた解析

a. 分岐年代推定

b. 祖先形質復元

c. 系統的独立比較

d. 系統的多様性の算出

整列・区分状態と置換モデルは固定され、
その不確実性は考慮されない

樹形の不確実性は多くの場合
一切考慮されない

ことが多い

ベイジアン MCMC による同時推定の実現

ベイズアン MCMC による同時推定の実現

- 整列状態と系統樹 (ALIFRITZ, BEAST, BALi-Phy)

ベイジアン MCMC による同時推定の実現

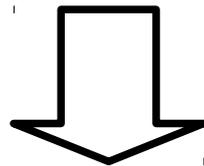
- 整列状態と系統樹 (ALIFRITZ, BEAST, BAli-Phy)
- 区分状態と系統樹 (BayesPhylogenies, PhyloBayes)

ベイジアン MCMC による同時推定の実現

- 整列状態と系統樹 (ALIFRITZ, BEAST, BAli-Phy)
- 区分状態と系統樹 (BayesPhylogenies, PhyloBayes)
- 置換モデルと系統樹 (MrBayes3.2)

ベイジアン MCMC による同時推定の実現

- 整列状態と系統樹 (ALIFRITZ, BEAST, BAli-Phy)
- 区分状態と系統樹 (BayesPhylogenies, PhyloBayes)
- 置換モデルと系統樹 (MrBayes3.2)



整列・区分・置換モデル選択・系統樹推定
を全て統合可能

まとめ

- 系統樹推定は多重推定・多重モデル選択である
- ベイジアン MCMC による系統樹推定では信頼性を過大評価しやすいが、柔軟なモデリングにより好ましくない制約を除去できる
- 同時推定化によって多重推定・多重モデル選択の問題を理論上は解決可能
- 実用的に利用できるプログラムのリリースが待ち望まれる