ベイズ統計学（2）

「マルコフ連鎖モンテカルロ」に登攀する

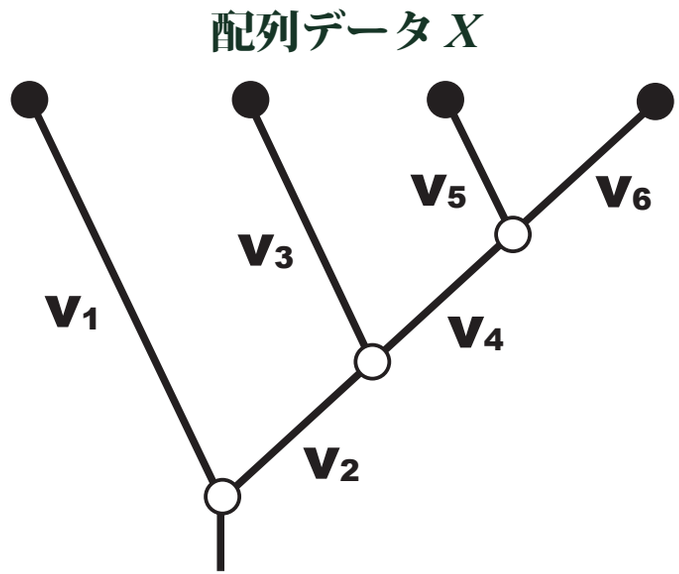
ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

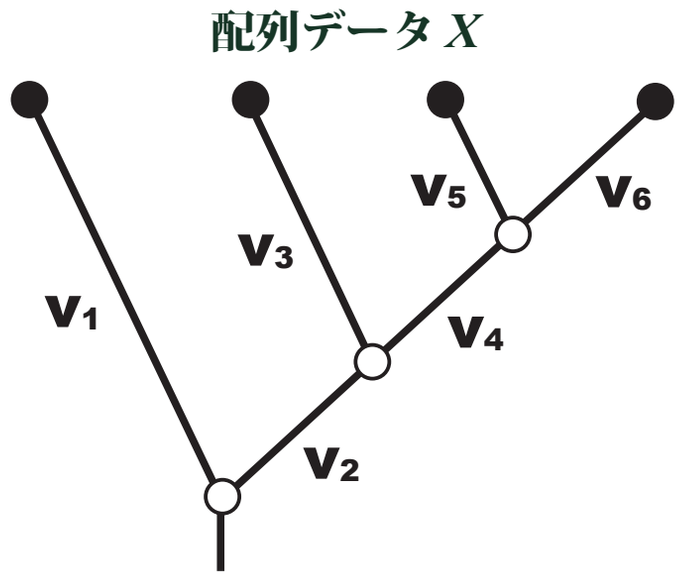
［例］分子系統樹のベイズ推定



ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

［例］分子系統樹のベイズ推定



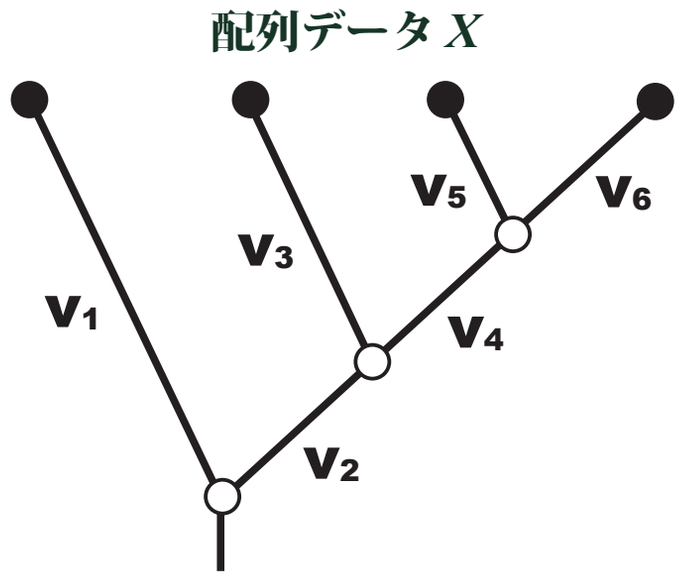
パラメータ群

- 樹形：τ
- 枝長：v
- 置換遷移確率：θ
- サイト間変異：α

ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

［例］分子系統樹のベイズ推定



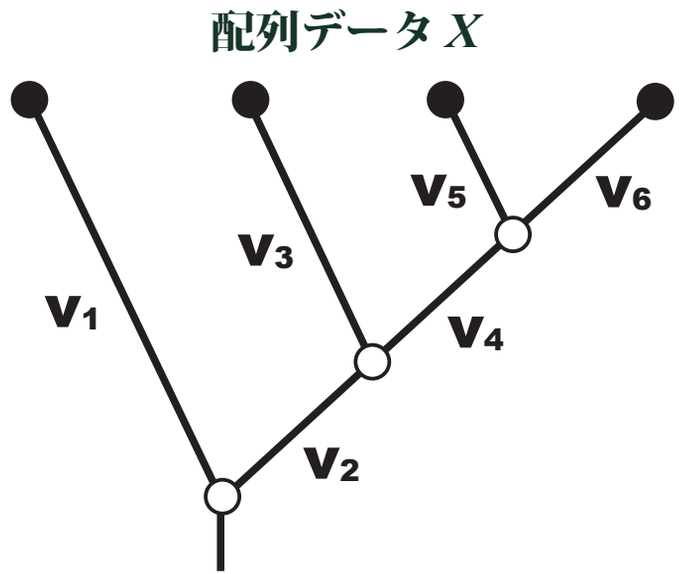
パラメータ群

- 樹形：τ → 推定対象
- 枝長：v
- 置換遷移確率：θ
- サイト間変異：α

ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

［例］分子系統樹のベイズ推定



パラメータ群

- 樹形：τ → 推定対象
- ~~枝長：v~~
- ~~置換遷移確率：θ~~
- ~~サイト間変異：α~~

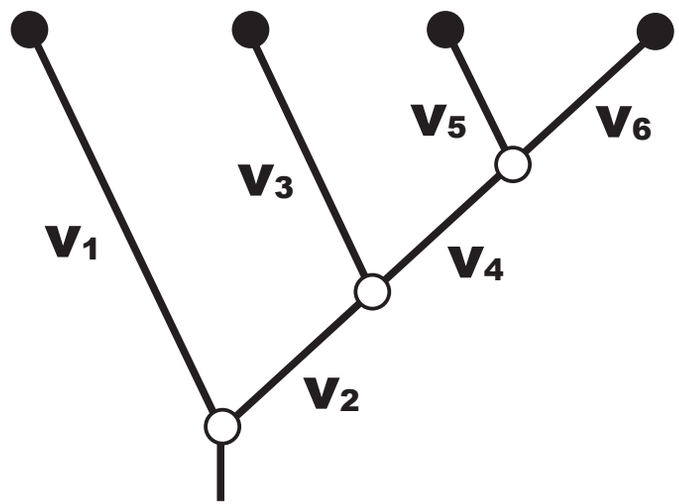
→ 対象外（撹乱母数）

ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

［例］分子系統樹のベイズ推定

配列データ X



パラメータ群

- 樹形：τ → 推定対象
- ~~枝長：v~~
- ~~置換遷移確率：θ~~
- ~~サイト間変異：α~~

→ 対象外（撹乱母数）

樹形の事後分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

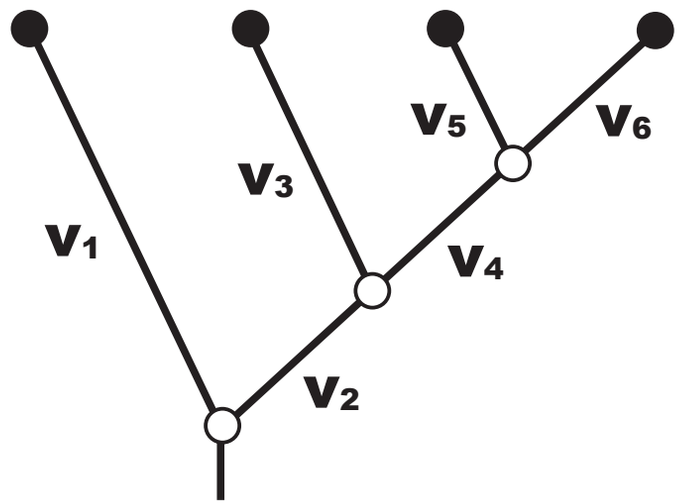
樹形の事前分布

ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

［例］分子系統樹のベイズ推定

配列データ X



パラメータ群

- 樹形：τ → 推定対象
- ~~枝長：v~~
- ~~置換遷移確率：$\boldsymbol{\theta}$~~
- ~~サイト間変異：$\boldsymbol{\alpha}$~~

→ 対象外（撹乱母数）

樹形の事後分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

樹形の事前分布

樹形の尤度

$$f(\boldsymbol{X}|\tau_i) = \int_{\boldsymbol{v}_i}\int_{\boldsymbol{\theta}}\int_{\alpha} f(\boldsymbol{X}|\tau_i, \boldsymbol{v}_i, \boldsymbol{\theta}, \alpha) f(\boldsymbol{v}_i) f(\boldsymbol{\theta}) f(\alpha) d\boldsymbol{v}_i d\boldsymbol{\theta} d\alpha$$

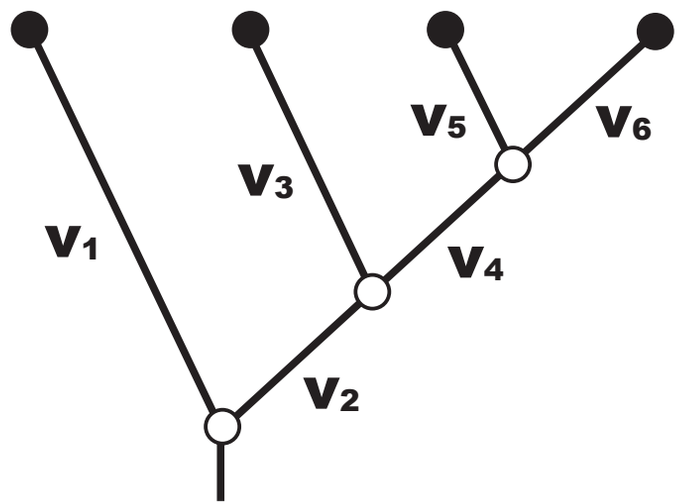
撹乱母数は期待値計算で消去

ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

［例］分子系統樹のベイズ推定

配列データ X



パラメータ群

- 樹形：τ → 推定対象
- ~~枝長：v~~
- ~~置換遷移確率：$\boldsymbol{\theta}$~~
- ~~サイト間変異：α~~

→ 対象外（撹乱母数）

樹形の事後分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

樹形の事前分布

樹形の尤度

$$f(\boldsymbol{X}|\tau_i) = \int_{\boldsymbol{v}_i}\int_{\boldsymbol{\theta}}\int_{\alpha} f(\boldsymbol{X}|\tau_i, \boldsymbol{v}_i, \boldsymbol{\theta}, \alpha) f(\boldsymbol{v}_i) f(\boldsymbol{\theta}) f(\alpha) d\boldsymbol{v}_i d\boldsymbol{\theta} d\alpha$$

撹乱母数は期待値計算で消去

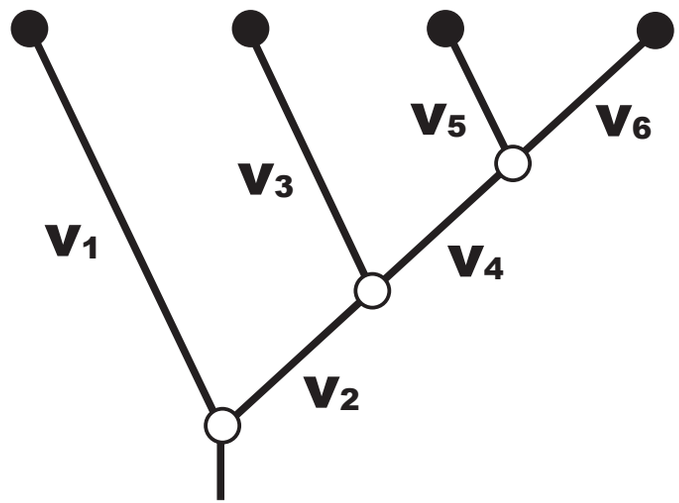
多重積分の重荷

ベイズ事後分布を求める

モデルが複雑になると計算量が膨大になる

［例］分子系統樹のベイズ推定

配列データ X



パラメータ群

- 樹形：τ → 推定対象
- ~~枝長：v~~
- ~~置換遷移確率：$\boldsymbol{\theta}$~~
- ~~サイト間変異：α~~

→ 対象外（撹乱母数）

樹形の事後分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

樹形の事前分布

組合せ論的爆発

樹形の尤度

$$f(\boldsymbol{X}|\tau_i) = \int_{\boldsymbol{v}_i}\int_{\boldsymbol{\theta}}\int_{\alpha} f(\boldsymbol{X}|\tau_i, \boldsymbol{v}_i, \boldsymbol{\theta}, \alpha) f(\boldsymbol{v}_i) f(\boldsymbol{\theta}) f(\alpha) d\boldsymbol{v}_i d\boldsymbol{\theta} d\alpha$$

撹乱母数は期待値計算で消去

多重積分の重荷

ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

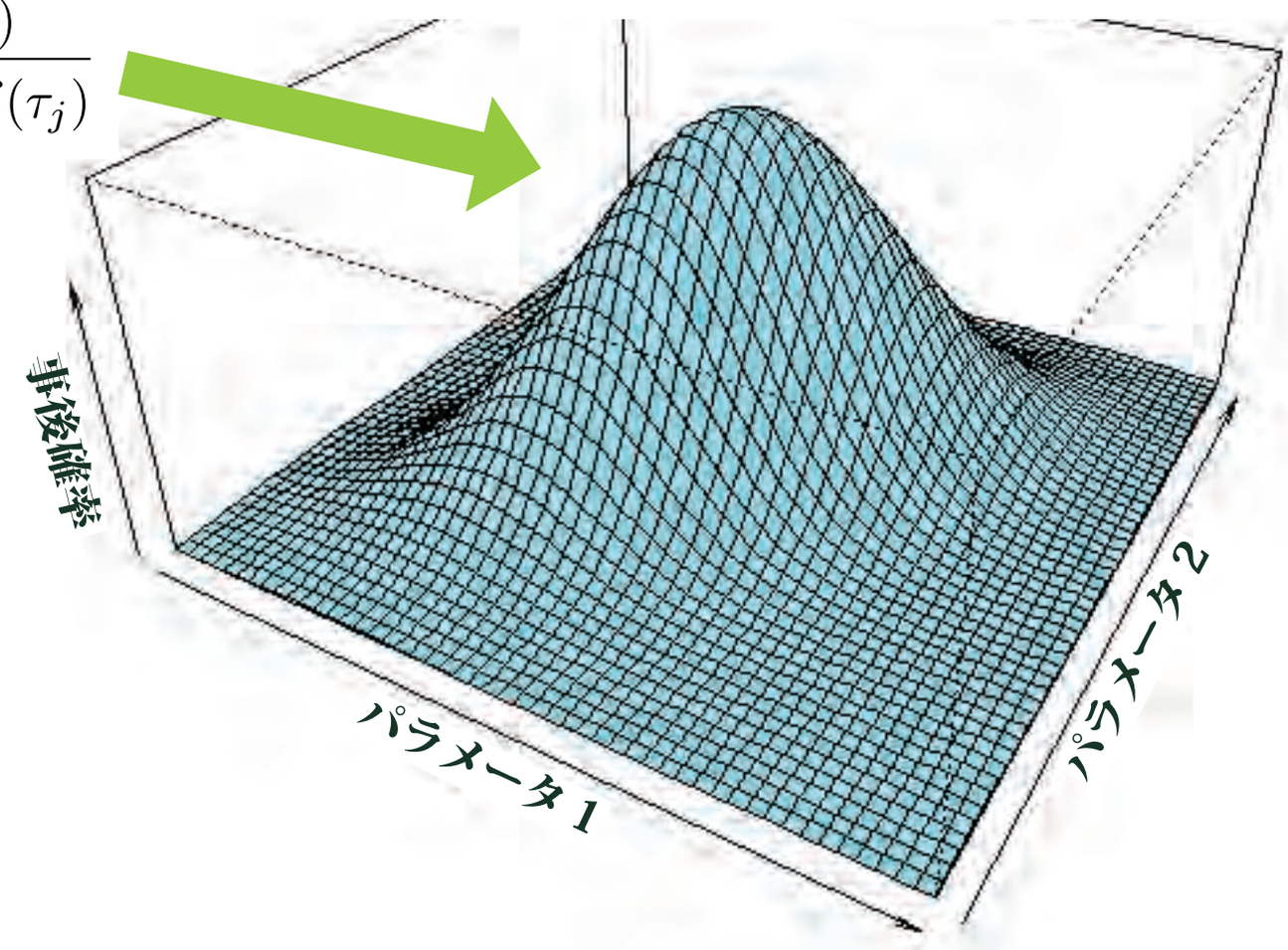
$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

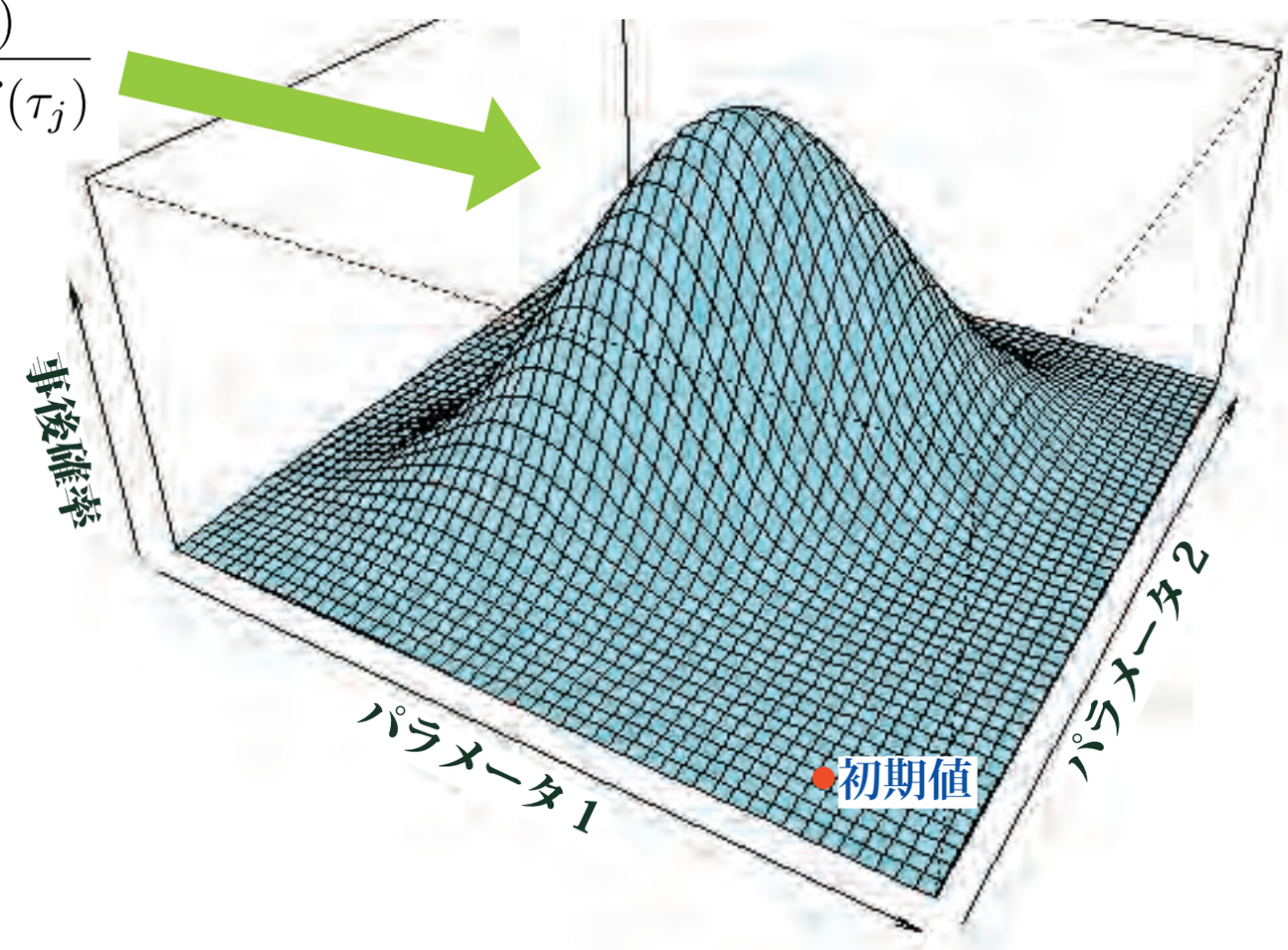


ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

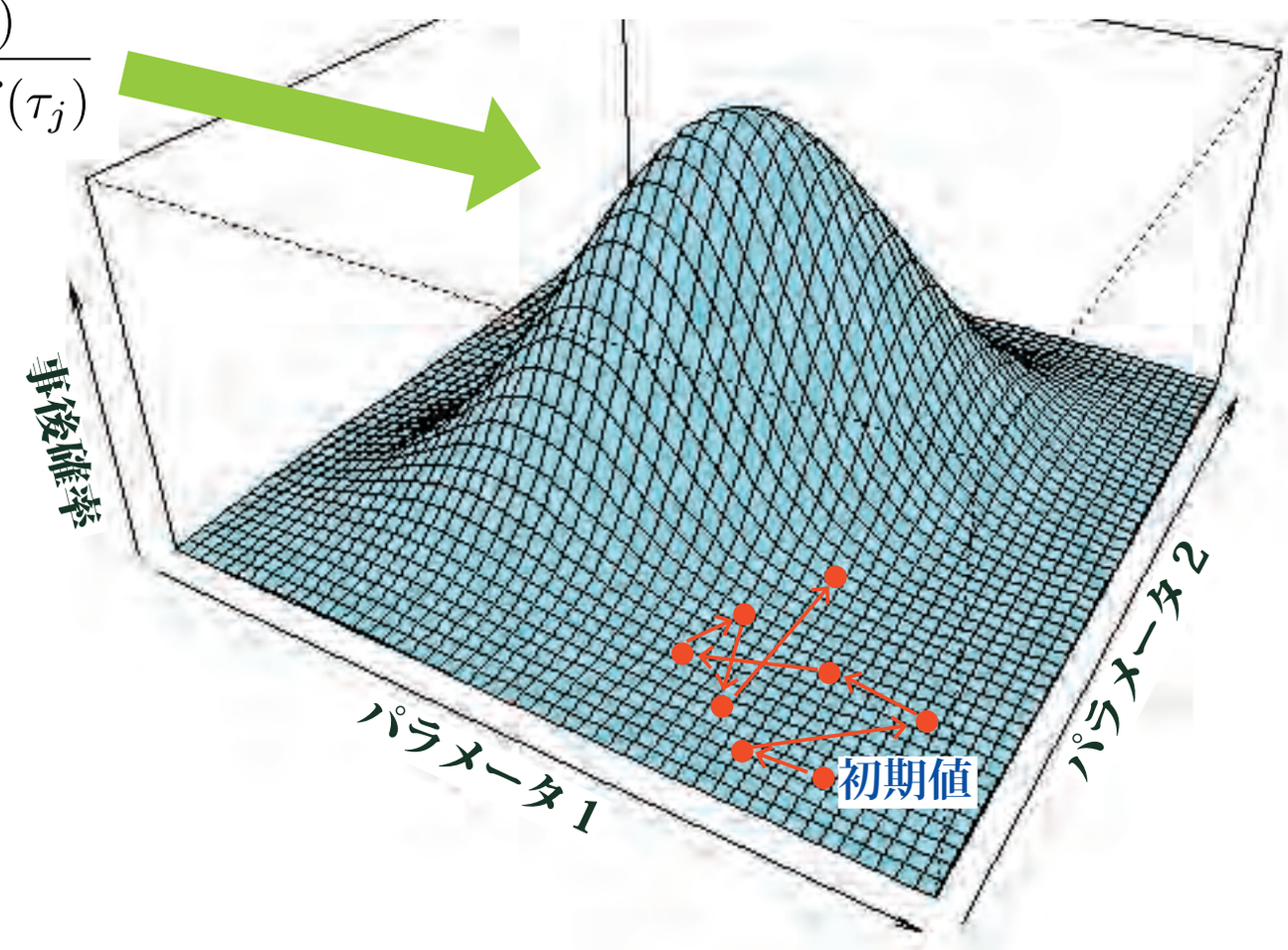


ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

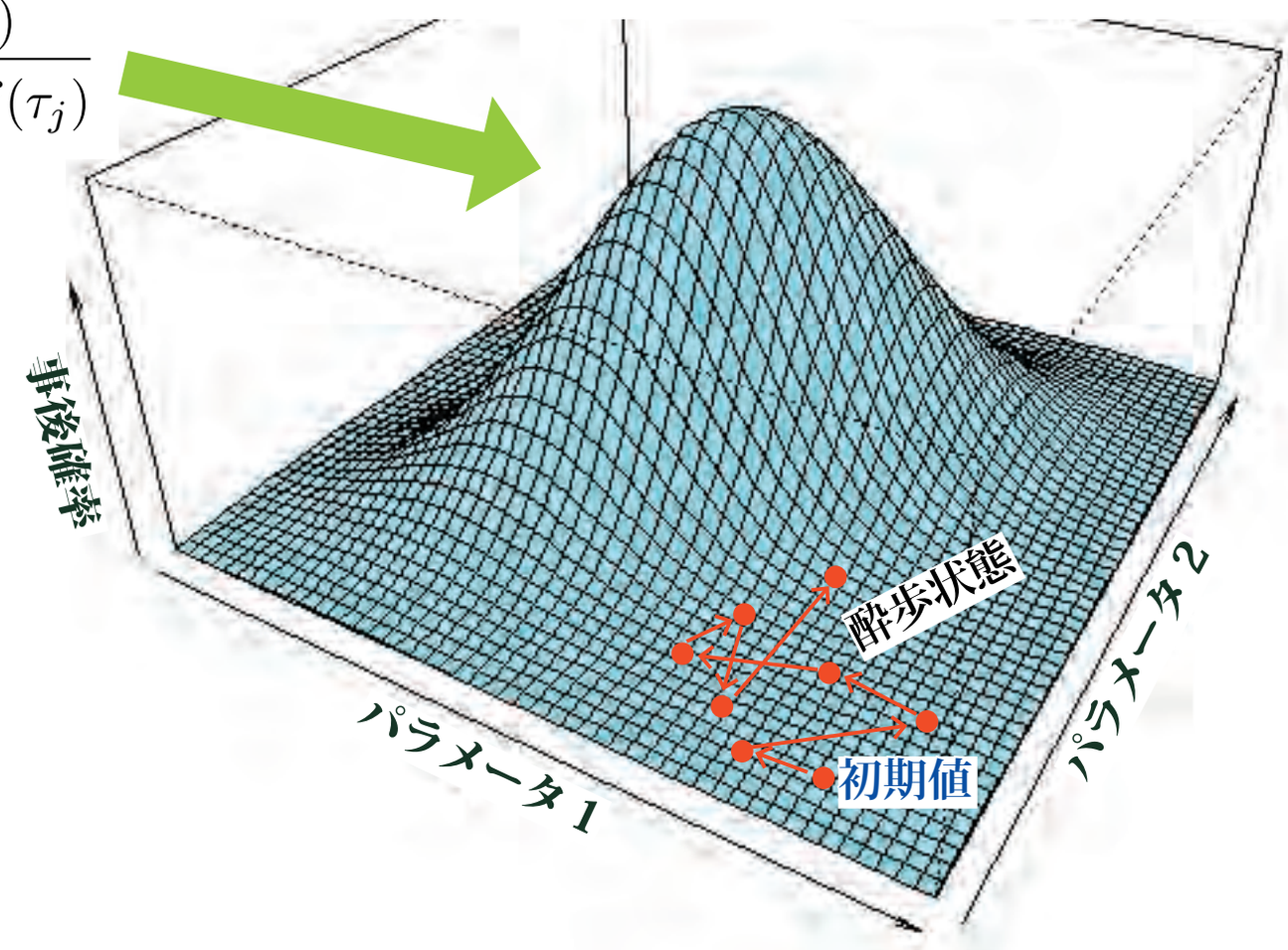


ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

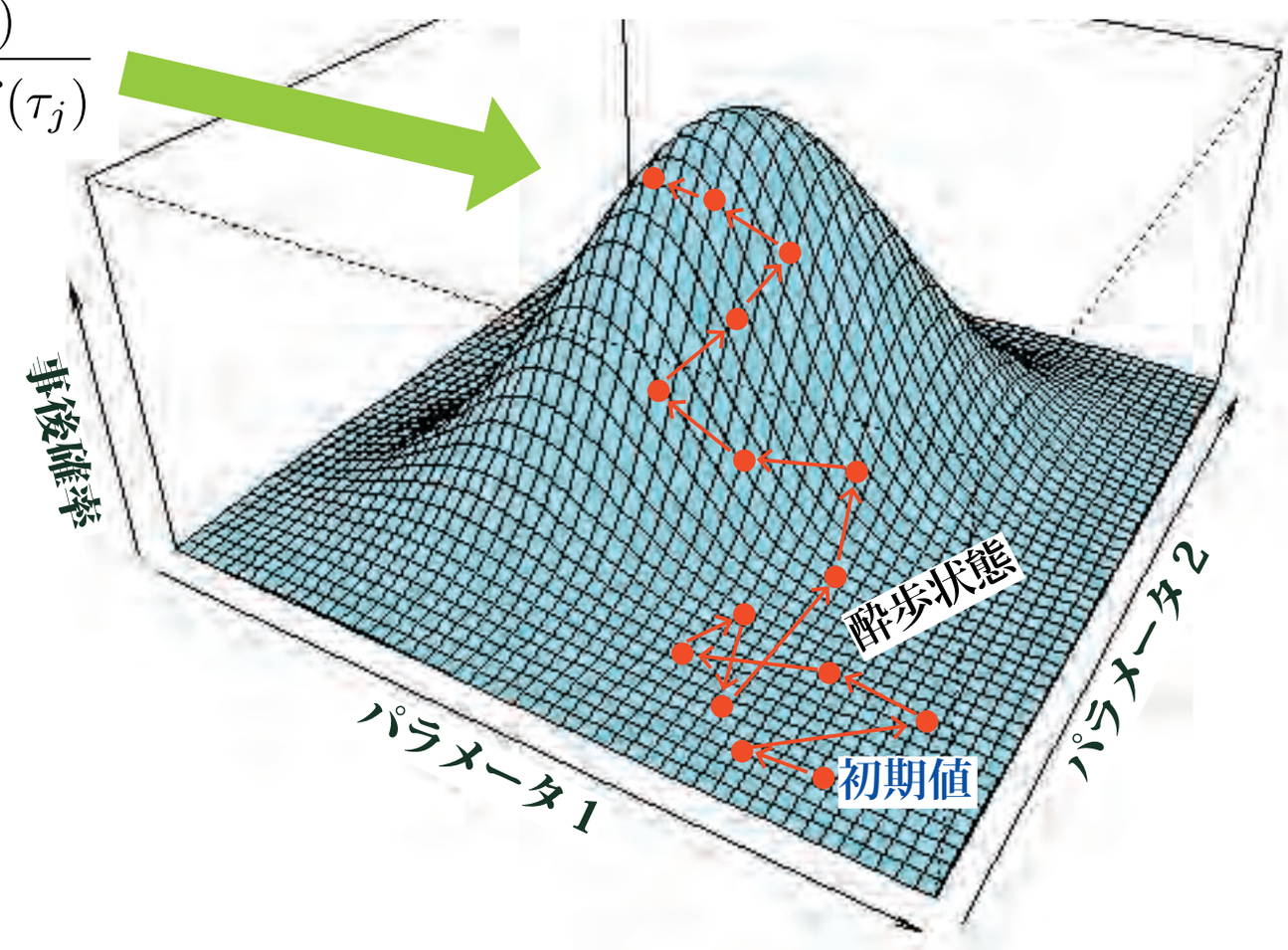


ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

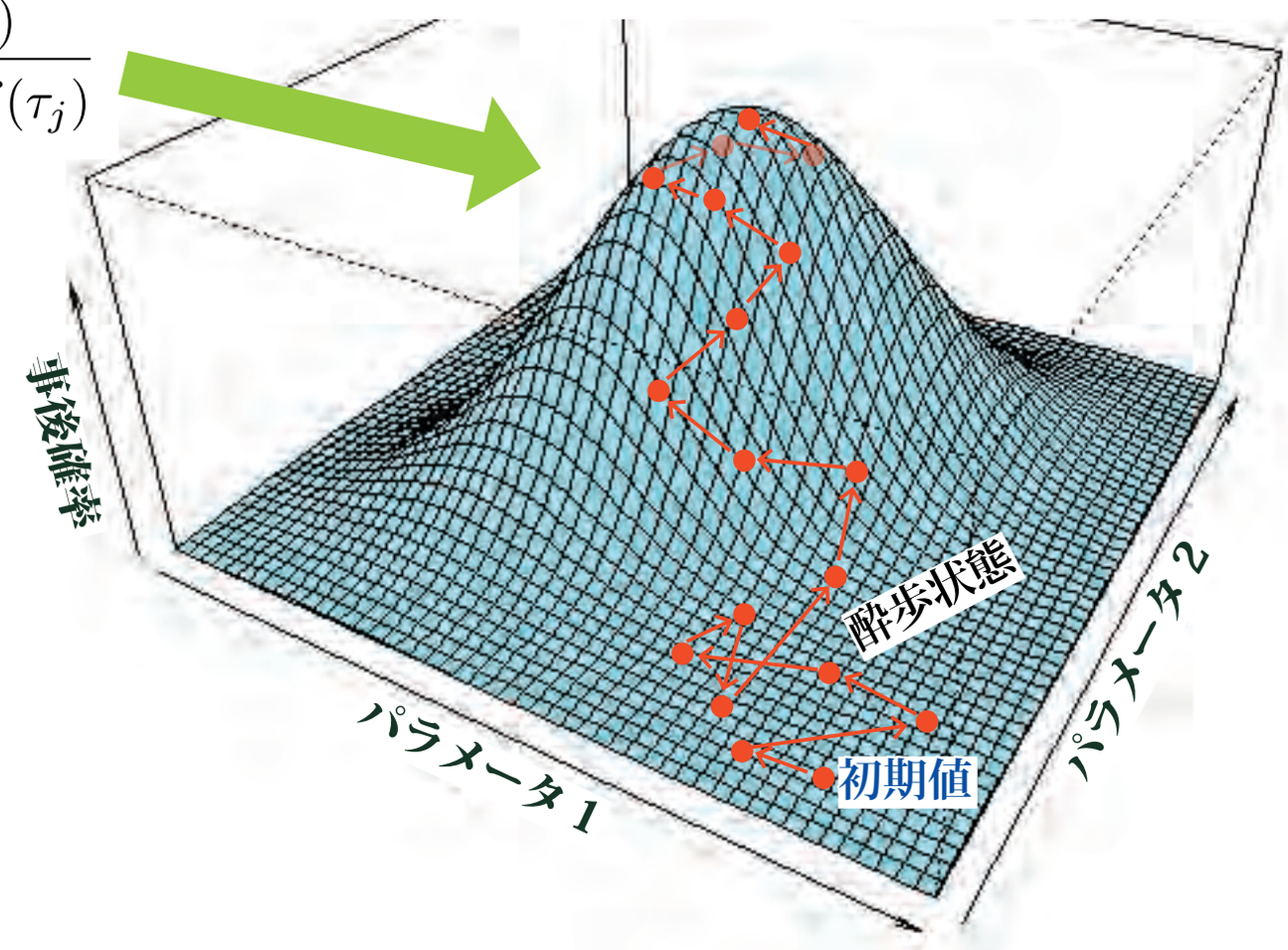


ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

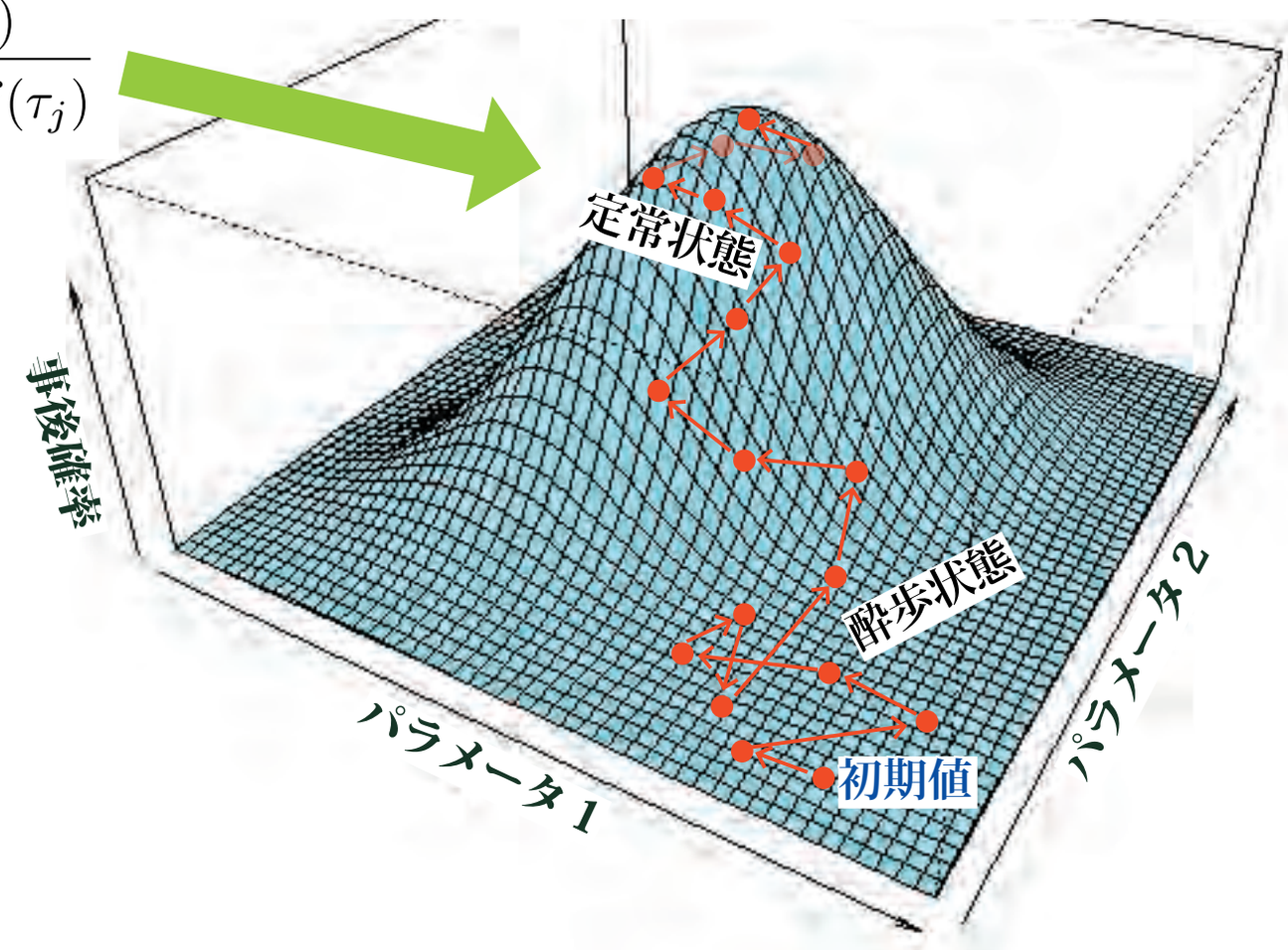


ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$



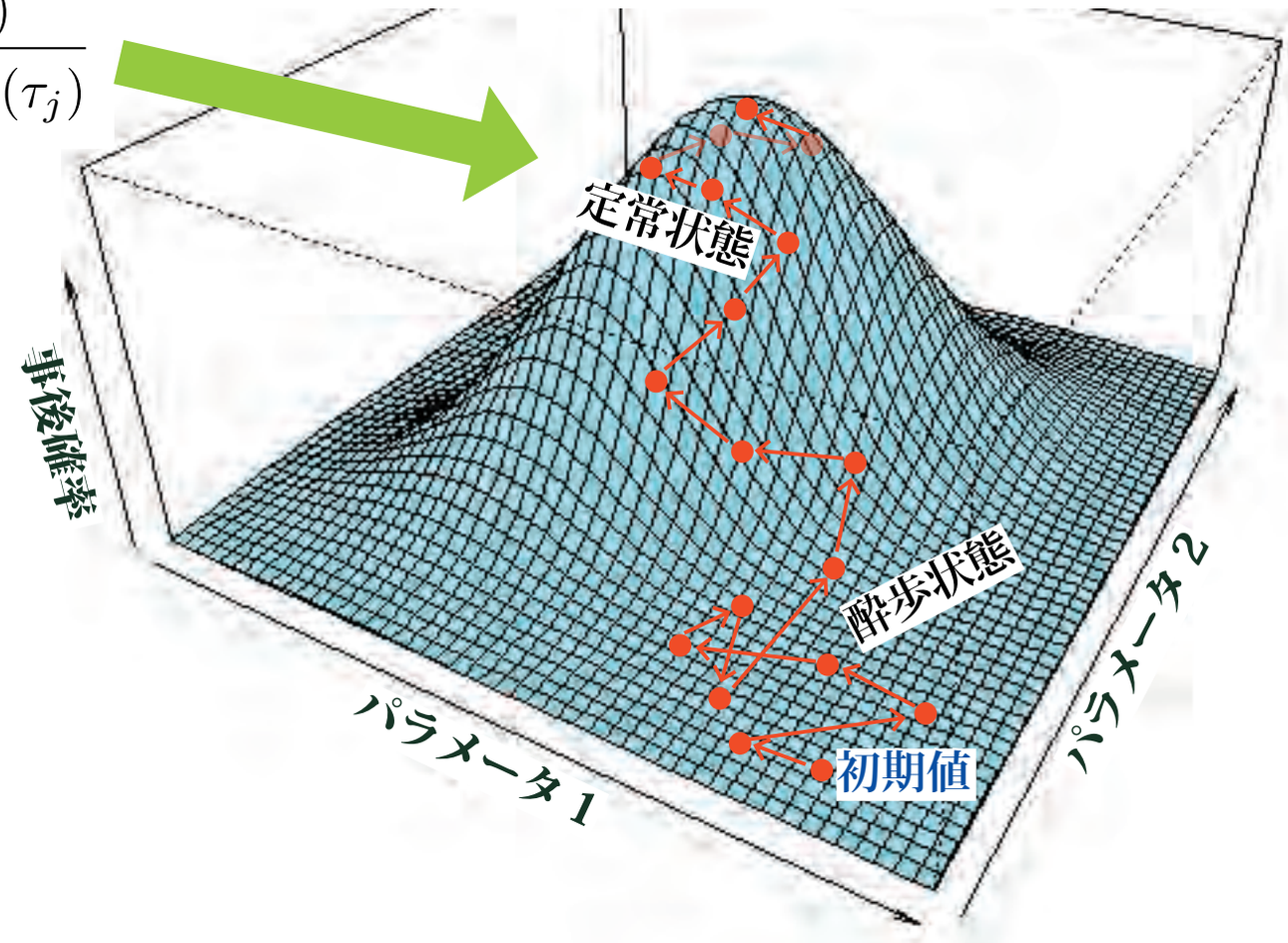
ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

パラメータ初期値を無作為に与え，マルコフ過程を用いて事後確率分布を系統的にサンプリングし，定常状態になるまで探索させる．



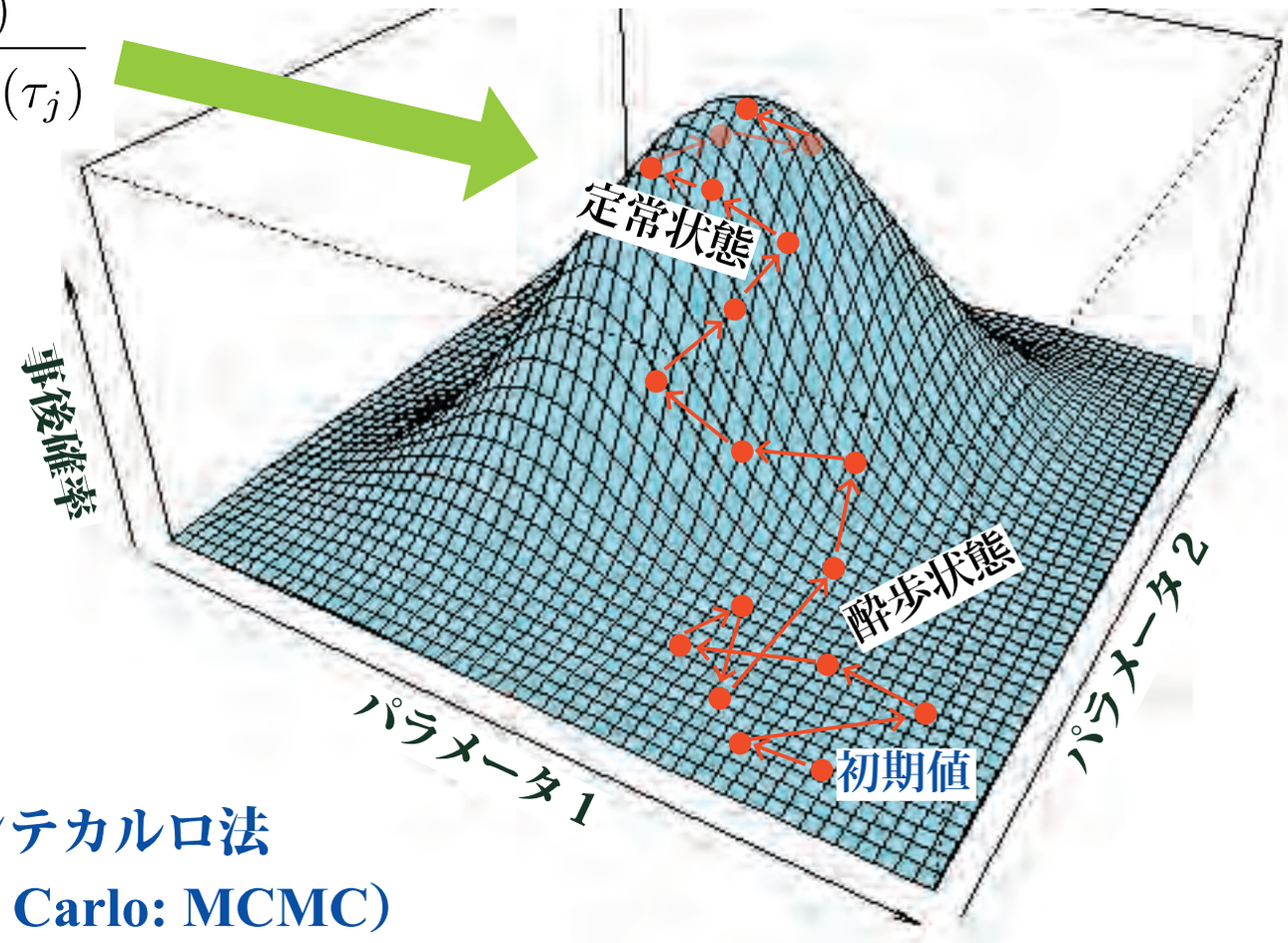
ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

パラメータ初期値を無作為に与え，マルコフ過程を用いて事後確率分布を系統的にサンプリングし，定常状態になるまで探索させる．



マルコフ連鎖モンテカルロ法
（Markov chain Monte Carlo: MCMC）

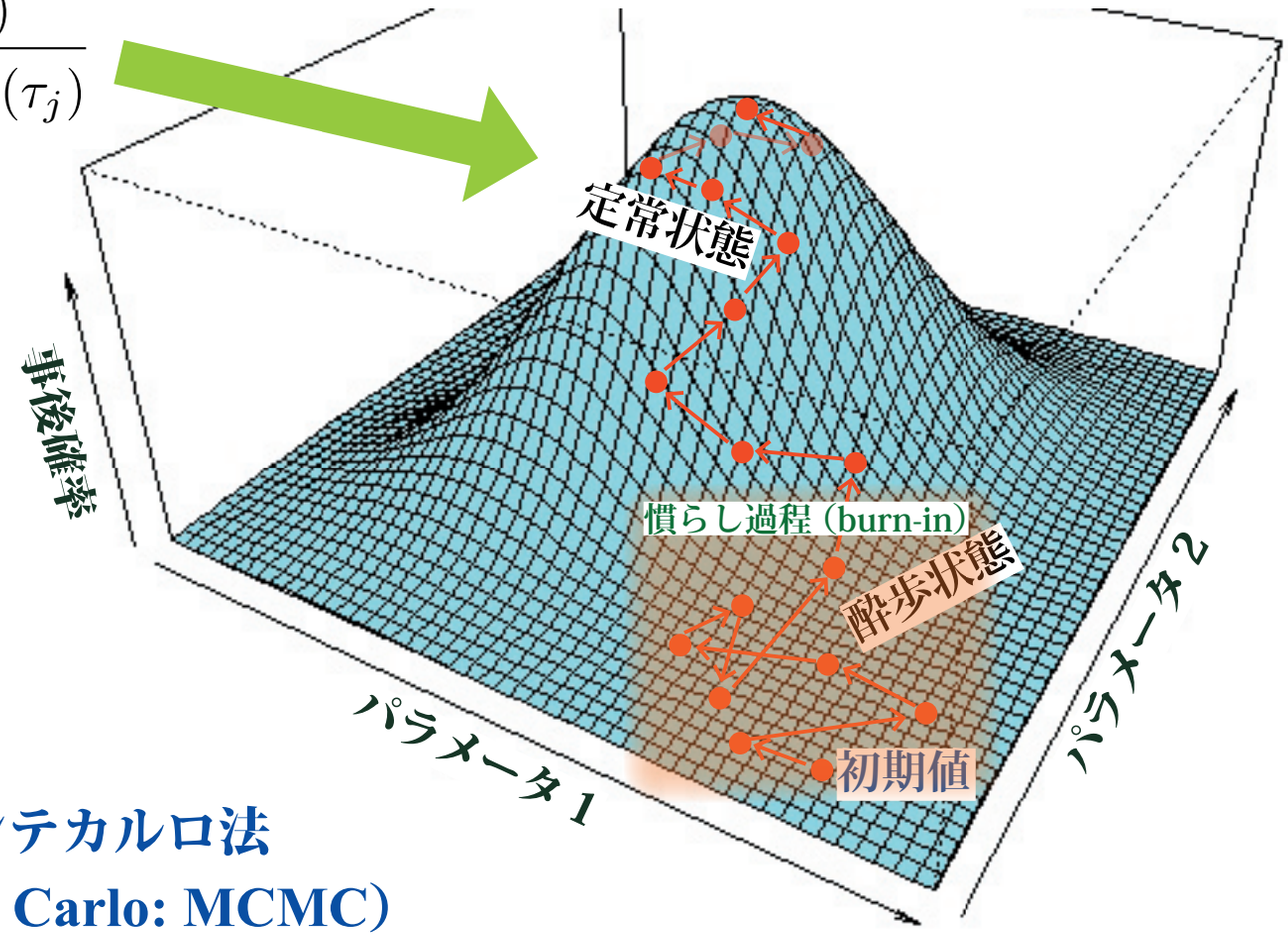
ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

パラメータ初期値を無作為に与え，マルコフ過程を用いて事後確率分布を系統的にサンプリングし，定常状態になるまで探索させる．



マルコフ連鎖モンテカルロ法
（Markov chain Monte Carlo: MCMC）

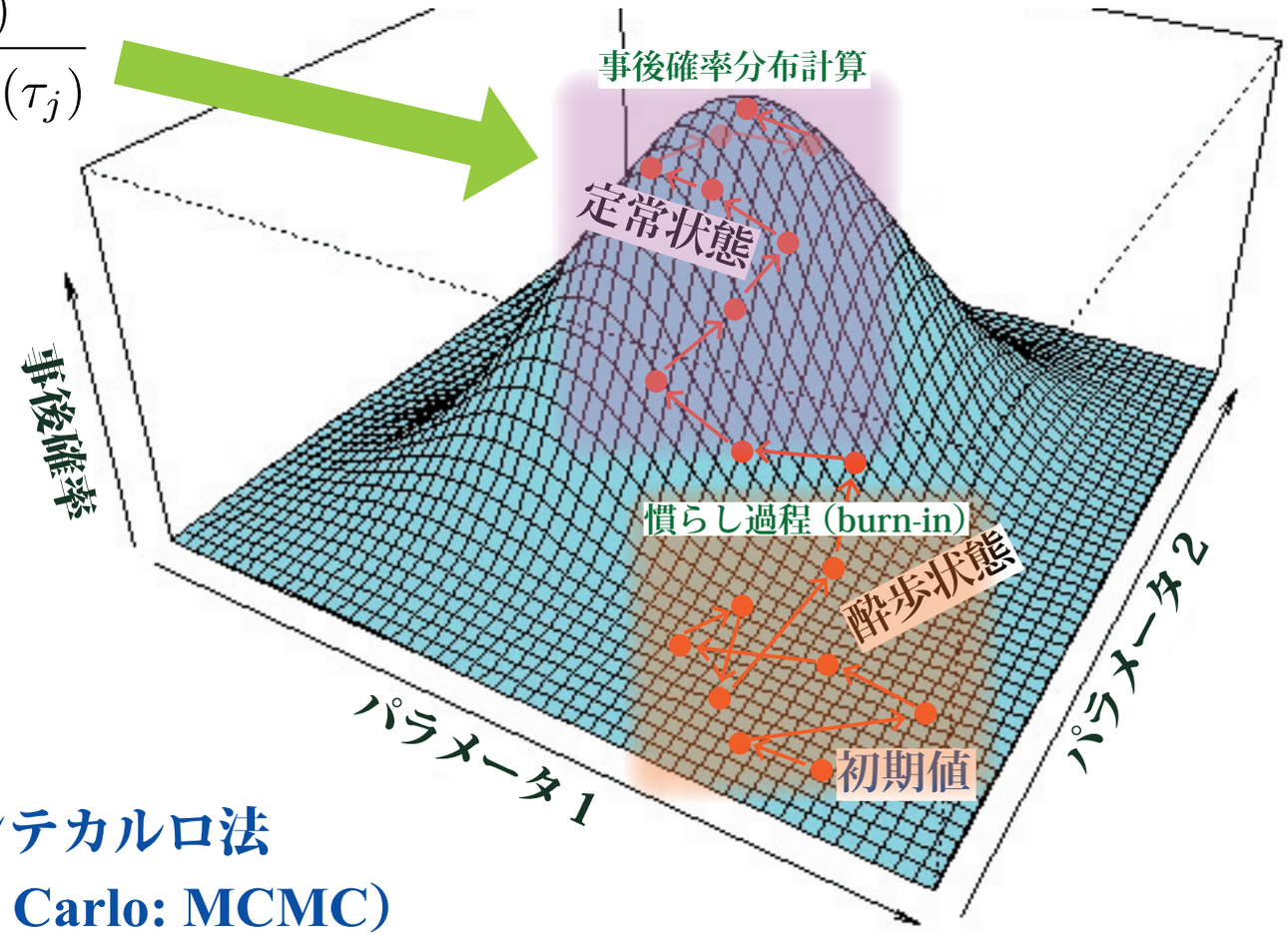
ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

事後確率分布

$$f(\tau_i|\boldsymbol{X}) = \frac{f(\boldsymbol{X}|\tau_i)f(\tau_i)}{\sum_{j=1}^{N} f(\boldsymbol{X}|\tau_j)f(\tau_j)}$$

パラメータ初期値を無作為に与え，マルコフ過程を用いて事後確率分布を系統的にサンプリングし，定常状態になるまで探索させる．



マルコフ連鎖モンテカルロ法
（Markov chain Monte Carlo: MCMC）

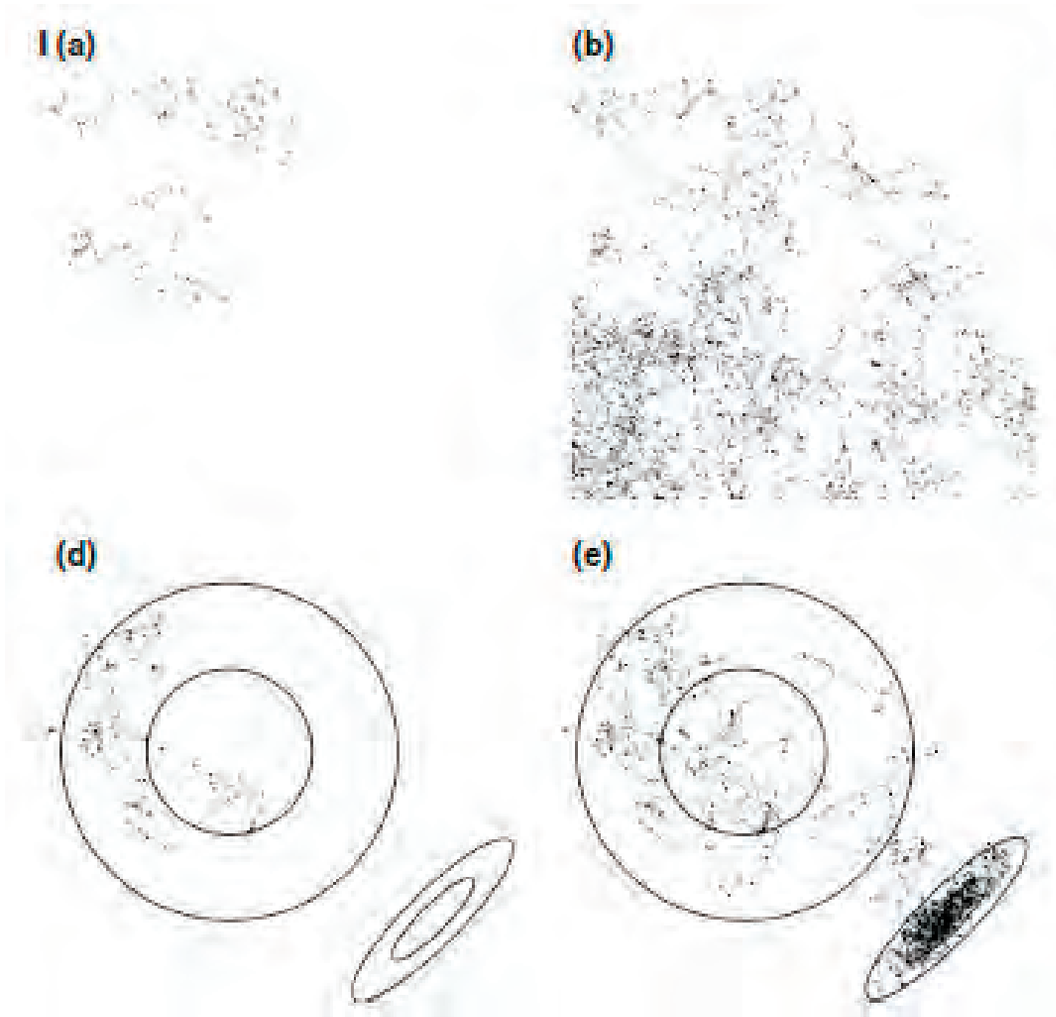
ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

100 ステップ

ベイズ事後分布を求める

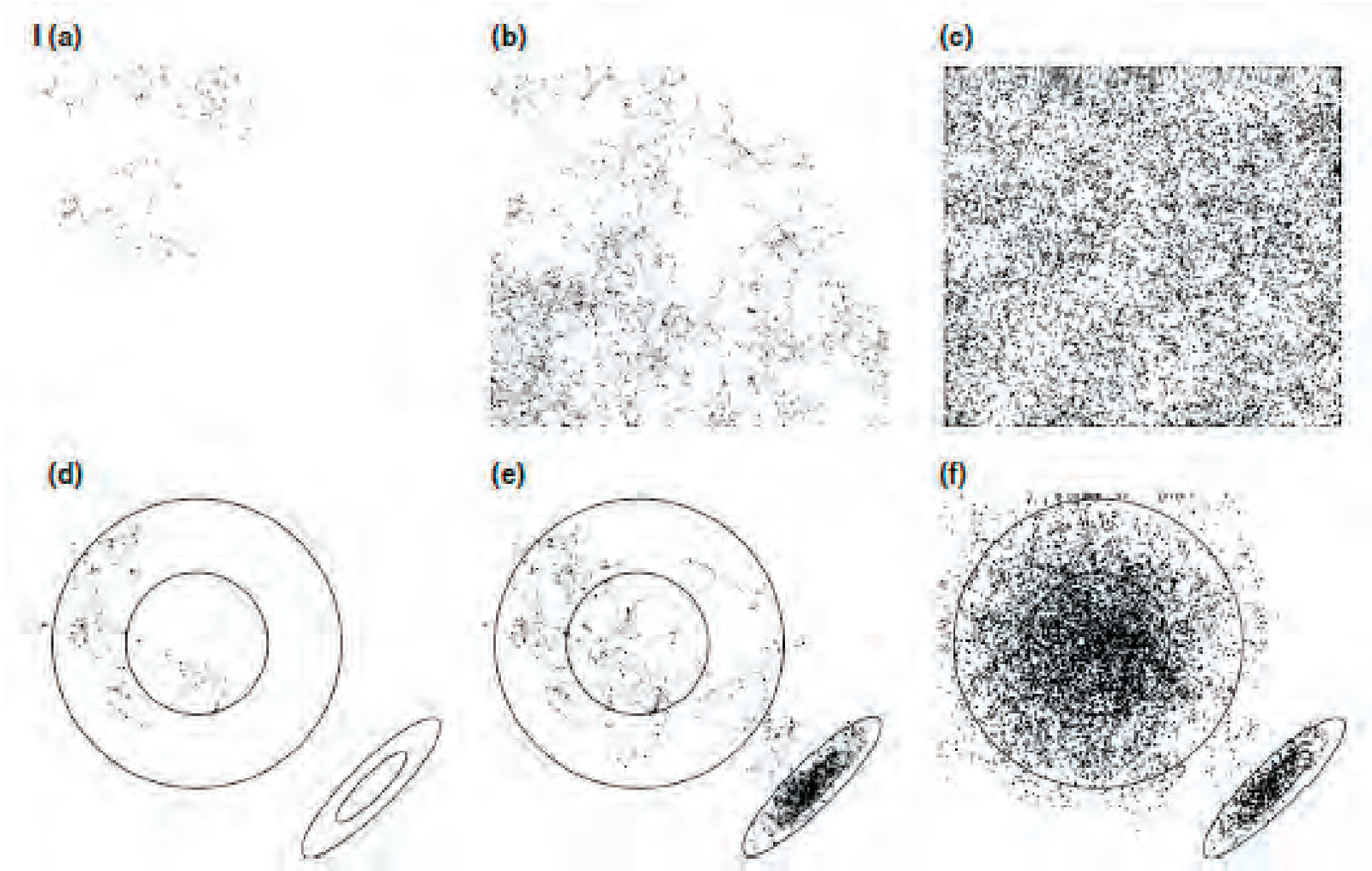
マルコフ連鎖モンテカルロ法（MCMC）という最終兵器



100 ステップ　　1,000 ステップ

ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器



100 ステップ　　1,000 ステップ　　10,000 ステップ

ベイズ事後分布を求める

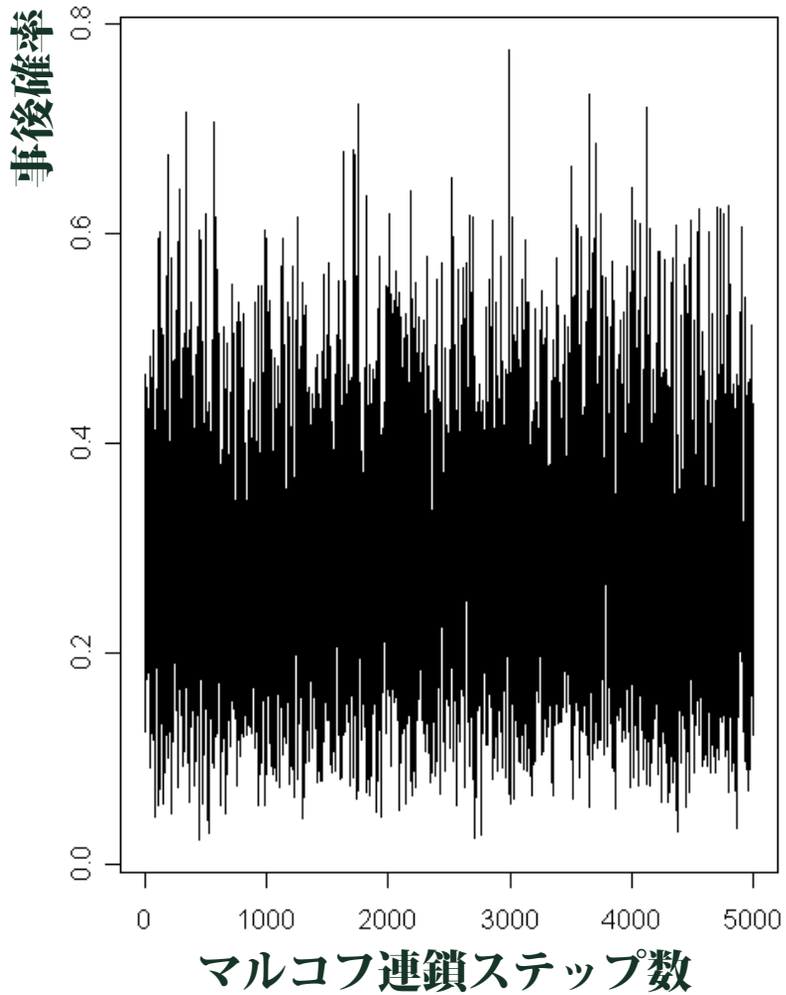
マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

［例］ベータ事前分布をもつ二項分布パラメータの事後分布

ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

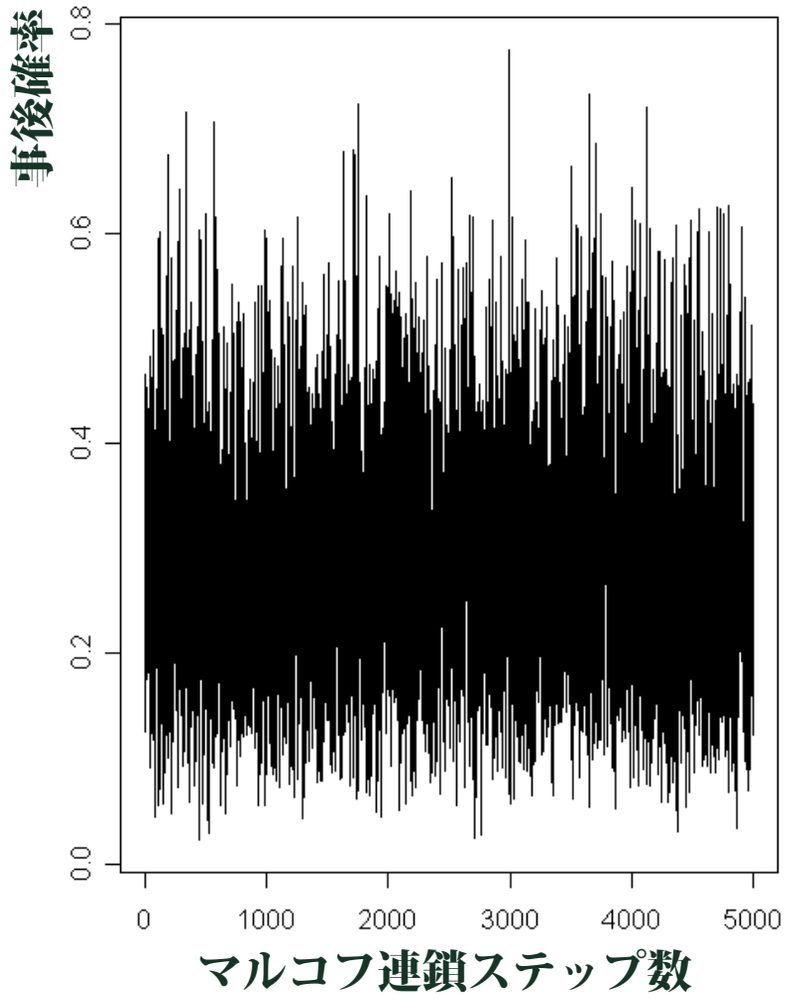
［例］ベータ事前分布をもつ二項分布パラメータの事後分布

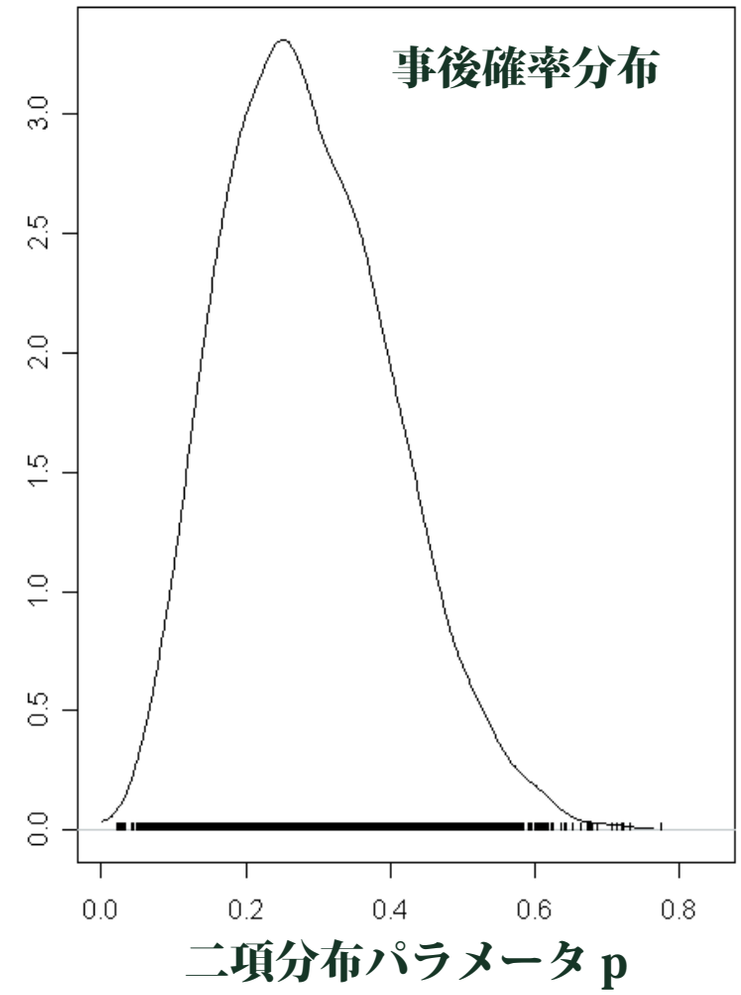


ベイズ事後分布を求める

マルコフ連鎖モンテカルロ法（MCMC）という最終兵器

［例］ベータ事前分布をもつ二項分布パラメータの事後分布





MCMC は福音か，それとも災厄か

ベイジアン最終兵器の光と影

MCMC は福音か，それとも災厄か

ベイジアン最終兵器の光と影

【光】「ベイジアン MCMC」が登場して，まだ 10 年ほどしか経っていない．にもかかわらず，その実用性はすでに生物学・医学の広い領域に浸透しつつある．生物統計学はもちろん，これから進化学・生態学・バイオインフォマティクスなどを学ぶとき，たいへん有効なツールである．とくに，現実に即した複雑なモデルを用いるとき，ベイジアン MCMC は現時点でほとんど唯一の計算手段である．

MCMC は福音か，それとも災厄か

ベイジアン最終兵器の光と影

【光】「ベイジアン MCMC」が登場して，まだ 10 年ほどしか経っていない．にもかかわらず，その実用性はすでに生物学・医学の広い領域に浸透しつつある．生物統計学はもちろん，これから進化学・生態学・バイオインフォマティクスなどを学ぶとき，たいへん有効なツールである．とくに，現実に即した複雑なモデルを用いるとき，ベイジアン MCMC は現時点でほとんど唯一の計算手段である．

【影】以下の問題点が指摘される：1）事前分布をどのように設定するのか，その妥当性あるいは結果への影響はどのようにして評価されるのか；2）マルコフ連鎖が収束したかどうかはいつどのようにして判定すればいいのか；3）ベイズ主義をめぐる“哲学的”な問題点は何一つ解決していない．ユーザーは眼前の問題解決に役立つ分析ツールのひとつとして使っているだけだから．