データ解析のための統計モデリング入門

第8章
マルコフ連鎖モンテカルロ法とベイズ統計モデル

outline

- **Introduction：MCMCとは**
- **8.1 例題：種子の生存確率（個体差なし）**
- **8.2 ふらふら試行錯誤による最尤推定**
- **8.3 MCMCアルゴリズムのひとつ：メトロポリス法**
  - 8.3.1 メトロポリス法でサンプリングしてみる
  - 8.3.2 マルコフ連鎖の定常分布
  - 8.3.3 この定常分布は何をあらわす分布なのか
- **8.4 MCMCサンプリングとベイズ統計モデル**
- **8.5 補足説明**
  - 8.5.1 メトロポリス法と定常分布の関係
  - 8.5.2 ベイズの定理
- **ロジスティック回帰をMCMCでやってみる**

■マルコフ連鎖モンテカルロ法（MCMC）

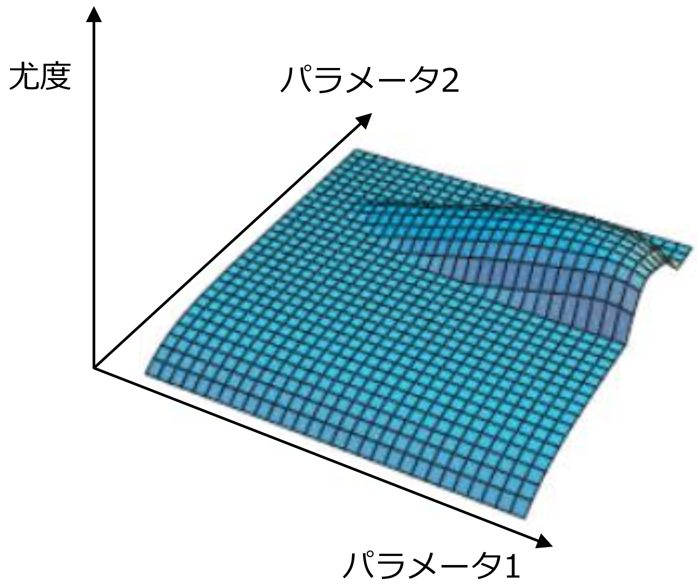
・**マルコフ連鎖**　：次の状態が現在の状態のみに依存するような確率過程
・**モンテカルロ法**：乱数を用いてシミュレーション等を行なう手法の総称

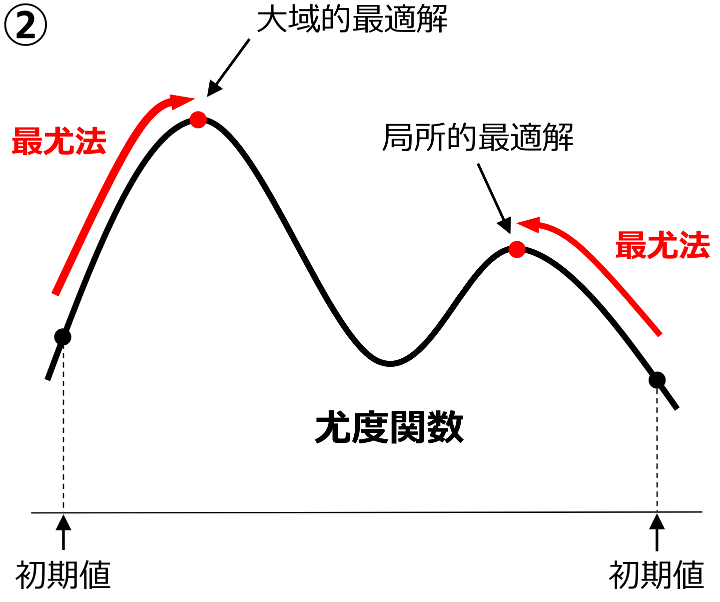
➘ ざっくりいえばMCMCとは

乱数を用いてマルコフ連鎖を生成し，パラメータの確率分布を推定する手法

【最尤法に対するMCMCの利点】

① 高次元のパラメータ空間を効率的に探索可能
② 尤度関数が多峰型になっていても，局所的な最適解にとらわれることがない

①


②


8.1 例題：種子の生存確率（個体差なし）

■ 各植物個体（全20個体）について，8個の種子の生死を調べたとする

- 個体 i の生存種子数：$\{y_1, y_2, \cdots, y_{20}\} = \{4,3,4,5,5,2,3,1,4,0,1,5,5,6,5,4,4,5,3,4\}$

  ・データの標本分散＝2.56

  ・$n=8, q=0.45$ の二項分布の分散＝1.98

- 特に過分散ではないようなので，生存種子数 y_i が**二項分布**に従うと仮定する

  ・データが離散値
  ・ゼロ以上で有限の範囲
  ・分散は平均の関数

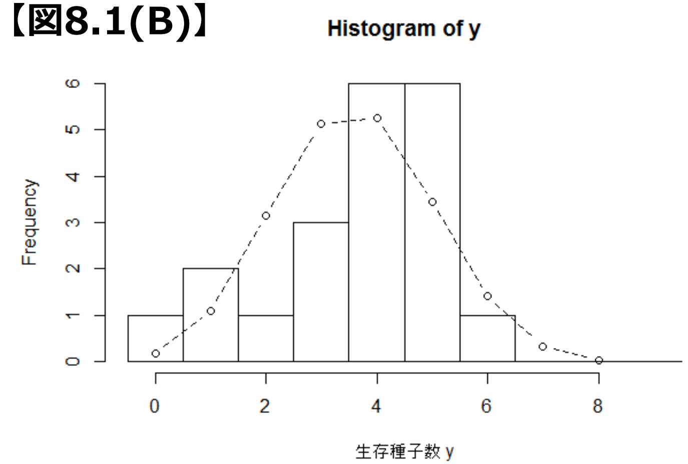
- 20個体に共通する種子生存確率を q とすると，ある個体 i の生存種子数が y_i である確率 $p(y_i|q)$ は以下

$$p(y_i|q) = \binom{8}{y_i} q^{y_i}(1-q)^{8-y_i}$$

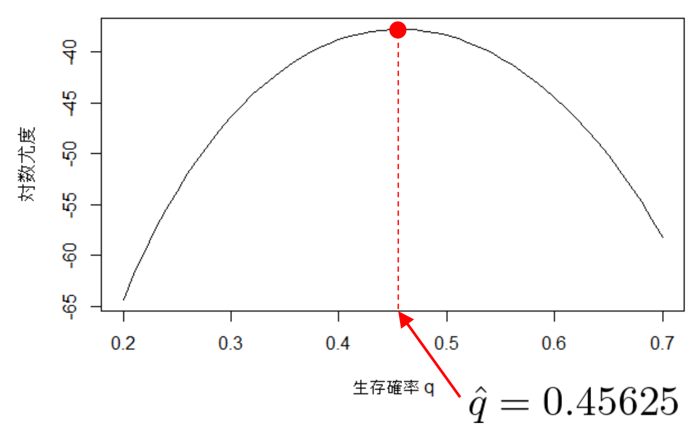
- 次の対数尤度を最大化する q が最尤推定値 $\hat{q}$

$$\log L(q) = \sum_{i=1}^{20} \log p(y_i|q) \;\Rightarrow\; \hat{q} = 0.45625$$

【図8.1(B)】



【図8.2】



■ 最尤推定量 $\hat{q}$ が解析的に求められない場合であっても，計算機を用いて繰り返し試行錯誤をすることによって q を少しずつ変化させることで，対数尤度が高くなる $\hat{q}$ を探しだすことが可能

➘ 効率が最も悪く，精度も良くない「**試行錯誤による最尤推定法**」を紹介

【試行錯誤による最尤推定法の手順】 ※q は離散化して0.01~0.99まで0.01きざみの値をとるとする

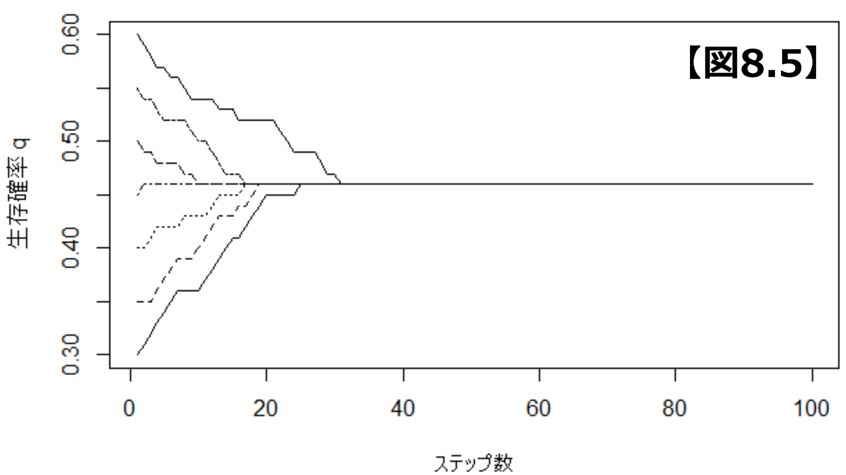
Step1. q の初期値を決める

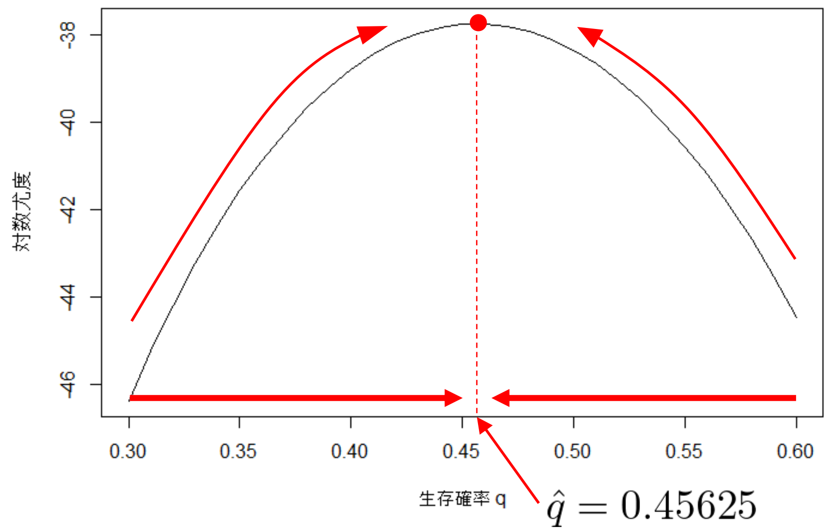
Step2. ランダムに隣の q'を選ぶ（確率0.5で左隣，確率0.5で右隣とする）

Step3. q'での対数尤度が q での対数尤度より大きければ q'に移動する．そうでなければ q にとどまる

➡ q の値が変化しなくなるまでStep2,3を繰り返す

➘ 手順に従って q を変化させていくと，初期値に関係なく，q の値は最尤推定値である $\hat{q} = 0.45625$ の方向に変化していく





■ 前節の「試行錯誤による最尤推定法の手順」を少し修正すると**メトロポリス法**になる

【メトロポリス法の手順】

Step1. q の初期値を決める

Step2. ランダムに隣の q' を選ぶ（確率0.5で左隣，確率0.5で右隣とする）

Step3. q' での対数尤度が q での対数尤度より大きければ q' に移動する

試行錯誤による最尤推定の手順

＋

Step4 $\boxed{\text{$q'$ において尤度が小さくなる場合であっても確率 r で q' に移動する}}$

（ただしこのとき r は尤度比に等しいものとし，$r = \dfrac{L(q')}{L(q)}$ とする．）

メトロポリス法にて新たに加わる手順

➘ メトロポリス法では，$\log L(q) > \log L(q')$ であっても確率 r で移動する

（対数尤度が悪化する度合いが小さいほど（＝ r が大きいほど）移動しやすい)

➘ メトロポリス法のようなルールに基づいて値を変化させていく方法は，
以下の理由から**MCMC（Markov Chain Monte Carlo）**と呼ばれている

- ひとつのステップの中で前の状態 q に基づいて新しい状態 q' を作り出しているので**マルコフ連鎖**になっている
- 一般に乱数を利用した計算アルゴリズムは**モンテカルロ法**と呼ばれるが，
  メトロポリス法のStep2，新たに追加されたStep4において乱数が使われている

■ 試行錯誤による最尤推定とメトロポリス法の違い

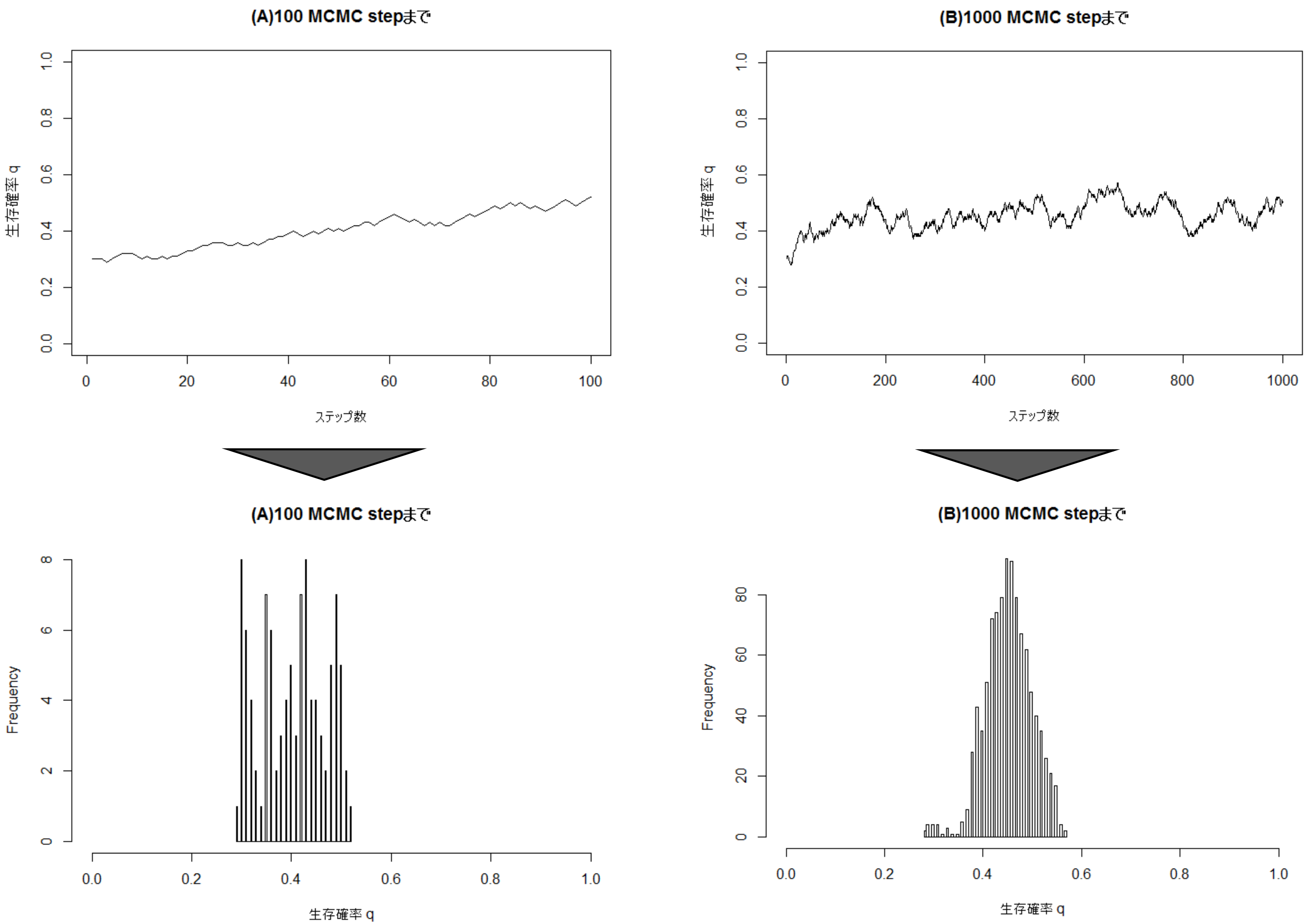
試行錯誤による最尤推定	⟺	**メトロポリス法**
対数尤度が大きくなる方向にしか q が移動しない		対数尤度が小さくなる方向にも q は移動し得る

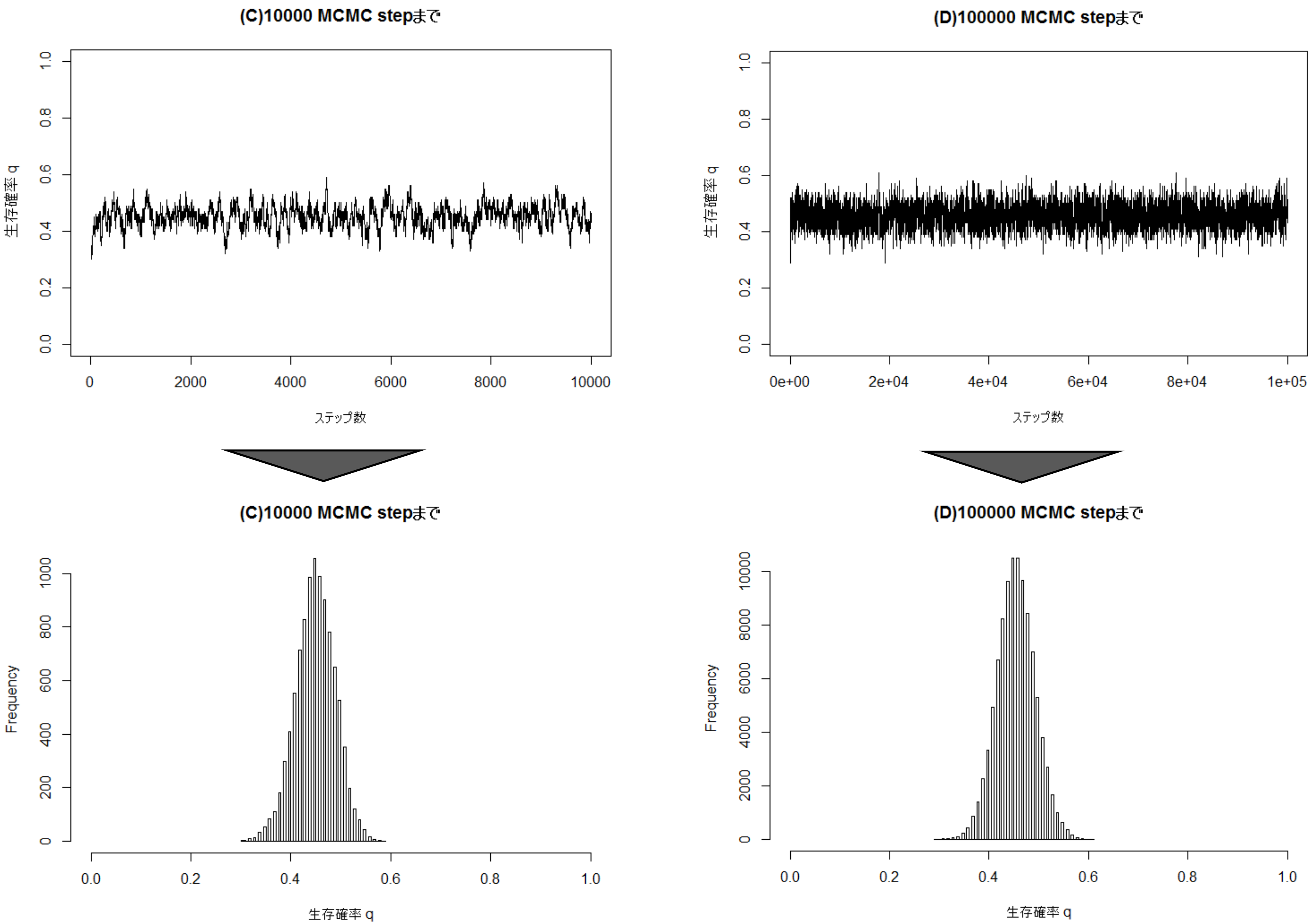
- 試行錯誤による最尤推定と比べて，メトロポリス法は対数尤度の大きいところに到達するまでに時間がかかる
- 試行錯誤による最尤推定では，一度対数尤度が最大になる q に到達すれば，そこから q が動くことはないが，メトロポリス法では，一度対数尤度が最大になる q に到達したとしても，そこから q が動き得る

メトロポリス法などのMCMCアルゴリズムの目的は特定の値の探索ではなく，ステップ数とともに変化するパラメーターの値の生成（これをMCMCサンプリングと呼ぶ）

- 各 q の値に対して，そこに到達した回数をカウントすると度数分布ができる（→ 図8.8参照）
- q を移動させる回数（ステップ数）を増やすと，この度数分布はある確率分布に近づいていく
- この「ある確率分布」を得ることがMCMCの目的

8.3.1 メトロポリス法でサンプリングしてみる【図8.8】




(C)10000 MCMC stepまで
生存確率 q
ステップ数
(D)100000 MCMC stepまで
生存確率 q
ステップ数
(C)10000 MCMC stepまで
Frequency
生存確率 q
(D)100000 MCMC stepまで
Frequency
生存確率 q

■ステップ数を増やしたときに, q の度数分布が近づいていく「ある確率分布」＝ **定常分布**

■**定常分布**：ある変数 q のマルコフ連鎖が一定の条件を満たしているときに,
そのマルコフ連鎖から発生する q の値が従う確率分布（※条件については文献を参照）

➘ 例題のメトロポリス法によって作り出されたマルコフ連鎖はその条件を満たしているので,
ここではとりあえず図8.8のような定常分布 $p(q|Y)$ を持つと考える

➘ ステップ数を増やすにつれて，サンプリングされた q の分布は, q の初期値に関係なく
定常分布に近づいていく（→ 図8.9参照）

➘ ただし，定常分布を近似できるような q の分布を得るためには十分な数のサンプリングが必要
（→ 図8.8参照）

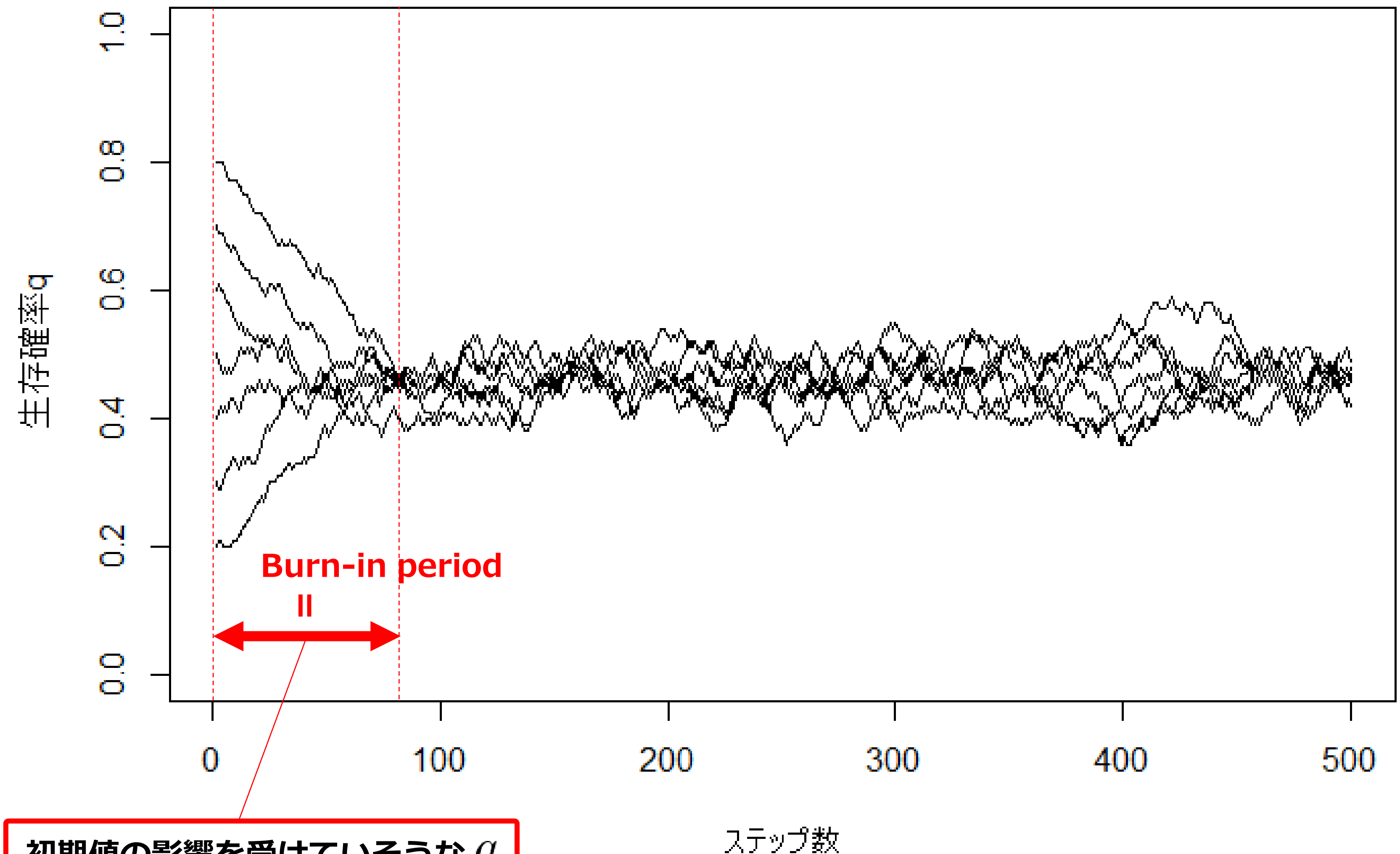
［ ∵ メトロポリス法により q を変化させると，ある q と次の q の間には相関があるため,
すなわち q を「ゆっくり」変化させているため，1000個程度の少ないサンプリングでは,
定常分布からのランダムサンプルには見えない ］

➘ メトロポリス法によって得られた十分に長いMCMCサンプルは定常分布からの
ランダムサンプルになっている（ある q と次の q の間には相関があるが，サンプリングされた順番は無視する）
（→ 8.5.1節参照）

■**より効率的に，かつより精度の高い定常分布の近似を得るには**

➘ メトロポリス法より良い（あるステップと次のステップでサンプルされた q の相関を低くするような）
MCMCアルゴリズムを使用する

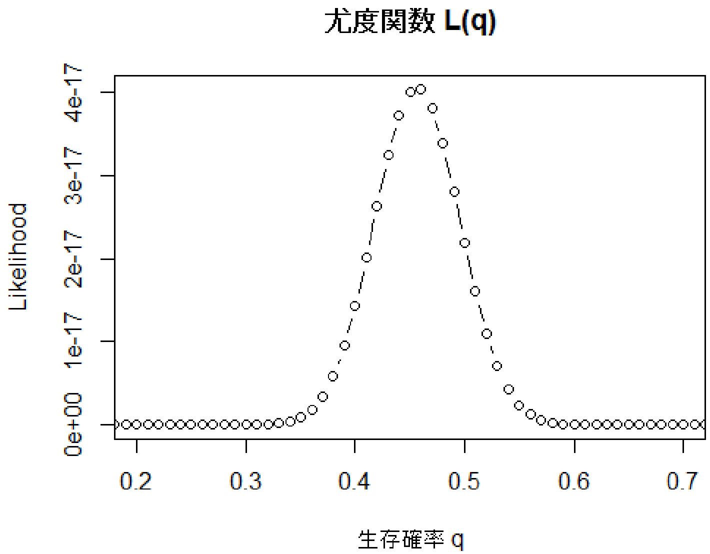
➘ 最初のステップの方でサンプルされた，初期値の影響を受けていそうな q は取り除く（→ 図8.9参照）

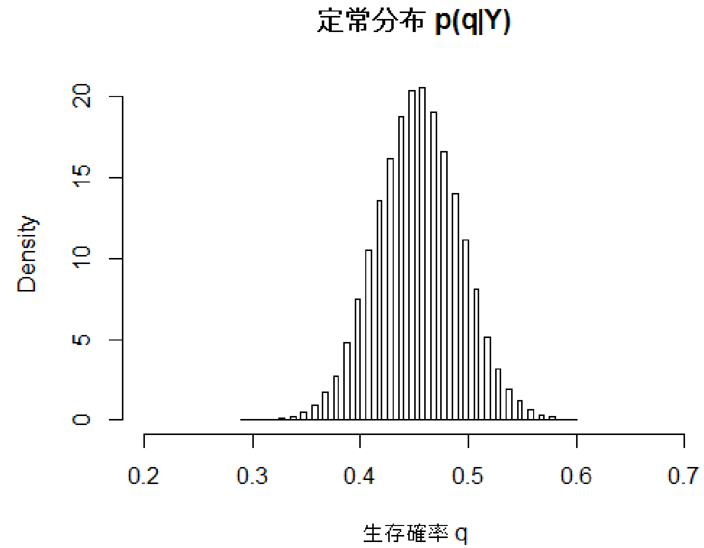


初期値の影響を受けていそうな q

➘ 取り除いた方が良いだろう

■本章の例題について言えば，
メトロポリス法により推定した定常分布 $p(q|Y)$ は尤度 $L(q)$ に比例する確率分布になっている





➘ 尤度 $L(q)$ に比例する離散化した q の確率分布は次のように定義される

$$p(q|Y) = \frac{L(q)}{\sum_q L(q)}$$

すべての $L(q)$ を足し合わせた値＝データ Y だけに依存する定数

➘ $p(q|Y)$は $L(q)$ と比例関係にある

➘ 推定された確率分布 $p(q|Y)$は q の尤もらしさである尤度 $L(q)$ に比例しているので，あるデータ Y に統計モデルをあてはめたときの q がとる値の確率分布と解釈できる

– 従って，MCMCサンプリングは統計モデルのあてはめの一種
– この分布の平均値・中央値・標準偏差・95%信頼区間といった統計量を評価することで，あてはめの結果として得られた q の分布を要約して示すことが可能

■MCMCサンプリングは，統計モデルを観測データに当てはめる方法の一つであり，
その結果として，与えられたデータとモデルの下でのパラメーターの確率分布が得られた

➘ このようにパラメーターを確率変数として扱うのが**ベイズ統計学**の特徴
（⇔ パラメーターには真の値が一つあると考える**頻度主義**）

■ベイズ統計学は，**ベイズの公式**の形式で推論を行なう統計学

➘ 本章の例題の統計モデルに対応するようにベイズの公式を導入してみる．
種子の生存確率 q をベイズの公式の形式で書くと次のようになる

$$\underbrace{p(q\mid Y)}_{\text{事後分布}} = \frac{\underbrace{p(Y\mid q)}_{\text{尤度}}\,\underbrace{p(q)}_{\text{事前分布}}}{\sum_q p(Y\mid q)\,p(q)}$$

– $p(q|Y)$：データ Y が得られたときに q が従う確率分布＝**事後分布**

– $p(Y|q)$： q の値が決まっているときにデータ Y が観測される確率
（この例題では，二項分布の積である尤度 $L(q)$ が相当するので，$p(Y|q) = L(q)$ となる）

– $p(q)$ ：データ Y がないときの q の確率分布＝**事前分布**

–右辺分母：規格化定数（$\sum_q p(Y|q)p(q) = p(Y)$）

➘ つまりベイズ統計モデルとは，次のような構造を持つ統計モデル

$$\text{事後分布} = \frac{\text{尤度}\times\text{事前分布}}{\text{データが得られる確率}} \propto \text{尤度}\times\text{事前分布}$$

■8.3.3項では，MCMCサンプリングによって定常分布 $p(q|Y)$ は尤度と比例関係にあるということが数値実験的に示されていた

➘ そこで以下の二つの式を見比べてみる

【定常分布 ∝ 尤度】 $p(q|Y) = \dfrac{L(q)}{\sum_q L(q)}$ ⟺ **【ベイズの公式】** $p(q|Y) = \dfrac{p(Y|q)p(q)}{\sum_q p(Y|q)p(q)}$

➘ 事前分布 $p(q)$ が q の値に依存せず定数となっていれば，二つの式は同じものであるとみなせる

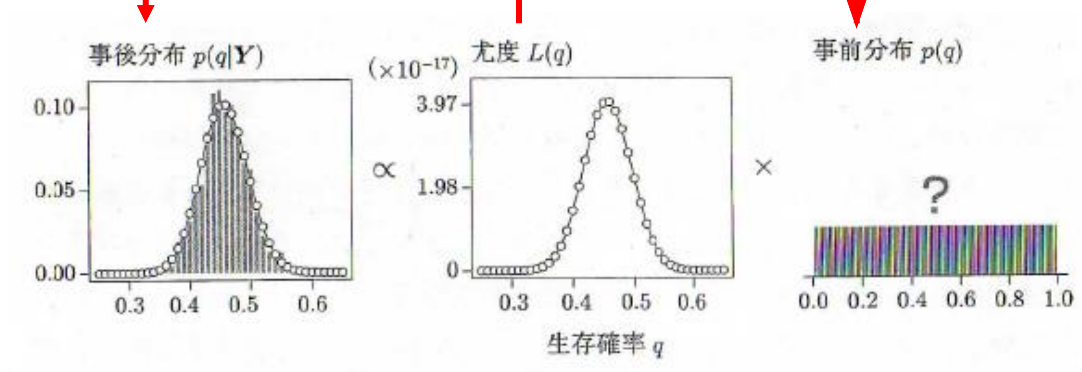
➘ 従って本章の例題では，パラメータ q の事前分布として例えば $p(q) = 1/99$ のような離散一様分布を仮定し，この事前分布とデータを使って，ベイズ統計学の枠組みの中で，パラメーター q の事後分布を推定していたということになる

➘ モデル設計の段階で事前分布・尤度をきちんと指定し，それに整合するようにMCMCサンプリングを実施して事後分布を推定するというのがベイズ統計モデリングの枠組み

【図8.13】

③ 事前分布を更新

② 事前分布にデータが持つ情報を追加



① 事前分布を仮定

■ **メトロポリス法を用いてパラメーター q のサンプリングを行なうと，それがなぜ定常分布 $p(q|Y)$ からのランダムサンプルになるのか**

➘ メトロポリス法で得られた q のサンプルが定常分布からのランダムサンプルであるためには，次の二つの条件が成立していなければならない

> ① q が任意の初期値から定常分布 $p(q|Y)$ に収束する
> ② ある q が定常分布 $p(q|Y)$ に従っていて，メトロポリス法で次の q^* を得たときに，この q^* も定常分布 $p(q|Y)$ に従っている

➘ 本章の例題の場合，これら二つの条件はどちらも成立している
ここでは簡単のため，条件②のみ検討する

Case(i) : ある時点でパラメーター q が $p(q|Y)$ からランダムに得られた値であり，メトロポリス法によって次の値として q^* が選ばれ，尤度が改善された（ $L(q) < L(q^*)$ ）とする

– このとき, q から q^* へと移動する確率 $p(q \rightarrow q^*)$ は次のようになる

$$p(q \rightarrow q^*) = 0.5 \times 1 = 0.5 \cdots (1)$$

– 逆に q^* から q へ移動する確率 $p(q^* \rightarrow q)$ は，尤度が悪くなるので次のようになる

$$p(q^* \rightarrow q) = 0.5 \times r = 0.5 \times \frac{L(q)}{L(q^*)} \cdots (2)$$

➘ (1), (2)式から次の関係が導ける： $L(q^*)p(q^* \rightarrow q) = L(q)p(q \rightarrow q^*) \cdots (3)$

➘ ここで $p(q|Y) = \frac{L(q)}{\sum_q L(q)}$ なので，次が得られる： $p(q^*|Y)p(q^* \rightarrow q) = p(q|Y)p(q \rightarrow q^*) \cdots (4)$

【補足】8.5.1 メトロポリス法と定常分布の関係（2/2）

Case(ii) : q から q^* へ移動しても尤度が改善されない場合（ $L(q) > L(q^*)$ の場合）

➘ Case(i)のときと同様に(4)式の関係が成立する

■Case(i), (ii)より，q から q^* への移動が尤度の改善になるかどうかに関係なく，(4)式の関係は成り立つ

➘ (4)式の関係は，**詳細釣り合いの条件**と呼ばれる

➘ (4)式の両辺について，すべての q について和をとると次が得られる

$$\sum_q p(q^*|Y)p(q^* \to q) = \sum_q p(q|Y)p(q \to q^*)$$

➘ ここで $p(q^*|Y)$ は q についての和とは無関係で，確率の定義から $\sum_q p(q^* \to q) = 1$ なので次が得られる

$$p(q^*|Y) = \sum_q p(q|Y)p(q \to q^*)$$

➘ 上式右辺：q が定常分布 $p(q|Y)$ に従っているときにメトロポリス法によって q が q^* になる確率
これが左辺の $p(q^*|Y)$ と等しいということは次を意味している

q が定常分布 $p(q|Y)$ に従っているならば，
メトロポリス法により選ばれた q^* もまた定常分布 $p(q^*|Y)$ に従っている

（これはメトロポリス法に限らず，すべてのMCMCアルゴリズムに共通する性質）

■ベイズの定理

・本章8.4節で登場した事後分布 $p(q|Y)$ や事前分布 $p(q)$ などの関係を記述した次式は**ベイズの定理**と呼ばれる

$$p(q|Y) = \frac{p(Y|q)p(q)}{\sum_q p(Y|q)p(q)}$$

➘ これは次の条件付き確率と同時確率の関係を整理したものに過ぎない

$$p(A|B)p(B) = p(A, B)$$

➘ 本章の例題に沿って q と Y の条件付き確率と同時確率の関係を整理してみる

– 条件付き確率と同時確率の定義より，データ Y が決まったときに q が特定の値をとる条件付き確率は次のように書ける：

$$p(q|Y) = \frac{p(Y, q)}{p(Y)}$$

– ここで右辺分子は $p(Y, q) = p(Y|q)p(q)$ と書くことができることを利用すると次が得られる

$$p(q|Y) = \frac{p(Y|q)p(q)}{p(Y)} = \frac{p(Y|q)p(q)}{\sum_q p(Y|q)p(q)}$$

– これが，よく見かける**ベイズの公式**である．