8.4 MCMCサンプリングとベイズ

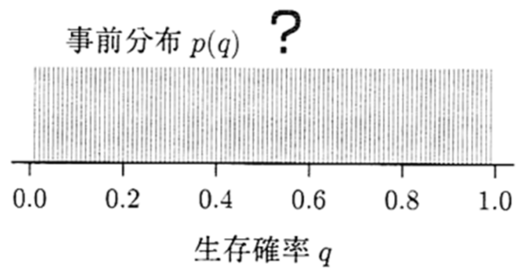
ベイズ統計モデルというのはこういう構造

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

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

↑8.3.3で出てきた

これを比べてみると、$p(q)$=定数 ということ



8.4 MCMCサンプリングとベイズ

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

これを比べてみると、$p(q)$=定数 ということ
つまり、こんな↓感じであれば辻褄が合いそう

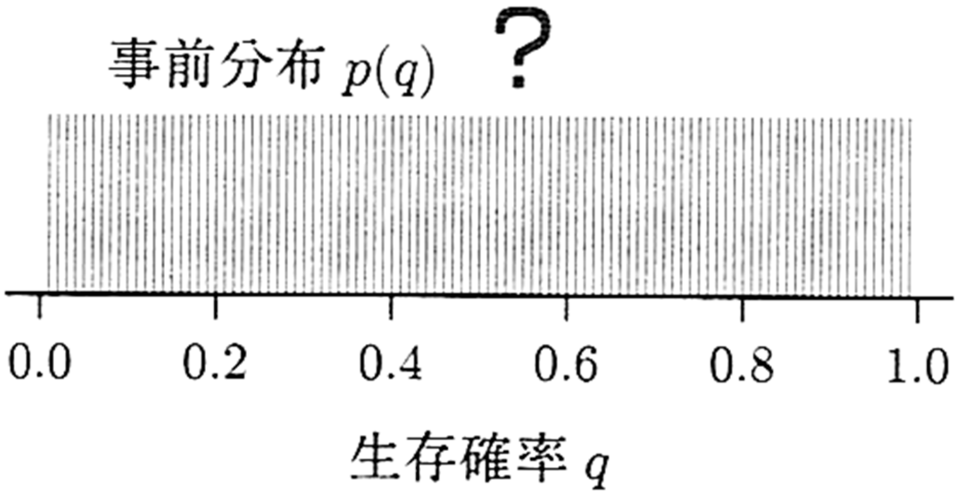


図8.12

8.4 MCMCサンプリングとベイズ

事後分布・尤度・事前分布の関係を
整理するしてみると、こんな↓感じ

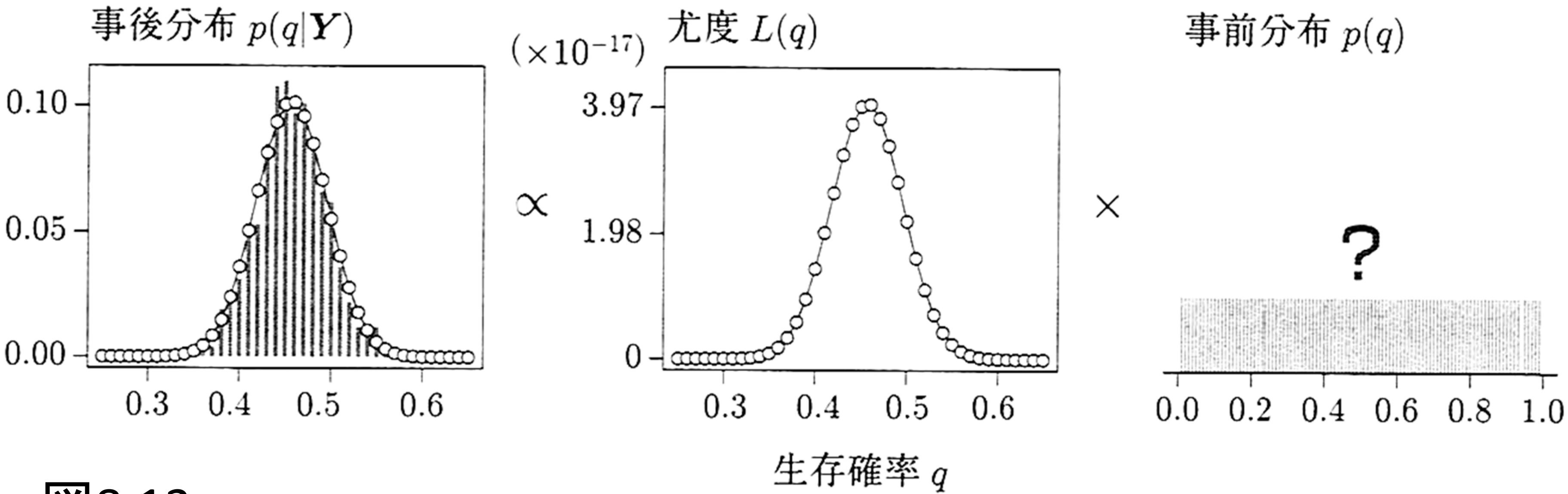
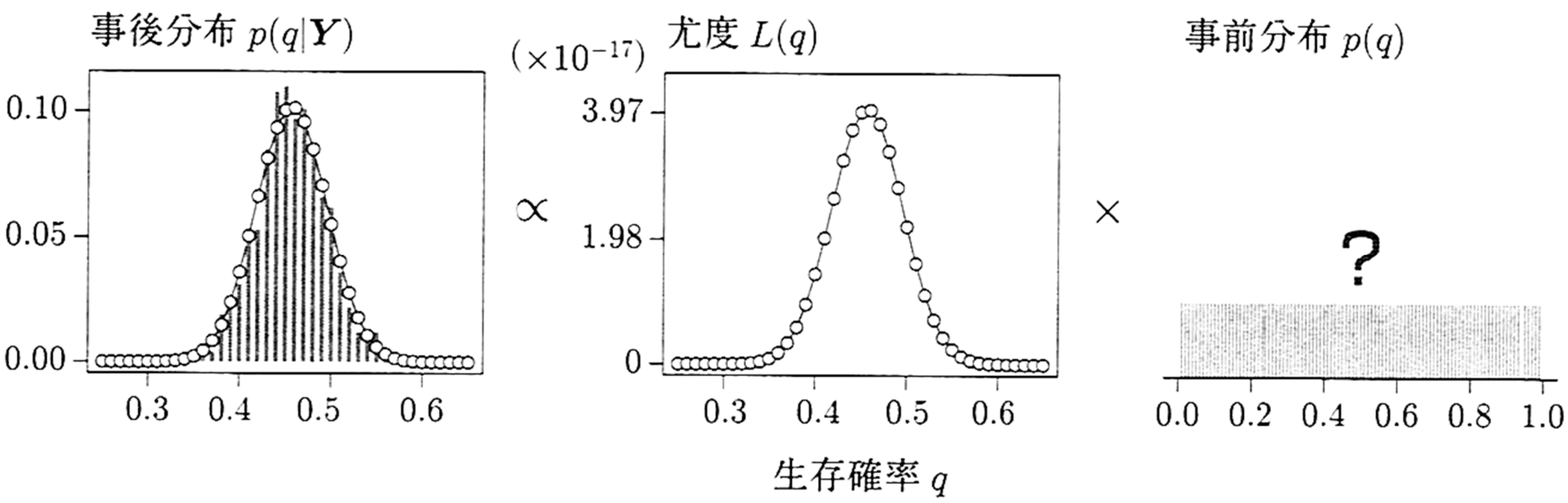


図8.13

8.4 MCMCサンプリングとベイズ

今回はあくまで説明のために、
あとづけで事前分布について考えましたたが
実際の当てはめでは尤度・事前分布を考えてから
MCMCサンプリングを行います



8.5.2 (補足)ベイズの定理

条件付き確率と事前分布の関係を整理したもの

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

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

ざっくり私の言葉で例を上げて説明します
病気である確率qと検査陽性Yについて
書きかえるとこうなる↓

$$p(\text{病}|\text{陽性})p(\text{陽性}) = p(\text{陽性}|\text{病})p(\text{病})$$

8.5.2 (補足)ベイズの定理

どういう状況か？

X病は不治の病で、罹患・発症すると昼夜激痛に苛まれたのち1ヶ月後に100%死亡する。
Aさんが X病検査キットを使ってみたら、陽性だった。
その検査キットの検出率は99%である。
また、擬陽性率は0.1%である。
ただし、X病は超珍病で、1億人に1人しかいない。

このとき、検査結果が陽性になったAさんが
本当にX病に罹患している確率は？ $p(\text{病}|\text{陽性})$

8.5.2 (補足)ベイズの定理

本当にX病に罹患している確率　p(病|陽性)

p(病|陽性)p(陽性) = p(陽性|病)p(病)だから

p(病|陽性) = p(陽性|病)p(病)/ p(陽性)

ここで

p(陽性|病)：その検査の正しい検出率＝0.99

p(病)：その病気の発生率＝1/100,000,000

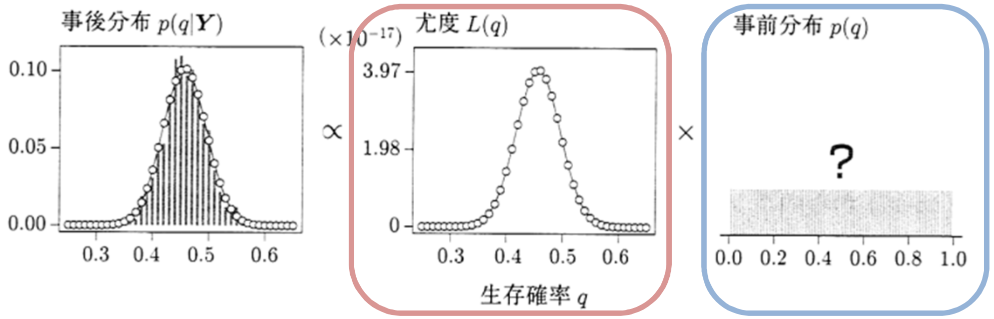
p(陽性)：何もなくても陽性＝擬陽性の率 ＝0.001

p(病|陽性) ＝0.99*(1/100,000,000)/0.001
＝0.0000099≒0.001%

8.5.2 (補足)ベイズの定理

99%という確率の検査で陽性なのに？
<u>病気の発生率</u>が 1/1億 であることが重要
　↑これが事前分布 prior distribution

検査の尤もらしさ＝**尤度**が
低い病気の発生率＝**事前分布**で調整された



8.6 まとめ

- 最尤推定は尤度最適化でパラメータ推定
  - ふらふら試行錯誤・数値計算

- MCMC法は尤度に比例する定常分布を得る
  - ただパラメータの分布という考え方は変
  - ベイズ統計の世界ならOK
- ベイズ統計なら、定常分布＝**事後分布**

これがベイズでMCMCして推定する理屈