2010/5/28

Bayesian Inference
4章前半 (～p59)

担当　川森 愛(生命科学院)

未知のパラメータ θ の推定

尤度 (likelihood)

事後分布
(posterior distribution)

$$[\theta|Y] = \frac{[Y|\theta][\theta]}{[Y]}$$

正規化定数
(normalizing constant)

もっと単純に．．．

$$[\theta|Y] \propto [Y|\theta][\theta]$$

θ の事後分布を考えるだけなら，正規化定数は重要でない．

- 事後分布の決定に必要ないといっても，平均や分散，あるいは特定の確率Pr(θ>0| ϒ)などを知るために[ϒ]の情報が必要．
- [ϒ]は以下の式で求まる．

$$[Y] = \int [Y|\theta][\theta]\, d\theta$$

しかし，上式の積分はしばしば困難である．

解決策

- 共役分布を使う (→ Section 4.1)

  共役分布：事前分布に尤度関数を掛けて事後分布を求めると，同じ形をとる関数．計算上の問題は解決するけれど，適用できる例は少ない．

- モンテカルロ法 (→ Section 4.2, 4.3)

  $[\theta|Y]$からθのサンプルをとってくれば，そこから事後分布の特徴もわかる．サンプルの平均値は事後分布の平均に対応するし，$\theta>0$の割合は$\Pr(\theta>0|Y)$に対応する．

  十分に大きいサンプルをとってくれば，精度も十分高くなる．

4.1 共役分布 (conjugacy)

分布を決定する超パラメータ Ψ

updating

Ψ_0 → Ψ_1

事前分布

(データを得る前)

事後分布

(データを得た後)

事前分布，事後分布の形が同じなので，
Ψのアップデートとして考えればよい．

4.1.1 Hautapu Trout

Hautapu川のブラウントラウト調査

- 13匹をマーキングして放流
- 調査区域はネットで覆って出入りなし
- 3回の調査で見つけたマーキング個体の数：

  10，7，8 (計25)
- 発見確率pを推定する

ブラウントラウトが何匹見つかるか →二項分布

p → ベータ分布 (パラメータa_0, b_0)

$$\frac{p^{a_0-1}(1-p)^{b_0-1}}{Be(a_0,b_0)}$$

$$[p|Y] \propto [Y|p] \times [p]$$

Ψ_0
$=\{a_0, b_0\}$

$$\propto p^{25}(1-p)^{39-25} \times p^{a_0-1}(1-p)^{b_0-1}$$

***2項分布**

***ベータ分布**
(ベータ関数の部分を省略)

Ψ_1
$=\{a_1, b_1\}$

$$= p^{a_1-1}(1-p)^{b_1-1}$$

***事後分布もベータ分布**

ただし， $a_1 = 25 + a_0$, $b_1 = 39\text{-}25+b_0$

The beta family of prior is conjugate for the binomial likelihood.

パラメータa_0, b_0 はどうやって決めればいいのか？

情報がないので，一様な分布(a_0=1, b_0=1)とする

⇒ Figure 4.1 (p. 50)のような事後分布が描ける

- 事後分布のモードは最尤推定量と同じ.
- 0.488から0.777の間に95%のpがある.
- 一度の潜水で半分以上の魚がみつかると，
  96%確信できる.

4.2 モンテカルロ法

事前分布や事後分布を確率分布で考えるベイズ法はとても良いアイディアだけど，計算が大変

➡ Markov chain Monte Carlo (MCMC) 法 が解決した!

MCMC法あれこれ

MCMCの基本的技術は1950年代に開発.
しかし1980年代後半まで注目されず.
生態学などに使われだしたのは2000年をすぎてから.

”生活史データに対してMCMC解析を行ないました”なんて使い方は間違い．MCMCはただの道具にすぎない．重要なのはどんなモデルを使ったか，という情報.

モンテカルロ法

平均寿命に対する最大寿命の比が知りたいとする.

寿命は指数関数で表わされるとする.

サンプル数 25

指数分布から計算して求めるのは大変!

⇒ 10万回乱数を生成して，結果をみればよい.
(3.2 GHz ラップトップで8秒くらい)

*1回1回のサンプリングが独立に行なわれるところがMCMC法と違う.

4.2.2 Rejection Sampling

分布c(x)からサンプリングしたものを，分布t(x)からサンプリングしたものに変換する.

(ヒストグラムが描かれた大理石があるとして，ヒストグラムだけ残してまわりを削り落とすイメージ．)

使い方1.　1度サンプリングしたものを，条件を変えて調べる.

例) Xをサンプリングしてきたけれど，X>0の期待値を知りたい.

→ $X \leqq 0$のサンプルを捨てる.

(5.5.2で実践する)

使い方2.　t(x)からサンプリングするのが難しいので，
簡単なc(x)からサンプリングする．

t(x) : target
c(x) : candidate　と呼ぶ．

- t(x)より常に大きいMc(x)を見つける．
- 棄却率の計算

$$w(X) = t(X) / Mc(X)$$

w(x)を確率として，ベルヌーイ試行により成功なら採択，
そうでなければ棄却する．

→ Figure 4.4