第四回ベイズ統計学・機械学習研究会

～自然共役事前分布・MAP推定・MCMC～

もくじ

- ベイズの定理のおさらい
- ベイズ統計学の流れと問題点
- 事後分布を求める3つの方法：概要
    - 自然共役事前分布
    - 点推定
    - MCMC（マルコフ連鎖モンテカルロ）
- 次回以降のための準備
    - 確率的プログラミング言語
    - Google Colabへの登録

ベイズの定理のおさらい

- 1740年ごろトーマス・ベイズが発見した条件付き確率に関して成り立つ定理
- $P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$
- ベイズ統計学はこの定理を応用して予測を行う

ベイズ統計学の流れ

尤度関数　　事前分布

事後分布

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

エビデンス

ベイズ統計学の流れ

尤度関数　事前分布

事後分布

$$P(\theta|D) = \frac{P(D|\theta)P(\theta)}{P(D)}$$

エビデンス

- 尤度関数
  - 分析したいデータの分布から決定する
    - コインの裏表：二項分布
    - テストの点：正規分布
    - 交通事故の件数：ポアソン分布
    - 身長と体重：線形回帰
- 事前分布
  - 尤度関数をもとに自分で決定する

- エビデンス（周辺尤度）
  - 尤度関数と事前分布の同時分布のとりうる値を合算したもの（周辺化）
  - 尤度関数と事前分布から計算可能
    - 分子が離散：取りうる値を合計
    - 分子が連続：分子を積分

ベイズ統計学の流れ

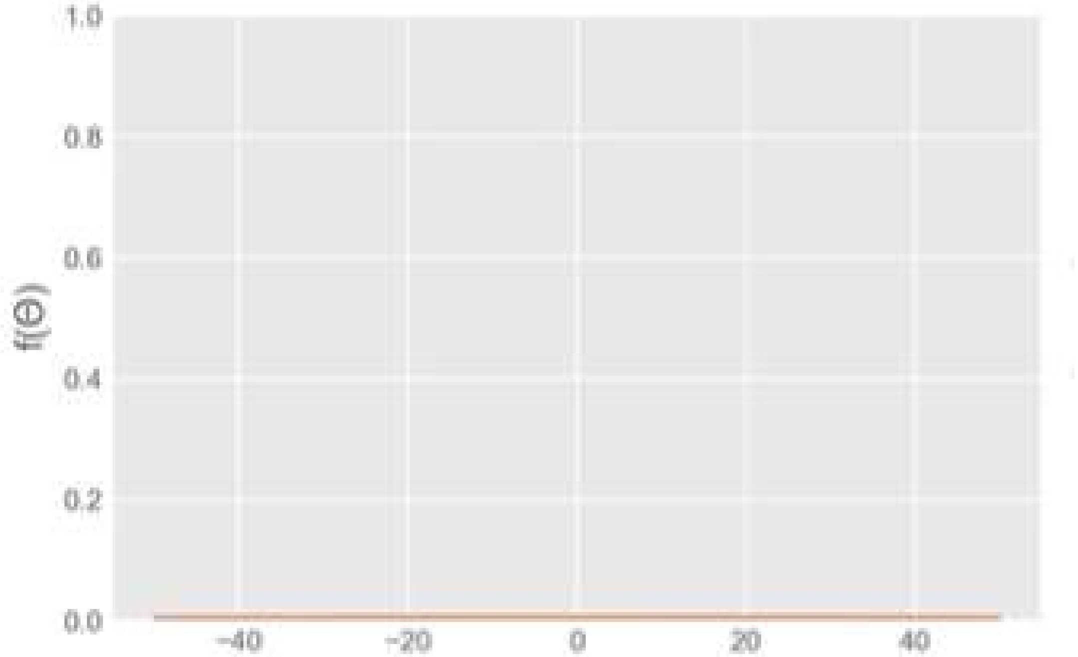
1. 下調べ：問題の背景を調べる
   - 原因と結果の関係の仮説を立てる
     - データの可視化など
2. 統計モデルの作成
   - １の情報の基づきモデルを作成
3. 事後分布の計算
   - 得られたデータと仮説から事後分布を計算

ベイズ統計学の問題点

- 事前分布を自分で決定する必要がある
  - 事前分布の選択に関する議論は決着していない
- 本研究会では 豊田(2015)『基礎からのベイズ統計学』の立場に従う
- 「公的分析では無情報事前分布を使用し、尤度への影響を最小限にする。」

無情報事前分布

- その事前分布を用いて得られる事後分布に，その事前分布ができるだけ影響しないような事前分布
- たとえば平均ゼロで標準偏差 10 の「すごくひらべったい正規分布」



事後分布を求める3つの方法

- 自然共役分布
- 点推定
- MCMC（マルコフ連鎖モンテカルロ）

自然共役分布

- 尤度関数ごとに相性の良い事前分布が存在する
- 尤度関数が単純（正規分布や二項分布などに）ならば使用可能
- 事後分布は自然共役分布の形になる
- 自分で事前分布のパラメータを決める必要があり、公的分析には向かない

自然共役分布一覧

自然共役分布	尤度
Beta分布	ベルヌーイ分布
Beta分布	二項分布
Gamma分布	ポアソン分布
正規分布	正規分布の平均
逆Gamma分布	正規分布の分散

MAP推定

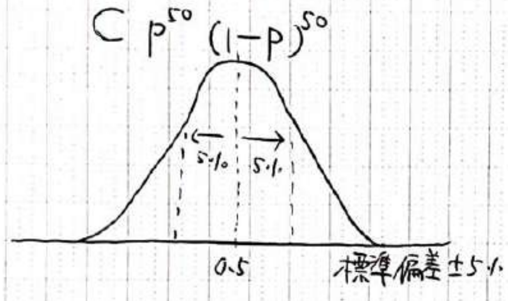
- 確率分布全体を求めるのはあきらめる
- 代わりに事後分布が最大の点を求める

MAP推定

- 第一回研究会資料より
- これは二項分布の自然共役分布であるベータ関数を事前分布として点推定

・今回つかう事前分布

$$C p^{50}(1-p)^{50}$$



100回投げたら、だいたい50%表が出るだろう。

コイン=A

事前分布 $C p^{50}\times(1-p)^{50}$ を仮定

$$p^{post}(p) = C\times p^{50}\times(1-p)^{50}\times D\times p^{2}\times(1-p)^{1}$$

$$= E\times p^{52}\times(1-p)^{51}$$

$$P_A = \frac{52}{52+51} = \frac{52}{103} = 50,49\%$$

コイン=Bも同様に。

$$p^{post}(p) = C p^{50}(1-p)^{50}\times D\times p^{60}\times(1-p)^{40}$$

$$= E\times p^{110}\times(1-p)^{90}$$

$$P_B = \frac{110}{110+90} = \frac{110}{200} = 55\%$$

ベイズ統計の力をつかうと、、

コイン=Aが表になる確率 50,49%

コイン=B 〃 55%

よりそれっぽい分析ができる!!

MAP推定

- MAP推定で無情報事前分布を用いると最尤推定法の結果と一致する
  - （一致しないってどこかで見た気がするので訂正するかも。。。）

Q. n回投げてk回表が出たコインの表が出る確率 p_0 を予測せよ。

コインの表が出る確率は、

$$\underbrace{{}_nC_k\, p^k (1-p)^{n-k}}_{\text{pの値によって確率はかわる}}$$ と表せる。

→これが最大になるpをさがす

$$L(p) = {}_nC_k \cdot p^k \cdot (1-p)^{n-k}$$

↑これを微分して最大になるpをさがす

L(p)を最大にするpもlogL(p)を最大にするpも同じなので、対数をとって

$$\log L(p) = \log {}_nC_k \cdot p^k \cdot (1-p)^{n-k}$$

$$= \log {}_nC_k + k\log p + (n-k)\log(1-p)$$

$$\frac{d}{dp}\log L(p) = \underbrace{\frac{k}{p} - \frac{(n-k)}{(1-p)}}_{\text{これが0になれば良いので}}$$

$$0 = \frac{k}{p} - \frac{(n-k)}{(1-p)}$$

$$p = \frac{k}{n}$$

おまけ

$(n-k)\log(1-p)$ の微分。

$\log(1-p)$ を合成関数だと考えて、

$y = \log x \quad x = 1-p$ をそれぞれ微分して、

$$\frac{dy}{dx} = y' = \frac{1}{x}$$ xを代入して、

$$= \frac{1}{1-p} \cdots ①$$

$$\frac{dx}{dp} = x' = -1 \cdots ②$$

①、②より

$$\frac{dy}{dx}\cdot\frac{dx}{dp} = \frac{dy}{dp} = \frac{1}{1-p} \times -1$$

$$= -\frac{1}{(1-p)}$$

CamScannerでスキャン

MCMC（マルコフ連鎖モンテカルロ）

- スライド５でエビデンス（周辺尤度）を積分で計算するとある
- そんなことはめったにできない
- コンピューターでエビデンス（周辺尤度）の近似値を求める
- 事後分布の形は判っているので、それをもとに乱数を発生させる

→MCMC（マルコフ連鎖モンテカルロ）

MCMC（マルコフ連鎖モンテカルロ）

- マルコフ連鎖モンテカルロは**マルコフ連鎖**と**モンテカルロ**に分けられる
- マルコフ連鎖
  - 次の状態が過去の状態に依存せず現在の状態のみによって決まる性質のこと
- モンテカルロ法
  - 確率変数のサンプリングをコンピューターを用いて行うことで、数学的問題を数値的に解く手法

次回以降のための準備

- 確率的プログラミング言語
- Google Colaboratory への登録

確率的プログラミング言語

- 確率的プログラミング言語とは
  - 統計的モデリングのためのツール
  - プログラミングの世界の力を借りて手計算では不可能なモデリングを可能にする
  - 確率プログラミング言語を使ってMCMCサンプリングを行う
- Stan
  - ベイズ推定を高速で行うことができる
  - RやPythonなど様々な言語から呼び出せる
  - C＋＋に変換して、RやPythonで呼び出せる形にコンパイル
- TensorFlow Probability
  - GoogleのTensorFlowをもとに作られた
  - Stanより高速
  - 開発途中
  - 日本語解説も少ない
- PyMC
- Pyro
  - Uberが開発

Google Colaboratoryへの登録

- Google Colaboratoryとは
  - ブラウザから Python を記述、実行できるサービス
    - 環境構築が不要
    - GPU への無料アクセス
    - 簡単に共有
- PyStanがすぐに使用可能なため利用します
  - （Rstanもいいかなと思ったんですけど。。。）
  - （Rstanの環境構築で3日潰したので嫌になりました。。。）

まとめ

- ベイズ統計学の事前分布の設定には注意が必要
- 本講義では無情報事前分布を使用
- MCMCで事後分布を数値的に求める