数理統計 補助資料 ～ベイズ推定～

2023年度2学期:　月曜1限, 水曜3限
担当教員:　石垣　司

ベイズ推定とは？

ベイズ統計学におけるパラメータの推定の総称

ベイズ推定の原理

- パラメータ $\boldsymbol{\theta}$ を確率変数と見なして, 事後分布 $p(\boldsymbol{\theta}|\boldsymbol{y})$ や事後分布の代表値 $\hat{\theta}$ を求める
- 考え方: 事前分布 $p(\boldsymbol{\theta})$ が所与のとき, 事後分布 $p(\boldsymbol{\theta}|\boldsymbol{y})$ はデータ所与後の統計モデルのパラメータの情報そのものである
  最小2乗推定, 最尤推定とは"原理"が異なる点に注意(本授業での3つ目の原理)

今回の授業のテーマ:
事後分布 $p(\boldsymbol{\theta}|\boldsymbol{y})$ やその代表値を求める方法の概要

#メモ　実際の推定方法の数理的内容は、一部を除いて学部2年生の学習内容を大きく超えてしまうので本授業では紹介できない。しかし、それでもなお、推定のブラックボックス化は避けてほしいという思いがあるのでこの授業を行う。そのため、本日の授業はイメージ優先となってしまう。

ベイズ推定の実現方法

ベイズ推定の難しさ

- 分母の積分の計算が解析的に解けるとは限らない
  一般の関数の積分 & パラメータの多重積分になっている

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{\prod_{i=1}^{N} p(y_i;\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{\int p(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

ベイズ推定の方法の大別

1. 共役事前分布の利用
2. 事後分布の解析的近似
3. マルコフ連鎖モンテカルロ法(MCMC法)
4. 最大事後確率推定(MAP推定)

1. 共役事前分布の利用 #1

共役事前分布

- 事前分布と事後分布が同じクラスの確率分布となるとき, そのような事前分布を共役事前分布とよぶ
  読み方: 共役(キョウヤク)
  例: 正規分布×正規分布＝正規分布

共役事前分布を利用した推定

- 長所: 解析的に事後分布が導出できる
  数値計算の必要なし. 近似無しの正確な事後分布を計算可能
- 短所: 利用できる分布が限られている
  前回の授業で導出した事後分布(正規分布&単回帰モデル)は共役事前分布を利用していたため解析的に導出できた
  共役事前分布が存在しない統計モデルのパラメータも多数

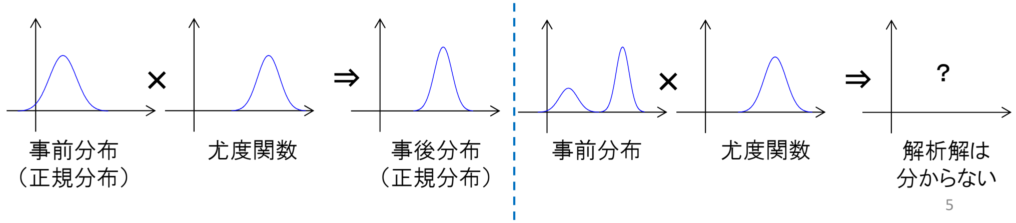
1. 共役事前分布の利用 #2

代表的な共役事前分布

#メモ　正規分布以外の分布の説明は授業の主旨とズレてしまうので詳細は割愛

推定したいパラメータ	共役事前分布	事後分布
正規分布の平均	正規分布	正規分布
（単回帰モデルの傾き）	正規分布	正規分布
正規分布の分散	逆ガンマ分布	逆ガンマ分布
ポアソン分布の平均	ガンマ分布	ガンマ分布
二項分布の確率関数	ベータ分布	ベータ分布
多項分布の確率関数	ディリクレ分布	ディリクレ分布

【正規分布の平均 μ の推定】

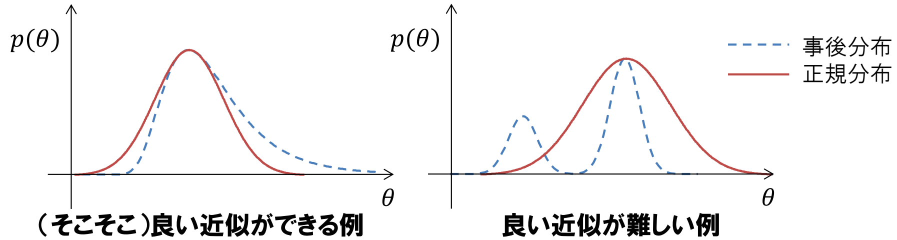


2. 事後分布の解析的近似

簡単な関数で事後分布を近似する

代表例：ラプラス近似

- 事後分布を正規分布により近似する（テイラー展開による2次近似）
- 長所：計算コストが低い
- 短所：複雑な形状の事後分布は良い近似ができない

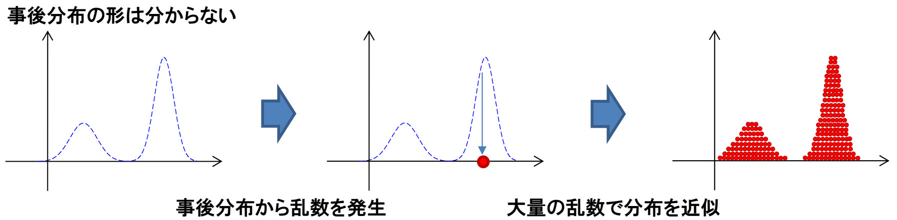


#メモ　計算コストとは事後分布の計算に必要な計算量や計算時間を指す。

3. マルコフ連鎖モンテカルロ法（MCMC法）

乱数により事後分布を近似する手法の総称

- 長所：実用的なほとんどの事後分布の推定に適用可能。乱数の数を増やすと，真の事後分布と一致する性質
  現在では最もスタンダードなベイズ推定のための手法
- 短所：計算コストが高い。処理を並列化できないため，本質的な高速化が困難

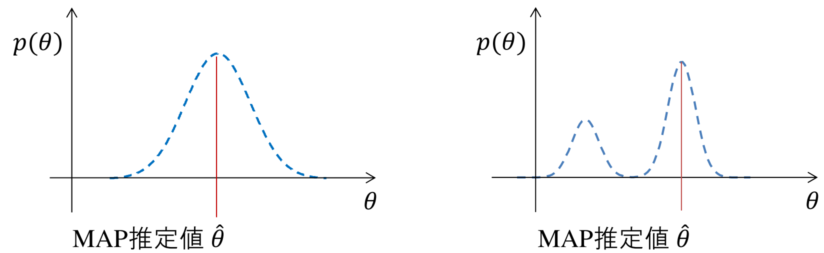


#メモ　1990年代以降，ベイズ統計学が普及してきた技術的な理由は計算機の高速化によってMCMC法が実用的になったため。現在でもより効率的な手法の開発が盛んに研究されている

4. 最大事後確率（MAP）推定

事後分布が最大となる点のみを推定

- 事後分布の点推定
  $\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}}\{p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})\}$ を探す（"$p(\boldsymbol{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$が最も大きくなる$\boldsymbol{\theta}$"という記号）
- 長所：MCMC法と比べて計算コストが低い
- 短所：事後分布の最大となる点のみを求めるので，他の手法と比べて事後分布の形状に関する情報が喪失
  下の両図の事後分布の形状は区別できない



MAP推定の例

確率分布 $N(\mu,\sigma^2)$ の平均 μ のベイズ推定

- **データ** $y_i \sim i.i.d.\, N(\mu,\sigma^2), (i=1,\dots,N)$
- **分散** σ^2 **は既知**
- **平均** μ **の事前分布：** $p(\mu)=N(\mu;m,v^2)$
- **尤度関数：** $p(\boldsymbol{y}|\mu)=\prod_{i=1}^{N}\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(y_i-\mu)^2\right\}$
- $\hat{\mu}=\underset{\mu}{\operatorname{argmax}}\{p(\boldsymbol{y}|\mu)p(\mu)\}$ **は** $\frac{\partial \log\{p(\boldsymbol{y}|\mu)p(\mu)\}}{\partial\mu}=0$ **となる点**
- MAP**推定量：**

$$\hat{\mu}=\frac{N\bar{y}v^2+m\sigma^2}{Nv^2+\sigma^2}$$

正規分布は左右対象で単峰のため事後分布の平均に一致

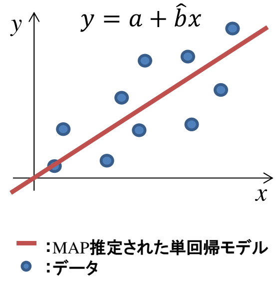
MAP推定と事後分布の推定

例：単回帰モデル

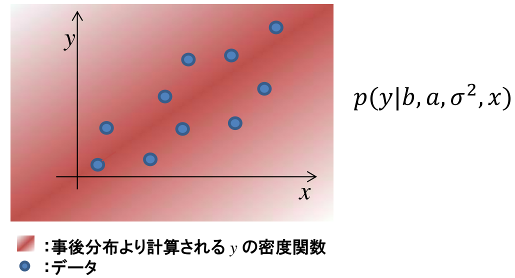
$$y=a+bx+e, e\sim N(0,\sigma^2)$$

- **分散** σ^2 **と切片** a **は既知で，傾き** b **のみの推定を考える**

【MAP推定】



【事後分布の推定】



- **当然ながら，事後分布を推定したほうが情報量は多い**

しかし，問題によって必要な情報は異なるので，
MAP**推定が有効なケースも多々ある。**

事後分布の信用区間 #1

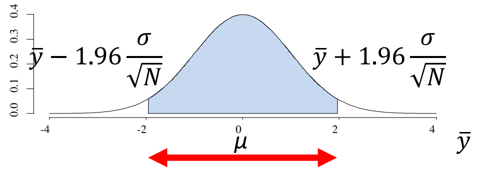
ベイズ統計学での区間推定に対応する区間（確信区間）

- **頻度主義的統計学での**95%**信頼区間**

  意味：サンプルサイズ N **の標本抽出を** M **回繰り返し，各回で区間を算出したとき，** $0.95\times M$ **回はパラメータの真の値を区間内に含んでいる（真のパラメータの値は定数）**

- **ベイズ統計学での**95%**信用区間**

  意味：パラメータの値がその区間内にある確率は95%**である（パラメータは確率変数）**



μ **の**95%**信頼区間** & 95%**信用区間**

正規分布を考えると信頼区間と信用区間の数学的表現は同一。
しかし，両者の意味・解釈は異なる

事後分布の信用区間 #2

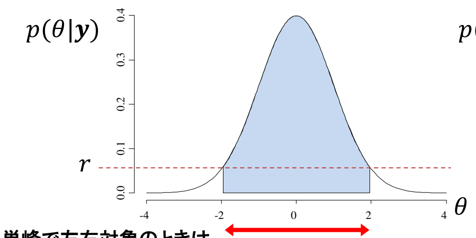
最高事後密度区間(HPDI: Highest Posterior Density Interval)

- **左右非対称の場合等の利用に適した区間**

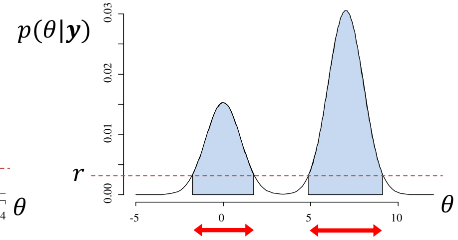
  ベイズ推定ではHPDI**が推定結果の解釈に良く用いられる**

- 95%HPDI**の定義：ある定数** r **と** $R=\{\theta|p(\theta|\boldsymbol{y})\ge r\}$ **で** $\int_R p(\theta|\boldsymbol{y})\,d\theta=0.95$ **となる区間**

  95%HPDI**は** ■ **の面積が**0.95 **となる** r **で決まる区間**



単峰で左右対象のときは信頼区間と一致



離散分布の最尤推定

ベルヌーイ分布の最尤推定

- 例：コインを 3 回投げて 3 回とも裏$(y=0)$が出た。このコインの表$(y=1)$が出る確率 θ（ベルヌーイ分布のパラメータ）を推定しなさい

確率関数：$\Pr(y;\theta)=\theta^{y}(1-\theta)^{1-y}$

尤度：$l(\theta|\boldsymbol{y})=\prod_{i=1}^{N}\theta^{y_i}(1-\theta)^{1-y_i}$

最尤推定量：$\hat{\theta}=\frac{1}{N}\sum_{i=1}^{N}y_i=\frac{1}{3}(0+0+0)=0$

$\therefore$ そのコインの表が出る確率はゼロである

サンプルサイズが小さい，または，生起確率の値がサンプルサイズと比べて極端に小さいとき，最尤推定の結果は常識や直感と合わない結果を導くことがある

ベルヌーイ分布のベイズ推定 #1

共役事前分布を用いたベルヌーイ分布の事後分布推定

- ベルヌーイ分布の共役事前分布はベータ分布

ベータ分布 $\mathrm{Beta}(\alpha,\beta)$

- 確率密度関数（連続型の確率分布）

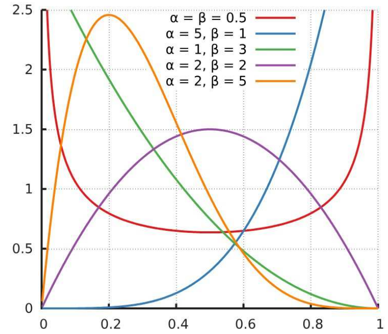
$f(x)=C(\alpha,\beta)x^{\alpha-1}(1-x)^{\beta-1}$

$(0\le x\le 1,)\ (\alpha,\beta>0)$

$C(\alpha,\beta)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$, $\Gamma(\alpha)$：ガンマ関数

$\Gamma(\alpha)=\int_0^\infty x^{\alpha-1}e^{-x}\,dx$, $\Gamma(N)=(N-1)!$

- パラメータ：α,β，期待値：$\frac{\alpha}{\alpha+\beta}$



Wikipedia「ベータ分布」

#メモ　$\alpha-1$ を成功回数，$\beta-1$ を失敗回数としたときのベルヌーイ確率 p の分布と考えるとわかりやすい。例えば，$\alpha=\beta=1$ のとき，その形は一様分布になる。試行回数0のときの確率 p の情報としてふさわしい

ベルヌーイ分布のベイズ推定 #2

ベルヌーイ分布の事後分布

- $y_i=1$ の出た回数を N_1, $y_i=0$ の出た回数を N_0, $z=\int p(\boldsymbol{y}|\alpha,\beta)p(\alpha,\beta)d\alpha d\beta$ とすると

$$P(\theta|\boldsymbol{y})=\frac{1}{z}\left\{\underbrace{\prod_{i=1}^{N}\theta^{y_i}(1-\theta)^{1-y_i}}_{\text{尤度}}\right\}\times\underbrace{C(\alpha,\beta)\theta^{\alpha-1}(1-\theta)^{\beta-1}}_{\text{事前分布}}$$

$$=\underbrace{C(N_1+\alpha,N_2+\beta)\theta^{N_1+\alpha-1}(1-\theta)^{N_0+\beta-1}}_{\text{事後分布}}$$

check!

事後分布は $\mathrm{Beta}(N_1+\alpha,N_2+\beta)$

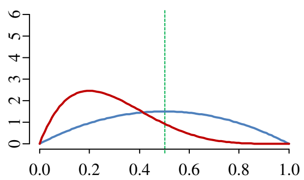
（パラメータが $N_1+\alpha$ と $N_0+\beta$ のベータ分布）

ベルヌーイ分布のベイズ推定 #2

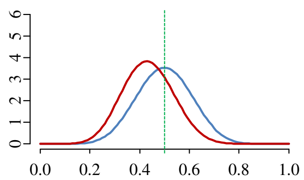
ベルヌーイ分布のベイズ推定

- 例：コインを 3 回投げて 3 回とも裏が出た。このコインの表の出る確率 p は $\mathrm{Beta}(2{,}2)$, $\mathrm{Beta}(10{,}10)$, $\mathrm{Beta}(25{,}25)$ に従うというそれぞれの確信に基づいて，確率 θ をベイズ推定しなさい

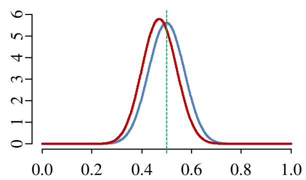
事前分布：$\mathrm{Beta}(2{,}2)$
事後分布：$P(\theta|\boldsymbol{y})$
$=Cp^{0+2-1}(1-p)^{3+2-1}$
$=\mathrm{Beta}(2{,}5)$



事前分布：$\mathrm{Beta}(10{,}10)$
事後分布：$P(\theta|\boldsymbol{y})$
$=Cp^{0+10-1}(1-p)^{3+10-1}$
$=\mathrm{Beta}(10{,}13)$



事前分布：$\mathrm{Beta}(25{,}25)$
事後分布：$P(\theta|\boldsymbol{y})$
$=Cp^{0+25-1}(1-p)^{3+25-1}$
$=\mathrm{Beta}(25{,}28)$



演習問題

ベルヌーイ分布の事後分布を導出しなさい

- **ただし, ここでは** $y_i = 1$ **の出た回数を** N_1, $y_i = 0$ **の出た回数を** N_0, $z = \int p(\boldsymbol{y}|\alpha, \beta)p(\alpha, \beta)d\alpha d\beta$ **とする**

$$P(\theta|\boldsymbol{y}) = \frac{1}{z}\left\{\prod_{i=1}^{N} \theta^{y_i}(1-\theta)^{1-y_i}\right\} \times C(\alpha, \beta)\theta^{\alpha-1}(1-\theta)^{\beta-1}$$

$$= C(N_1 + \alpha, N_2 + \beta)\theta^{N_1+\alpha-1}(1-\theta)^{N_0+\beta-1}$$

ヒント：

- **まともに** z **の積分を計算しては大変。** $\int_{-\infty}^{\infty} C(\alpha, \beta)\theta^{\alpha-1}(1-\theta)^{\beta-1}d\theta = 1$ **の関係を利用する**