情報認識
「ベイズ推定法（第9章）」

- 担当教員:　杉山　将（計算工学専攻）
- 居室:　W8E-505
- 電子メール:　sugi@cs.titech.ac.jp

これまでの復習

- 最大事後確率則: $\arg\max_{y} p(y \mid x)$
- ベイズの定理より $p(y \mid x) \propto \underbrace{p(x \mid y)}_{\text{条件付き確率}} \underbrace{p(y)}_{\text{事前確率}}$
- 事前確率は各カテゴリの標本の割合で推定

$$\hat{p}(y) = n_y / n$$

- 本講義の主題: 条件付き確率をうまく推定したい
- 簡単のため，条件付きでない確率密度関数 $p(x)$ を $\{x_i\}_{i=1}^{n}$ から推定する問題を考える.

確率密度関数の推定

- パラメトリック法：モデルのパラメータを推定
  - 最尤推定法，EMアルゴリズム，赤池の情報量規準
  - ベイズ推定法，最大事後確率法，経験ベイズ法
- ノンパラメトリック法：モデルを使わず直接確率密度関数を推定
  - カーネル密度推定法
  - 最近傍密度推定法

パラメトリック法

- θ :パラメータ
- Θ :パラメータの定義域
- パラメトリックモデル $q(x;\theta)$:有限次元のパラメータで記述された確率密度関数の族
- パラメトリック法:パラメトリックモデルを用いて確率密度関数を推定する方法

最尤推定法

- 尤度：訓練標本 $\{x_i\}_{i=1}^n$ がモデル $q(x;\theta)$ から生起する確率を θ の関数とみたもの：

$$L(\theta) = \prod_{i=1}^{n} q(x_i;\theta)$$

- 最尤推定法：尤度を最大，即ち，手元にある訓練標本が最も生起しやすいようにパラメータ値を決める方法

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \Theta} L(\theta)$$

パラメータの確率的取り扱い

- 最尤推定の枠組みでは，モデル $q(x;\theta)$ のパラメータ θ を決定論的な変数として扱った．
- もし，パラメータも確率変数とみなせば，次のような確率が定義できる．
  - 事前確率 $p(\theta)$
  - 事後確率 $p(\theta \mid X)$
  - 尤度 $p(X \mid \theta)$

$$X = \{x_1, x_2, \ldots, x_n\}$$

- これまではモデルを $q(x;\theta)$ と表記してきたが，ベイズ推定の枠組みでは $q(x \mid \theta)$ と表す．

ベイズ推定法

- ベイズ推定法(Bayesian inference): モデルをパラメータの事後確率に関して平均することによって推定する方法

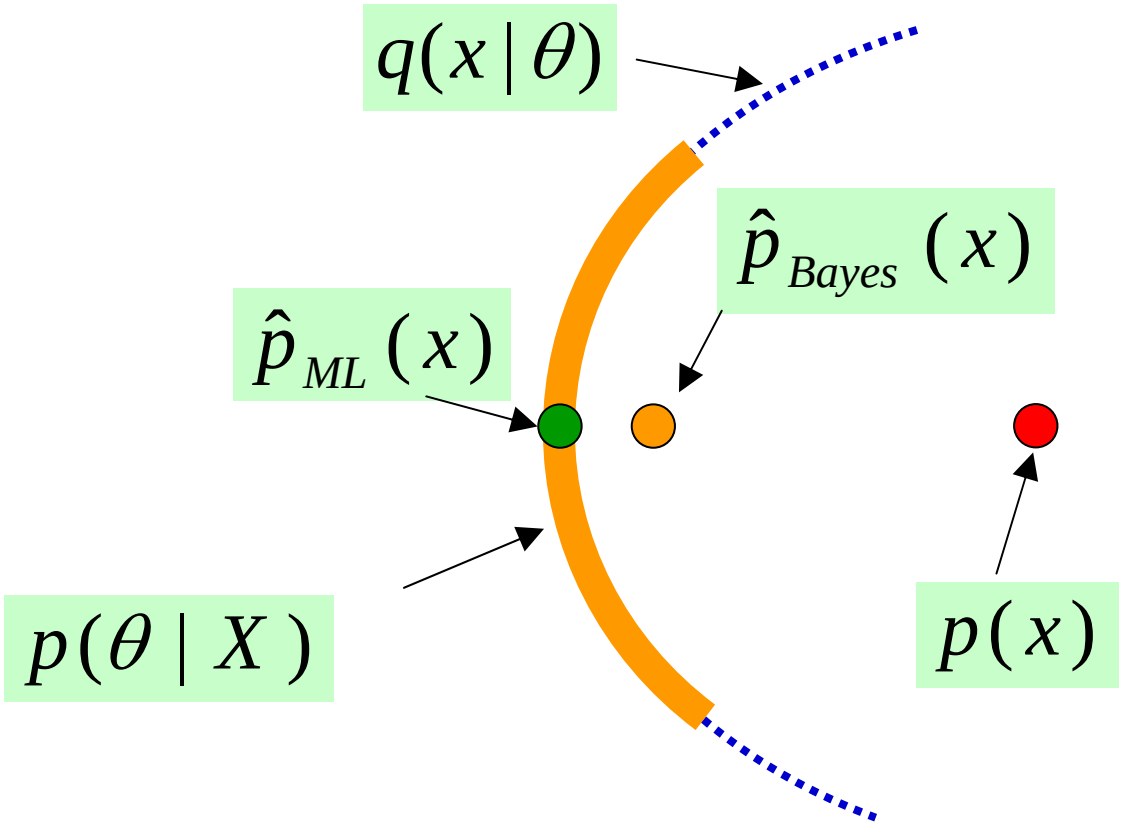
$$\hat{p}(x) = \int_{\Theta} q(x \mid \theta)\, p(\theta \mid X)\, d\theta$$

- 最尤推定: 1つの代表パラメータで推定
- ベイズ推定: 無数の確率密度関数の平均で推定

ベイズ推定法のイメージ

$$\hat{p}_{Bayes}(x) = \int_{\Theta} q(x \mid \theta)\, p(\theta \mid X)\, d\theta$$

- 一般に $\hat{p}_{Bayes}(x)$ はモデルに含まれない.



ベイズ推定法の計算

- 尤度は

$$p(X \mid \theta) = \prod_{i=1}^{n} q(x_i \mid \theta)$$

- 事後確率は，ベイズの定理を用いれば

$$p(\theta \mid X) = \frac{p(X \mid \theta)\, p(\theta)}{p(X)}$$

$$= \frac{p(X \mid \theta)\, p(\theta)}{\int_{\Theta} p(X \mid \theta')\, p(\theta')\, d\theta'}$$

$$= \frac{\prod_{i=1}^{n} q(x_i \mid \theta) p(\theta)}{\int_{\Theta} \prod_{i'=1}^{n} q(x_{i'} \mid \theta') p(\theta')\, d\theta'}$$

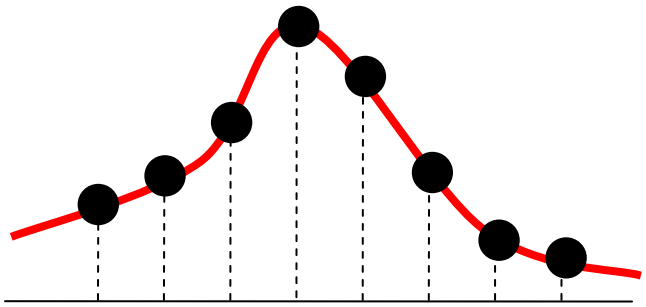
ベイズ推定法の計算（続き）

$$\hat{p}(x) = \int_{\Theta} q(x \mid \theta) \frac{\prod_{i=1}^{n} q(x_i \mid \theta) p(\theta)}{\int_{\Theta} \prod_{i'=1}^{n} q(x_{i'} \mid \theta') p(\theta') d\theta'} d\theta$$

- ベイズ推定法は，事前確率とモデルさえ与えられれば，原理的には計算できる．
- しかし，実際にはパラメータに関する積分を計算しなければならない．

ベイズ推定法の計算（続き）

- パラメータ空間 Θ の次元が低いとき，格子上の値を使えば，（台形公式などにより）簡単に近似計算できる.
- しかし，格子点の数はパラメータ空間の次元に指数的に比例するため，次元が高いとき，計算効率が非常に悪い.



最大事後確率推定法

- 最大事後確率推定法(maximum a posteriori probability estimation, MAP推定法):パラメータに関して積分するのをやめて，事後確率を最大にするパラメータ1点を用いる

$$\hat{p}(x) = q(x \mid \hat{\theta}_{MAP})$$

$$\hat{\theta}_{MAP} = \arg\max_{\theta} p(\theta \mid X)$$

- ベイズの定理より

$$\hat{\theta}_{MAP} = \arg\max_{\theta} \left[\sum_{i=1}^{n} \log q(x_i \mid \theta) + \log p(\theta) \right]$$

最大事後確率推定法と最尤推定法

- MAP推定法はベイズ推定の近似解法.
- しかし，代表パラメータ $\hat{\theta}_{MAP}$ で分布を推定するため，ベイズ推定よりはむしろ最尤推定に近い.
- MAP推定法は，対数尤度に対数事前確率を加えたものを最大にする.
- そのためMAP法は，罰則付き最尤推定法(penalized MLE)とも呼ばれる.

$$\hat{\theta}_{MAP} = \arg\max_{\theta} \left(\underbrace{\sum_{i=1}^{n} \log q(x_i \mid \theta)}_{\text{対数尤度}} + \underbrace{\log p(\theta)}_{\text{対数事前確率}} \right)$$

演習

- 以下の設定でMAP推定量と最尤推定量を求めよ.
  - モデル：分散1のガウスモデル

$$q(x \mid \mu) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2} \right)$$

  - 標本： $\{x_i\}_{i=1}^{n},\ x_i \in R$
  - 事前確率：ガウス分布

$$p(\mu) = \frac{1}{\sqrt{2\pi\beta^2}} \exp\left(-\frac{\mu^2}{2\beta^2} \right)$$

$\beta > 0$

（小さい $|\mu|$ の方が出やすい）

- MAP推定量と最尤推定量の関係を調べよ.

事前確率の役割

- 最尤推定法は訓練標本が少ないとき精度がよくないことがある.
- MAP推定法では事前確率の大きい方に解が引っ張られるので，訓練標本が少ない場合でも精度がよいことがある.

まとめ

- ベイズ推定法：パラメータも確率変数とみなして，パラメータの事後分布に関してモデルを平均する
- パラメータに関する積分の計算が大変
- 最大事後確率推定法（MAP法）：ベイズ推定法の積分を最大の点で近似する
- MAP法はベイズ推定よりもむしろ最尤推定に近い（罰則付き最尤推定法とも呼ばれる）

小レポート

- Octaveなどを用いた実験:
  - 平均0.5, 分散1の正規分布からn 個標本を生成せよ.
  - 演習で用いたモデル, 事前分布に対して, MAP推定法を用いて分布の平均を推定せよ.

$$\hat{\mu}_{MAP} = \frac{1}{n+\beta^{-2}} \sum_{i=1}^{n} x_i$$

  - 事前分布の β の値を変化させ, 推定結果がどのように変化するかを調べよ. また, 最尤推定法($\beta = \infty$)とも比較せよ.
  - 真の平均, 標本数などを変化させたとき, 実験結果がどのように変化するかを考察せよ

Octaveのサンプルプログラム

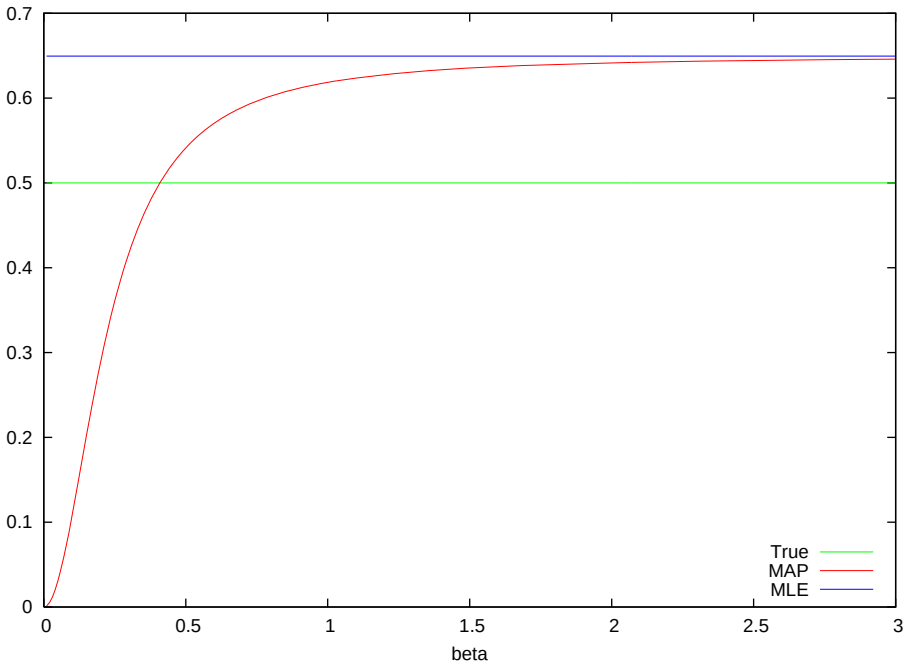
ex8.m

```
clear all
n=20; mu=0.5; sigma=1;
xx=sigma*randn(n,1)+mu;
betas=[0.01:0.01:3];

mu_MLE=mean(xx); sigma_MLE=std(xx,1);
for i=1: length(betas);
  mu_MAP(i)=sum(xx)/(n+betas(i).^(-2));
end

figure(1); clf; hold on;
plot(betas,mu*ones(size(betas)),‘g-')
plot(betas,mu_MAP,'r-')
plot(betas,mu_MLE*ones(size(betas)),'b-')
xlabel('beta'); legend('True','MAP','MLE',4)
print -deps MAP.eps
```

実行例



試験について

- 2月7日（月）10時40分～12時10分
- S222講義室
- 試験内容：
  - 専門用語の英単語（日本語→英語）
  - 自由記述問題（以下より2問選択）
    - 最尤推定法について
    - ノンパラメトリック法について
    - ベイズ推定法について
  - 自分が考えたパターン認識の例について
- 教科書，ノートの持ち込みは不可．