情報認識
「パラメトリック法：ベイズ推定」

- 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

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　$p(\theta \mid x_1, x_2, \ldots, x_n)$
  - 尤度　$p(x_1, x_2, \ldots, x_n \mid \theta)$

ベイズ推定法

- **ベイズ推定法(Bayes estimation method)**：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

$$\hat{p}(x) = \int_{\Theta} q(x;\theta)\, p(\theta \mid x_1, x_2, \ldots, x_n)\, d\theta$$

- 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

ベイズ推定法の計算

■ 尤度は

$$p(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} q(x_i;\theta)$$

■ 事後確率は, ベイズの定理を用いれば

$$p(\theta \mid x_1, x_2, \ldots, x_n) = \frac{p(x_1, x_2, \ldots, x_n \mid \theta)\, p(\theta)}{p(x_1, x_2, \ldots, x_n)}$$

$$= \frac{p(x_1, x_2, \ldots, x_n \mid \theta)\, p(\theta)}{\int_{\Theta} p(x_1, x_2, \ldots, x_n \mid \theta')\, p(\theta') d\theta'}$$

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

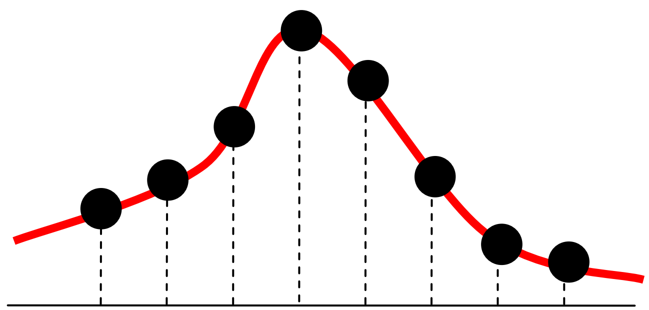
ベイズ推定法の計算 (続き)

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

- 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, (台形公式やシンプソンの公式により) 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.



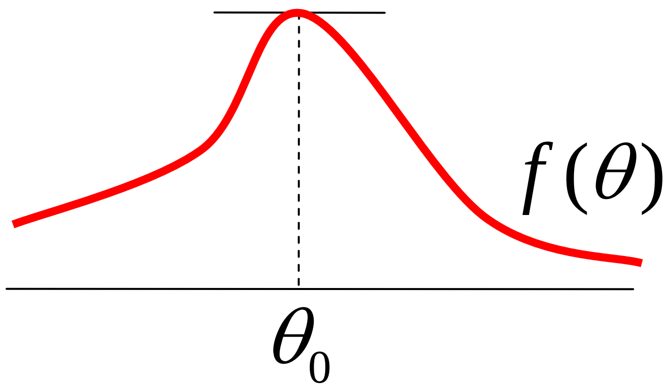
積分の近似

- 以下では， $f(\theta)$ の積分 $\int_{\Theta} f(\theta) d\theta$ を近似する問題を考える．
- $\theta_0 = \arg\max_{\theta \in \Theta} f(\theta)$ は最大値なので次式を満たす

$$\left. \frac{\partial}{\partial \theta_i} f(\theta) \right|_{\theta = \theta_0} = 0$$

- $\log f(\theta)$ に対しても同様に

$$\left. \frac{\partial}{\partial \theta_i} \log f(\theta) \right|_{\theta = \theta_0} = 0$$



テイラー展開

■ **テイラー展開**(Taylor expansion)：

$$f(\theta) = f(\theta_0) + (\theta - \theta_0) f'(\theta_0) + \tfrac{1}{2}(\theta - \theta_0)^2 f''(\theta_0) + \cdots$$

(θ の1次元のとき)

■ **テイラー展開を適当な次数で打ち切れば，関数 $f(\theta)$ を多項式で近似できる．**

■ **θが多次元のとき，**

$$f(\theta) = f(\theta_0) + (\theta - \theta_0)^T h + \tfrac{1}{2}(\theta - \theta_0)^T H (\theta - \theta_0) + \cdots$$

$$h_i = \frac{\partial}{\partial \theta_i} f(\theta)\bigg|_{\theta=\theta_0} \qquad H_{i,j} = \frac{\partial^2}{\partial \theta_i \partial \theta_j} f(\theta)\bigg|_{\theta=\theta_0}$$

：**ヘシアン行列** (Hessian matrix)

積分の近似 (続き)

- $\log f(\theta)$ を θ_0 の周りで2次のテーラー展開

$$\log f(\theta) \approx \underbrace{\log f(\theta_0) + 0 + \tfrac{1}{2}(\theta-\theta_0)^T H(\theta-\theta_0)}_{\log \tilde{f}(\theta)}$$

$$H_{i,j} = \frac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(\theta) \bigg|_{\theta=\theta_0}$$

- 指数をとれば,

$$f(\theta) \approx \tilde{f}(\theta) = f(\theta_0)\exp\left(\tfrac{1}{2}(\theta-\theta_0)^T H(\theta-\theta_0)\right)$$

ラプラス近似

■正規分布の確率密度関数の積分は1：

$$\int_D p(x)dx = 1$$

$$p(x) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu) \right)$$

■これより，

$$\int_\Theta \exp\left(-\tfrac{1}{2}(\theta-\theta_0)^T (-H)(\theta-\theta_0) \right) d\theta = \sqrt{\frac{(2\pi)^{\dim\theta}}{|-H|}}$$

■積分のラプラス近似(Laplace approximation)：

$$\int_\Theta f(\theta)d\theta \approx \int_\Theta \tilde{f}(\theta)d\theta = f(\theta_0)\sqrt{\frac{(2\pi)^{\dim\theta}}{|-H|}}$$

ラプラス近似 (続き)

- $\log f(\theta)$ を2次関数で近似することは, $f(\theta)$ を (正規化されていない) ガウス関数で近似することに対応.
- そのため, ラプラス近似はガウス近似 (Gaussian approximation) とも呼ばれる.
- $f(\theta)$ の形状がガウス分布の形状に近いとき, ラプラス近似は精度がよい.

小レポート

- $f(\theta) = p(\theta;1) + p(\theta;2)$

$$p(\theta;\sigma) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{\theta^2}{2\sigma^2} \right)$$

- $Z = \int_{-\infty}^{\infty} f(\theta) d\theta$ **の値をラプラス近似で求めよ（真の値は2である）．**