2.3.5 逐次推定

茨城大学工学部情報工学科
05T4072R
相原　功昌

逐次推定

- 逐次:
  順を追って次々に物事がなされるさま
- 逐次推定:
  データ点を1つずつ処理し、処理したデータは廃棄して推定する方法

→全てのデータ点を一度に処理できない大規模なデータ集合、オンラインな応用分野で重要

平均の最尤推定量について

- 平均の最尤推定量 μ_{ML} の場合
  N個の観測値に基づいて推定した結果: $\mu_{ML}^{(N)}$
  最後のデータ点 X_N がどれくらい影響したかを調べる
  →式(2.126)

平均の最尤推定量について（2）

$$\mu_{ML}^{(N)} = \frac{1}{N}\sum_{n=1}^{N} \mathrm{x}_n$$

$$= \frac{1}{N}\mathrm{x}_N + \frac{1}{N}\sum_{n=1}^{N-1} \mathrm{x}_n$$

$$= \frac{1}{N}\mathrm{x}_N + \frac{N-1}{N}\mu_{ML}^{(N-1)}$$

$$= \frac{1}{N}\mathrm{x}_N + \frac{N}{N}\mu_{ML}^{(N-1)} - \frac{1}{N}\mu_{ML}^{(N-1)}$$

$$= \mu_{ML}^{(N-1)} + \frac{1}{N}\left(\mathrm{x}_N - \mu_{ML}^{(N-1)}\right) \quad \cdots(2.126)$$

$$\mu_{ML}^{(N-1)} = \frac{1}{N-1}\sum_{n=1}^{N-1} \mathrm{x}_n$$

$$(N-1)\mu_{ML}^{(N-1)} = \sum_{n=1}^{N-1} \mathrm{x}_n$$

$$\frac{(N-1)}{N}\mu_{ML}^{(N-1)} = \frac{1}{N}\sum_{n=1}^{N-1} \mathrm{x}_n$$

平均の最尤推定量について（3）

- N-1個のデータ点を観測した時点でのμの推定値：$\mu_{ML}^{(N-1)}$
- 誤差信号：$(\mathrm{x}_N - \mu_{ML}^{(N-1)})$
- データ点x_Nを観測すると、1/Nに比例する小さな量だけ、誤差信号の方へ$\mu_{ML}^{(N-1)}$を移動させて修正した値を$\mu_{ML}^{(N)}$としている
- Nが増えるにつれて、後続のデータ点からの影響は小さくなる

汎用的な逐次学習の定式化

- この方法で逐次アルゴリズムを導出する

→いつもできるとは限らない

汎用的な逐次学習の定式化が必要となる

→Robbins-Monroアルゴリズム

Robbins-Monroアルゴリズム（1）

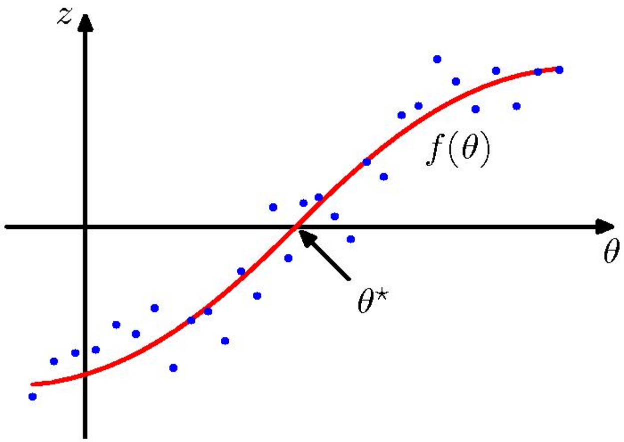

図2.10

- 確率変数のパラメータ: θ
- θ に依存した確率変数: z
- 回帰関数 $f(\theta)=E[z\,|\,\theta]$を定義

$$
\begin{aligned}
f(\theta) &\equiv E[z\,|\,\theta] \\
&= \int zp(z\,|\,\theta)dz \qquad \cdots(2.127)
\end{aligned}
$$

- このとき、$f(\theta)=0$ の根 θ^* を逐次的に求めるためのアルゴリズム

Robbins-Monroアルゴリズム(2)

- 仮定
  - Z の条件付分散は有限である
    $E[(z-f)^2 \mid \theta] < \infty$
  - $\theta > \theta^*$では$f(\theta) > 0$
  - $\theta < \theta^*$では$f(\theta) < 0$
- 定義
  - N-1個のデータを観測した後の推定値: $\theta^{(N-1)}$
  - このパラメータの下でのN個目の Z の観測値: $z\left(\theta^{(N-1)}\right)$
  - このときのN回目のパラメータ:

$$\theta^{N} = \theta^{N-1} - a_{N-1} z\left(\theta^{(N-1)}\right) \quad \cdots (2.129)$$

Robbins-Monroアルゴリズム（3）

- 係数 a_N が
  - $\lim_{N \to \infty} a_N = 0$ ・・・(2.130)
  - $\sum_{N=1}^{\infty} a_N = \infty$ ・・・(2.131)
  - $\sum_{N=1}^{\infty} a_N^2 < \infty$ ・・・(2.132)

を満たすなら、式(2.129)で与えられる θ^N は目標の根に確率1で収束する

Robbins-Monroアルゴリズム（4）

- 式(2.130)
  →この過程が極限値に収束できるように、解の逐次的な修正量を減らすことを保証
- 式(2.131)
  →アルゴリズムが根以外に速すぎる収束をしないことを保証
- 式(2.131)
  →蓄積されたノイズの分散を有限に抑え、収束を阻害しないことを保証

一般的な最尤推定問題（１）

- Robbins-Monroアルゴリズムを用いて、一般的な最尤推定問題を逐次的に解く
- 最尤推定解：θ_{ML}
  →負の対数尤度関数の停留点(誤差の最小)
  →式(2.133)を満たす

$$-\frac{\partial}{\partial\theta}\left\{\frac{1}{N}\sum_{n=1}^{N}\ln p(x_n \mid \theta)\right\}\Bigg|_{\theta_{ML}} = 0 \quad \cdots (2.133)$$

一般的な最尤推定問題（2）

$$-\frac{\partial}{\partial\theta}\left\{\frac{1}{N}\sum_{n=1}^{N}\ln p(x_n|\theta)\right\}\Bigg|_{\theta_{ML}} = 0$$

微分と総和の演算を交換すると

$$-\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\theta}\ln p(x_n|\theta)\Bigg|_{\theta_{ML}} = 0$$

期待値の近似を厳密にするために

$N\to\infty$の極限を考えると

$$-\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N}\frac{\partial}{\partial\theta}\ln p(x_n|\theta) = E_x\left[-\frac{\partial}{\partial\theta}\ln p(x|\theta)\right] \quad \cdots (2.134)$$

一般的な最尤推定問題(3)

- 最尤推定解を求める

＝回帰関数の根を求める

- Robbins-Monroアルゴリズムを適用

  →式(2.135)

$$\theta^{(N)} = \theta^{(N-1)} - a_{N-1} \frac{\partial}{\partial \theta^{(N-1)}} \left[-\ln p(x_N \mid \theta^{(N-1)}) \right] \quad \cdots (2.135)$$

ガウス分布の平均の逐次推定（1）

- パラメータ: $\theta^{(N)}$

→ガウス分布の平均の推定量: $\mu_{ML}^{(N)}$

- 確率変数: z

$$z = -\frac{\partial}{\partial \mu_{ML}} \ln p(x \mid \mu_{ML}, \sigma^2)$$

$$= -\frac{1}{\sigma^2}(x - \mu_{ML}) \qquad \cdots (2.136)$$

ガウス分布の平均の逐次推定（2）

- 式(2.136)を式(2.135)に代入

$$\mu_{ML}^{(N)} = \mu_{ML}^{(N-1)} - a_{N-1}\{-\frac{1}{\sigma^2}(x-\mu_{ML})\}$$

- 係数a_Nを $a_N = \sigma^2 / N$ とすると
  式(2.126)の形式になる

- 多変量の場合も同様に、同じ手法が適用できる