2005年6月25日

# 赤池情報量規準
# Akaike Information Criterion

矢野浩一
総合研究大学院大学
統計科学専攻
博士課程3年

# ［テーマ］赤池情報量規準

- 赤池弘次によって1973年に提唱されたモデルの当てはまりの良さを表す量。Akaike Informatio Criteron (AIC)という
- 以下のように定式化されている（AICが小さいほどモデルの当てはまりが良い）

$$AIC = -2\log l(\Theta) + 2p$$

ここで $l(\Theta)$ は尤度関数、

$p$ はモデルで推定されるべきパラメータの数

# Kullback-Leiber情報量（1）

- Kullbackによって1952年に提唱された情報量

$$K = \int p(x) \log \frac{p(x)}{q(x)} dx \geqslant 0$$

$$K = \underline{\int p(x) \log p(x) dx} - \boxed{\int p(x) \log q(x) dx}$$

基本的にはこちらは不明

こちらの計算に注力

$p(x)$ :真のモデル

$q(x)$ :統計モデル

# Kullback-Leiber情報量（2）

- イメージ

KL情報量＝真のモデル－統計モデル ≧０

しかし、世間一般では
真のモデルは未知
（なことが多い）

この項目の最小化をすれば、
**KL**情報量は小さくなる（はず）

# 最尤法（1）

以下の関数を対数尤度関数といい

$$\frac{1}{N}\sum \log q(x;\theta)$$

この対数尤度関数をパラメータ $\theta$ に対して最大化する。この方法を最尤法という

<u>大数の法則</u>からサンプル数が大きくなると

$$\frac{1}{N}\sum \log q(x) \to \int p(x)\log q(x)\,dx$$

# 最尤法（2）

- 最尤推定（尤度の最大化）とはつまり、何をやっているかと言うと・・・・

↓

## KL情報量を最小化すること

# 最尤法の問題点

- パラメータ数が多くなると勝手に対数尤度関数が大きくなる

$$\frac{1}{N}\sum \log q\left(x\,;\theta_{1,}\theta_{2,}\cdots\theta_{p}\right)$$

↓

推定値の当てはまりのよさと関係なく大きな値になる傾向

↓

そのためにパラメータ数pだけ対数尤度関数を補正すればよい。つまり $\log l\left(\Theta\right)-p$

# 最小赤池情報量基準

- 結局、AICは歴史的事情のため（＊）、以下のような形で定式化された

$$AIC = -2\log l(\Theta) + 2p$$

マイナスがついているので、AICを「最小化」したものが一番当てはまりがよいことになる

（＊）尤度比検定との兼ね合いから-2を掛けたと赤池（１９８２）にある。

# その他の情報量規準

- 必ずしも万能ではないAIC
  - しかし、AICを使えば問題ないことが多い
  - AICにどのような制限があるのかは小西・北川（2004）を参照
- どうしてもAICでダメな場合はその他の情報量規準を使えば解決することもある
  - TIC, BIC, RIC, GIC, EIC, etc.
  - 詳しくは小西・北川（2004）を参照

# 参考文献


- 赤池弘次（1982）「統計とエントロピー」数学セミナー
- 尾形良彦（2004）「赤池情報量規準とベイズモデル」講義資料
- 小西貞則・北川源四郎（2004）「情報量規準」朝倉書店
- 坂元慶行・石黒真樹夫・北川源四郎（1983）「情報量統計学」 共立出版
- Shibata, Ritei, (1989), "Statistical Aspects of Model Selection," From Data to Model (J. C. Willems ed.), Springer-Verlag.