

---

# 情報理論による実証分析と 測定論の橋渡し 実証分析の「意味」早わかり

---

福井義高

青山学院大学大学院国際マネジメント研究科

平成27年9月7日

# 今日、何を伝えたいか

- 実証分析で行っていることの意味を考える
- Kullback-Leibler情報量からVuong検定まで統一的に理解することを目指す
  - 論文第5～12節の部分

# Kullback-Leibler (KL) 情報量

- モデル(仮説)が真の確率分布とどれだけ近いかを測る量

$$\begin{aligned} D &= \int_{-\infty}^{\infty} p(x) [\log p(x) - \log f(x)] dx \\ &= E_p [\log p(x)] - E_p [\log f(x)] \geq 0 \end{aligned}$$

- 実証研究者が目指すべきは、真の(条件付)確率分布とモデルの「距離」であるKL情報量をゼロにすること

# 最尤法の考え方

$$y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2)$$

$$f(y_i | x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - x_i' \beta)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\varepsilon_i^2}{2\sigma^2}}$$

## ■ 尤度関数／対数尤度関数の最大化

$$L(y | x; \theta) = f(y_1 | x_1; \theta) \times \dots \times f(y_n | x_n; \theta)$$

$$\begin{aligned} l(\theta) &= \log L(\theta) = \log f_1(\theta) + \dots + \log f_n(\theta) \\ &= l_1(\theta) + \dots + l_n(\theta) \end{aligned}$$

# 最尤法とKL情報量(1)

- 我々の目的は第二項であるモデルの平均対数尤度  $E_p[\log f(x)]$  の最大化に帰着
- 我々は真の分布を知らないなので、真の分布  $p$  をデータに基づく経験分布  $f$  で置換え

$$E_f[l_i(\hat{\theta})] = \frac{1}{n} \sum l_i(\hat{\theta}) = \frac{l(\hat{\theta})}{n}$$

最尤法推定値 = KL情報量第二項推定値  $\times n$

- 最尤法とは真の分布との距離をKL情報量の意味で最小化する統計手法

## 最尤法とKL情報量(2)

- 最尤推定値は平均対数尤度推定値としてはバイアスあり
- バイアス推定値を「ペナルティ」とし対数尤度から引いて、マイナス2をかけたのがTIC (竹内情報量規準)

$$TIC = -2 \left[ l(\hat{\theta}) - tr(\hat{I}\hat{J}^{-1}) \right] = -2l(\hat{\theta}) + 2tr(\hat{I}\hat{J}^{-1})$$

- モデル構築すなわち説明変数選択にあたって規準となるスカラー指標

# TIC簡略版としてのAIC

- バイアス推定値は標本に大きく依存
- モデルが真の分布に「十分」近ければ、
$$TIC = -2l(\hat{\theta}) + 2tr(\hat{I}\hat{J}^{-1}) \simeq -2l(\hat{\theta}) + 2(k+1)$$
- このバイアス推定不要かつパラメータ数え上げという「機械的」作業でモデル選択を可能にしたのがAIC

# AICへの批判(1)

- AICは  $\log s^2 + \frac{2k}{n}$  の大小で選択
- 自由度修正済  $R^2$ は  $\log s^2 + \frac{k}{n}$  の大小で選択
- 帰無仮説モデル  $M_0$  とネストするモデル  $M_A$   
$$y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_{k_0} x_{k_0 i} + \varepsilon_i$$
$$y_i = \beta_1 + \beta_2 x_{2i} + \dots + \beta_{k_0} x_{k_0 i} + \dots + \beta_{k_A} x_{k_A i} + \varepsilon_i$$
- 対数尤度差  $\times 2$  が漸近的に自由度  $k_A - k_0$  のカイ二乗分布に従う



## AICへの批判(2)

- AICは真のモデルを選ぶ確率が1に収束するという意味での一貫性を持たない
- 一方、 $\log s^2 + \frac{\log n \times k}{n}$  の大小で選択するBICは一貫性を持つ
- AICよりBICが好ましい？

# AIC批判への反批判(1)

- 帰無仮説モデルがネストされている場合、通常の仮説検定とは、帰無仮説が一定の有為水準で棄却された(なかった)場合にモデル $M_A$ ( $M_0$ )を選ぶモデル選択問題
  - AICは仮説検定と同じ統計量(対数尤度差)を異なる閾値で使っている
  - AICは有意水準設定という難問にKL情報量概念に基づいて一定の答えを与えた

# AIC批判への反批判(2)

モデリングの目的は「よい」モデルを求めることであって、「真の」モデルを求めることではない．．．統計的モデルが複雑なシステムをある目的に沿って近似したものであることを思い起こせば、真の次数を推定するという問題設定自体が適当でない．．．モデルは複雑な現象の近似であるという立場からあえていえば、真の次数は無限大

たとえ、有限の真の次数が存在するものと仮定しても、よいモデルの次数が真の次数と一致するとは限らない．．．データ数が少ない場合には、推定されるパラメータの不安定さを考えると、より低次のモデルを用いた方がより高い予測精度が得られる可能性がある（北川・小西 2004）

# AICと仮説検定の融合(1)

- AICはネストされたモデル間の比較にしか使えない仮説検定の「弱点」を解決
- ノンネスト・モデル間のAICを比較した場合、その差がどれほど意味のあるものなのか
- こうした問題意識のもと、Linhart (1988)、Kishino and Hasegawa (1989)及びVuong (1989) はノンネスト・モデル間の対数尤度差が漸近的に正規分布に従うことを利用した検定を提案

## AICと仮説検定の融合(2)

- ノンネスト・モデルを比較する場合、どちらにも帰無仮説という「優先的」地位を与えず、両方とも「間違った」モデルであり、真のモデルとの距離が同じか否かという観点で検定
- 対数尤度推定値の差を標準偏差推定値で割った統計量  $\Delta l / \sqrt{\widehat{\text{var}}(\Delta l)}$  が漸近的に分散1の正規分布に従うことを利用
  - 推定にバイアス修正を行わないのが Vuong検定

# 会計実証分析におけるAICとVuong検定(1)

- AICは伝統的な仮説検定と対立するものではなく、むしろモデル選択と仮説検定を統一的に理解するうえでも有用な概念
  - Vuong検定もこの文脈で理解可能
- しかし、(被説明)変数の変動を予測あるいは「説明」することがモデル構築の目的ならば、「フルモデルだけを考えて、必要に応じて回帰係数が不安定にならないような工夫... をすれば十分」(下平 2007)

## 会計実証分析におけるAICとVuong検定(2)

- ノンネスト・モデルが互いに排他的な仮説を反映している場合、「当てはまり」が良いからといって、フルモデルは正当化できない
- しかし、フルモデルを明確に否定する理論的根拠を欠く、データの情報価値の大小を議論の焦点とする会計実証分析において、ノンネスト・モデル選択の有用性は限定的
  - 純利益と包括利益のどちらがmore relevantか等、二者択一の問題設定自体が疑問