

科学研究費助成事業 研究成果報告書

平成 28 年 10 月 31 日現在

機関番号：17102

研究種目：挑戦的萌芽研究

研究期間：2013～2015

課題番号：25540013

研究課題名(和文) 規準化した情報量規準による統計モデルの絶対評価

研究課題名(英文) Evaluation of stochastic models via regularized information criteria

研究代表者

西井 龍映(Nishii, Ryuei)

九州大学・マス・フォア・インダストリ研究所・教授

研究者番号：40127684

交付決定額(研究期間全体)：(直接経費) 1,800,000円

研究成果の概要(和文)：一般の統計モデル評価のための情報量基準 AIC は、統計学やその応用分野に大きな貢献を果たしてきた。ただ AIC は期待対数尤度の推定値であるため、回帰分析における決定係数のように絶対評価値ではない。本研究では統計モデルの絶対評価値を得るため、評価対象モデルと最も簡単なモデルのそれぞれの AIC の差に基づく AIC 決定係数を定義した。また分散不均一の場合の決定係数をヘテロ決定係数を定義した。さらに時空間データに基づく回帰モデルについて、2段階母数推定およびそのモデル評価指標を提案した。

研究成果の概要(英文)：The information criterion AIC proposed for evaluations of statistical models has contributed greatly to statistics and the application fields. AIC which is an unbiased estimate of the expected log likelihood, however, is not an absolute scale like the coefficient of determination in regression analysis. In this research, AIC-coefficient of determination is introduced based on the difference of AIC's of a current model and the simplest model for driving an absolute scale for the model evaluation. In addition, the heteroscedastic coefficient of determination is introduced in regression analysis with heteroscedasticity. Furthermore, a two stage procedure is introduced for model estimation and adaptive information criteria of regression models based on spatio-temporal data.

研究分野：統計モデリング

キーワード：モデル評価 カルバック ライブラー情報量 決定係数 期待対数尤度 情報量基準

1. 研究開始当初の背景

回帰分析においてモデルの良し悪しを評価するために一般的に用いられる基準は決定係数 R^2 である。決定係数は $0 < R^2 < 1$ を満たす規準化された量であるため、異なるデータであっても決定係数が同じであれば目的変数の予測値と実測値の線形関係の程度がわかる絶対量となっている。ただ決定係数は回帰モデルに説明変数を追加すると単調に大きくなるので、説明変数のモデル選択には自由度調整済み決定係数が用いられる。これは漸近的には AIC と同等なモデル選択基準となる。

一般の統計モデルの評価には、日本の研究者から情報量基準 AIC が提案され、統計学やその応用分野に大きな貢献を果たしてきたことは周知の事実である。それは負の期待対数尤度の不偏推定量として提案された。本研究では統計モデルを既存の情報基準を用いて、数値比較が容易な評価値を提案したい。

2. 研究の目的

Kullback-Leibler (KL) 情報量は期待対数尤度と真のモデルのエントロピーとの差であたえられる。AIC は前者だけの不偏推定量として提案されているため、KL 情報量を持っている非負性が失われている。またモデル同士を比較する相対量であり絶対量ではない。つまり、2 つのデータセットの AIC が共に 100 となっても、それぞれのモデルの予測能力が近いことを意味しない。ここではモデルの AIC 等の評価基準に基づいて、絶対評価が可能なモデル評価指標を導出することを目的とする。また同時に統計モデルの適切な評価指標の導出も検討する。

3. 研究の方法

決定係数は独立な等分散の誤差を持つ回帰モデルを前提としている。平均に関する回帰モデル、および平均が定数という**最も簡単なモデル**を考え、2 つのモデルによる推定分散の比で定義されている。この見方を一般化し、「評価対象のモデルは**最も簡単なモデルからどれだけ改善したか**」を評価する値として、決定係数や既存の情報量基準を用いて評価指標を提案する。また時空間回帰モデルにおいては新しい推定法および評価指標を提案する。

4. 研究成果

1) 情報量基準に基づく決定係数の提案
〔雑誌論文〕Nishii (2014) において、次の情報量基準に基づく決定係数を提案した。

一般の統計モデルにおいて、評価したいモデルと最も簡単なモデルの情報量基準のそれぞれの値、例えば AIC と AIC_0 を用いて決定係数 R_{AIC} を定義する。

$$R_{AIC} = 1 - \exp(AIC/n - AIC_0/n)$$

ただし n は標本数を表す。この評価指標 (AIC—決定係数と呼ぶ) は両モデルにおける期待対数尤度の不偏推定量の差として与えられている。同様に R_{BIC} 等も定義される。同論文では回帰分析における決定係数、AIC—決定係数、BIC—決定係数の間の不等式も得た。

AIC—決定係数の数値が 1 に近いと良いモデルと評価され、回帰モデルでは評価しやすい値を与える。ただ最も簡単なモデルが、すでにある程度すぐれている場合は、提案モデルの相対的良さが見えにくい基準であることがわかった。例えば分散についての一般化線形モデルで、分散が一定であるというモデルを最も簡単なモデルとした場合である (同論文の数値例参照)。この点は今後の改善点である。

2) 分散不均一の場合の回帰モデルの評価

決定係数は誤差分散が一定の回帰モデルにおける推定分散の比と見ることができ。〔雑誌論文〕Qin and Nishii (2015) では、これを分散不均一の場合に一般化した。すなわち個別の誤差分散がそれぞれ推定できるケースを考える。このとき、分散不均一の場合の決定係数を、モデルに基づく平均の予測値と実測値の差の 2 乗和に対し、推定分散の逆数による重みを付けて和を取る**ヘテロ決定係数**を定義した。これは評価値が 0 以上 1 以下の値をとるため、わかりやすい指標となっている。この提案指標は誤差分散が一般化線形モデルで与えられるケースを考察した研究動機に基づく。同論文の実データ解析によると、BIC の評価が決定係数よりヘテロ決定係数に近いことがわかった。

3) 時空間データの回帰モデルの推定法とモデル評価法

時空間で相関を持つデータに対する回帰モデルを考える。従来手法は (a) 誤差に時空間の相関を取り入れた時空間モデルを考え、(b) 最尤法により回帰係数、空間相関、分散等の母数推定を行い、(c) 情報量基準でモデルを評価する、ことが通常のステップである。ただ ステップ (b) のためには逐次的手法が必要であり、必ずしも安定的に推定できるわけではない。そこで時空間回帰モデルに対して次の 2 段階母数推定およびそのモデル評価指標を提案した。

- (1) データが時空間に対して独立であると仮定した回帰モデル (独立モデル) により回帰係数と分散の推定を行う。
- (2) 各標本における目的変数を独立モデルで予測し、実際は時空間で依存すると考えられる複数の予測値で実測値に回帰する。
- (3) (1), (2) で用いた全ての母数の数により AIC 等で時空間モデルを評価する。

上記のステップ (1), (2)における推定値, (3) の評価値はすべて陽な形で瞬時に求められる。これは同時方程式モデルのように逐次推定を繰り返す必要がある手法に比べれば格段の違いである。また実データに應用して母数推定を行った結果, 時空間モデルを最尤法により地道に求めた母数推定結果と比べても遜色がない。また適応的にモデル自由度を定義しているが, 数値例では合理的な結果を得ている。この結果は現在発表準備中である。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計7件)

P. Qin and R. Nishii (2015). Statistical prediction of Dst index by solar wind data and t-distributions, IEEE Transactions on Plasma Sciences, 43, 11, 3908-3915, DOI:10.1109/TPS.2015.248566.

R. Nishii and S. Tanaka (2015). Unified modeling based on SVM and SVR for prediction of forest area ratio by human population density and relief energy, Proc. of IEEE IGARSS 2015, DOI:10.1109/IGARSS.2015.7326331.

S. Tanaka and R. Nishii (2015). Incorporation of gridded data into the analysis of remotely-sensed images: Basic quantitative strategy to analyze deforestation by population growth, Proc. of IEEE IGARSS 2015, DOI:10.1109/IGARSS.2015.7326332.

R. Nishii (2014). Regression analysis and its development. A Mathematical Approach to Research Problems of Science and Technology- Theoretical Basis and Developments in Mathematical Modeling. Mathematics for Industry 5, R. Nishii et al. (eds.), Springer, 249-262.

P. Qin, T. Yamasaki and R. Nishii (2014). Statistical detection of the influence of solar activities to weak earthquakes, Pacific Journal of Math-for-Industry, 6:6.

R. Nishii and S. Tanaka (2013). Modeling and inference of forest coverage ratio using zero-one inflated distributions with spatial dependence. Environmental and Ecological Statistics 20(2), 315-336.

小平剛央, 中本尊元, 小池真人, 天野浩平, 西井龍映, 秦攀 (2013). 逐次実験計画法による車体構造の複合領域最適化手法. 自動車技

術会論文集 44(2), 535-541.

[学会発表](計13件)

江田智尊, 松井秀俊, 西井龍映, 持田恵一, 恩田義彦, 櫻井哲也, 吉田拓広, 植物バイオマス増産のための遺伝子データ解析, 情報・統計科学シンポジウム, 統計科学研究会, 福岡市, Dec. 4, 2015.

S. Tanaka and R. Nishii, Review on the methodology to explore functional forms of deforestation, 共同研究集会, 統計数理研究所, 立川市, Nov. 6, 2015.

R. Nishii and S. Tanaka, Unified modeling based on SVM and SVR for prediction of forest area ratio by human population density and relief energy, IEEE International Geoscience and Remote Sensing Symposium 2015, Milan, Italy, July 29, 2015.

S. Tanaka and R. Nishii, Incorporation of gridded data into the analysis of remotely-sensed images: Basic quantitative strategy to analyze deforestation by population growth, IEEE International Geoscience and Remote Sensing Symposium 2015, Milan, Italy, July 29, 2015.

R. Nishii, Deforestation modeling based on statistical and machine learning approaches, ISM Symposium on Environmental Statistics(招待講演), 統計数理研究所, 立川市, Feb. 24, 2015.

M. Hosotsubo and R. Nishii, Statistical analysis of the awarding of grants-in-aid for scientific research to faculty members of Japanese national universities and significance to policy making, Remote Sensing and Ecological Science 2014, Gifu, Aug. 30, 2014.

R. Nishii, P. Qin and D. Uchi, Contextual unmixing of geospatial data based on Bayesian modeling, IEEE International Geoscience and Remote Sensing Symposium 2014, Quebec, Canada, July 15, 2014.

S. Tanaka and R. Nishii, Re-evaluation of topographic attributes with human population in deforestation framework with spatial dependency, IEEE International Geoscience and Remote Sensing Symposium 2014, Quebec, Canada, July 17, 2014.

D. Uchi, R. Nishii, and P. Qin, Logistic regression and MCMC for contextual classification of hyperspectral imagery, IMS-APRM 2014, Taipei, July 3, 2014.

西井龍映, 時系列回帰モデルのモデル選択とその応用, 日本応用数理学会 2013 年度年会 (招待講演), 福岡市, Sept. 9, 2013.

S. Tanaka and R. Nishii, Effect evaluation of topographic attributes on forest coverage ratio based on digital elevation model, IEEE International Geoscience and Remote Sensing Symposium 2013, Melbourne, Australia, July 24, 2013.

R. Nishii, P. Qin and D. Uchi, Nonlinear Bayesian unmixing of geospatial data based on Gibbs sampling, IEEE International Geoscience and Remote Sensing Symposium 2013, Melbourne, Australia, July 25, 2013.

P. Qin, K. Yanase and R. Nishii, Multi-hour-ahead prediction of Dst index using nonlinear autoregressive models with exogenous variables, Japan Geoscience Union Meeting 2013, Japan Geoscience Union Meeting, 千葉市, April 28, 2013.

〔図書〕(計1件)

R. Nishii et al. (eds.) (2014). A Mathematical Approach to Research Problems of Science and Technology, Mathematics for Industry. Vol. 5, Springer, 507 頁. ISBN 978-4-431-55060-0

6 . 研究組織

(1) 西井 龍映 (Nishii, Ryuei)

九州大学・マス・フォア・インダストリ
研究所・教授

研究者番号 : 40127684