多数の例外的データが存在する回帰問題のための最小記述長原理の拡張

鈴木 英之進† 志村 正道††

データの集合からモデルを推定する回帰問題においては，どのようなモデルを仮定してもモデルで説明できない例外的データが存在する場合がある．従来用いられてきたモデルの妥当性を評価する規準は，例外的データが存在する回帰問題には適用できないか，適用できても例外的データが多いときには不適切である．本論文では，最小記述長原理を，例外的データを説明する例外的モデルと例外的データ以外の通常的データを説明する通常的モデルを仮定することにより拡張し，この拡張した最小記述長原理に基づく情報量規準を提案する．情報量規準では，モデルの妥当性を，2つのモデルの記述長，例外的モデルを仮定したときの例外的データの記述長，および通常的モデルを仮定したときの通常的データの記述長についての和の短さで表した．この情報量規準を，多数の例外的データが存在する典型的な回帰問題である，地震データから地中の構造を推定する問題に適用した．実験の結果，情報量規準を用いたときの推定結果は，従来用いられてきたカットつき最小2乗法に基づく評価規準や経験的評価規準を用いたときの推定結果より，正確であることが分かった．

Extension of Minimum Description Length Principle for Regression with Numerous Exceptional Data

EINOSHIN SUZUKI† and MASAMICHI SHIMURA††

Regression, which is an estimation problem of the model from a set of data, often contains exceptional data which cannot be explained by any models. Existing criteria which evaluate the goodness of models are either inapplicable to this kind of problem, or are applicable but inadequate when exceptional data are numerous. In this paper, we extend the minimum description length principle by assuming exceptional model and normal model for explaining exceptional data and normal data respectively. Based on this extended minimum description length principle, we propose the information criterion. In this criterion, a model is judged better if it has a shorter add-sum of the description length of the two models, the description length of the exceptional data relative to the exceptional model, and the description length of the normal data relative to the normal model. The criterion has been applied to the problem of estimating an underground structure from a set of earthquake data, which is a typical regression with numerous exceptional data. As the result of the experiments, it has been shown that the information criterion enables a more accurate estimation compared to the criterion based on the least trimmed squares method and the empirical evaluation criterion, which are currently employed for this problem.

1. はじめに

回帰問題とは，データの集合からその背景にあるモデルを推定する問題の一種である．回帰問題として形式化できる問題は，与えられた振舞いを満たす構造を求める設計問題や，観察されたデータの集合から構造を推定する構造推定問題など，数多く存在する．回帰問題は一般に，さまざまなモデルについてデータの説明を求める順問題を解いてモデルの妥当性を評価し，最も良いモデルを求めることにより解かれる．しかし，現実の回帰問題では，順問題を解くと妥当でない結果が得られてしまうデータや，データと理論値の差である残差の絶対値が非常に大きいデータ，あるいは順問題が解けないデータが存在し，これらのデータをモデルが説明できないという意味で例外的データと呼ぶ．したがって，現実の回帰問題でモデルの妥当性を評価するためには一般に，例外的データへの対処が必要となる．従来用いられてきたモデルの評価規準では，例

† 横浜国立大学工学部電子情報工学科
Division of Electrical and Computer Engineering, Faculty of Engineering, Yokohama National University
†† 東京工業大学大学院情報理工学研究科計算工学専攻
Department of Computer Science, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

外的データを，単に除外したり，異常値と見なしたり，単純な経験的手法を用いて処理してきた．しかし，例外的データが多数存在する場合には，例外的データを除外してそれ以外の通常的データに基づいてモデルの妥当性を評価することは困難であると考えられる．また，この場合，例外的データを異常値として処理するためには，多くの経験と結果に対する深い知識が必要である．経験的評価規準は理論的な根拠を持たず，一般には信頼性に欠ける．

近年，モデル推定問題において，モデルの妥当性を評価するための規準として提案されているのが，最小記述長原理[1]に基づく評価規準である．この評価規準では，モデルの妥当性をモデルの記述長とそのモデルを仮定したときのデータの記述長についての和の短さで表す．最小記述長原理は，いくつかの優れた統計的性質を持ち，さまざまな分野で用いられていて成果をあげている[2)~5)]．しかし，例外的データが存在する回帰問題では，順問題が解けないデータについてはモデルを仮定したときのデータの記述長は計算できず，残差の絶対値が非常に大きいデータについてはモデルを仮定したときのデータの記述長は非常に大きな値となってしまうため，最小記述長原理を例外的データが存在する回帰問題にそのままの形で適用することは困難である．

本論文では，多数の例外的データが存在する回帰問題におけるモデルの新しい評価規準として，例外的データを扱えるように拡張した最小記述長原理に基づく情報量規準を提案する．情報量規準では，通常的データと例外的データを個別に説明するモデルを仮定し，モデルの妥当性を，2つのモデルについての記述長，およびそれぞれのモデルを仮定したときの通常的データと例外的データの記述長についての和の短さで表す．この情報量規準を評価するためには，種々の人工的な問題に適用することが考えられるが，ここでは情報量規準の有用性を示すために，多数の例外的データが存在する現実の回帰問題である，地震データから地中の構造を推定する問題に適用した．近畿地方北部・中国地方東部における地震データを用いた実験の結果，情報量規準は従来用いられてきたカットつき最小2乗法に基づく評価規準や経験的評価規準[6),7)]に比べて，妥当性が高いことを確かめることができた．

2. 回帰問題におけるモデルの評価規準

2.1 回帰問題

回帰問題とは，データ集合 D からその背景にある真のモデル M を推定する問題の一種である．データ集合 $D = \{d_1, d_2, \cdots, d_n\}$ は，n 個のデータ $d_1, d_2, \cdots, d_n$ から構成される．本論文では，データ $d_i = \{d_{i1}, d_{i2}, \cdots, d_{im}\}$ は，m 個の数値 $d_{i1}, d_{i2}, \cdots, d_{im}$ から構成されるものとする．順問題 f とは与えられたモデルのもとでデータの説明を求めるものであり，データ d_i とモデル M を入力として，データと理論値の差である残差 $r_i(M)$ および理論値の誤差などを表す付加情報 $h(d_i, M)$ を出力する問題である．すなわち，$f(d_i, M) = \{r_i(M), h(d_i, M)\}$ となる．データを構成する数値 d_{ij} とその理論値の差を $r_{ij}(M)$ で表すと，残差 $r_i(M)$ は，$r_i(M) = \{r_{i1}(M), r_{i2}(M), \cdots, r_{im}(M)\}$ となる．

回帰問題は通常，さまざまなモデルを仮定し，そのモデルに基づいて順問題を解いて，残差などからモデルの妥当性を評価し，最も良いモデルを求めることにより解かれる．データ集合 D に対するモデル M の妥当性を $goodness(M, D)$，悪さを $badness(M, D)$ で表す．ここで，モデルとデータが妥当である場合，順問題を解いて得られる理論値や付加情報も妥当であり，理論値はデータにほぼ等しくなるので残差を構成する各数値 $r_{ij}(M)$ はほぼ0に等しくなる．この場合，モデルはデータを説明するといい，このようなデータを通常的データと呼ぶことにする．逆にモデルかデータのどちらかが妥当でない場合には，理論値や付加情報が妥当でないか，順問題が解けないか，あるいは残差を構成する数値の絶対値が大きい．この場合，モデルはデータを説明しないといい，そのようなデータを例外的データと呼ぶことにする．ここで，例外的データの中で，理論値や付加情報が妥当でないか，順問題が解けないデータを不能データ，残差を構成する数値の絶対値が大きいデータを異常データと呼ぶことにする．本論文では，データ集合中に多数の例外的データが存在する回帰問題を扱うことにする．

2.2 従来の評価規準

モデルの妥当性を評価する規準としては，最尤法に基づく評価規準，最小2乗法やカットつき最小2乗法に基づく評価規準，ロバスト推定法に基づく評価規準，およびベイズ的アプローチに基づく評価規準などがある．以下，これらの評価規準を簡単に説明しておく．

最尤法に基づく評価規準では，式(1)に示すようにモデルの妥当性 $goodness$ を，モデル M を仮定したときにデータ集合 D を得る条件つき確率 $\mathrm{P}(D|M)$ で表す．

$$goodness(M, D) \equiv \mathrm{P}(D|M) \tag{1}$$

$$= \prod_{i=1}^{n} \prod_{j=1}^{m} \mathrm{P}(d_{ij}|M) \tag{2}$$

各データ d_i を生じた事象も各データを構成する数値 d_{ij} を生じた事象も互いに独立であると仮定した場合，式(2)に示すように右辺はモデル M を仮定したときにデータを構成する数値 d_{ij} を得る確率 $\mathrm{P}(d_{ij}|M)$ の積となる．以下，各データを生じた事象も各データを構成する数値を生じた事象も互いに独立であると仮定する．

最小2乗法に基づく評価規準では，式(3)に示すようにモデルの悪さ *badness* を，残差を構成する数値 $r_{ij}(M)$ の2乗をその分散 $\sigma_{ij}(M)^2$ で割った値についての和で表す．なお，最小2乗法は，最尤法において残差を構成する数値 $r_{ij}(M)$ が正規分布に従う場合に相当する[8]．

$$badness(M,D) \equiv \sum_{i=1}^{n}\sum_{j=1}^{m}\frac{r_{ij}(M)^2}{\sigma_{ij}(M)^2} \tag{3}$$

現実の回帰問題においては，残差を構成する数値の中にいくつかの絶対値が非常に大きい数値が存在することが多く，式(3)に示す *badness* の値は主に一部の絶対値が非常に大きい数値によって支配されてしまう．したがって，最小2乗法に基づく評価規準は，そのままの形では現実の回帰問題に対して適切でない．カットつき最小2乗法は，このような問題に対して考えられたロバスト推定法[8]の一種であり，式(4)に示すように，絶対値がある閾値 θ 以上の数値を含む残差を除外する方法である．

$$badness(M,D) \equiv \sum_{i\ (|r_{ij}(M)|<\theta)}\sum_{j=1}^{m}\frac{r_{ij}(M)^2}{\sigma_{ij}(M)^2} \tag{4}$$

カットつき最小2乗法以外にも種々のロバスト推定法が存在し，それらは異常データを小さく重みづけすることによって，上述の問題点を回避している[8]．しかし，この方法では，対象問題に関する深い知識や経験を用いて通常的データと異常データとの区別を修正しながら推定を繰り返す必要があり，そのような知識や経験がない場合には，この方法は適用できない．

上述の評価規準では，モデルの妥当性は，モデル M を仮定したときにデータ集合 D を得る条件つき確率 $\mathrm{P}(D|M)$ で表されていた．これに対しベイズ的アプローチに基づく評価規準では，モデルの妥当性は，データ集合 D を仮定したときにモデル M を得る確率である事後確率 $\mathrm{P}(M|D)$ で表される．事後確率は式(5)に示すベイズの式で与えられるが，データ集合 D は問題に対して不変であるために，データ集合の出現確率 $\mathrm{P}(D)$ は無視することができる．したがって，ベイズ的アプローチに基づく評価規準では，モデルの妥当性 *goodness* は，式(6)に示すように，モデル M を仮定したときにデータ集合 D を得る条件つき確率 $\mathrm{P}(D|M)$ と，モデル M の出現確率である事前確率 $\mathrm{P}(M)$ についての積で表される．

$$\mathrm{P}(M|D) = \frac{\mathrm{P}(M)\mathrm{P}(D|M)}{\mathrm{P}(D)} \tag{5}$$

$$goodness(M,D) \equiv \mathrm{P}(M)\mathrm{P}(D|M) = \mathrm{P}(M)\prod_{i=1}^{n}\prod_{j=1}^{m}\mathrm{P}(d_{ij}|M) \tag{6}$$

データを構成する数値のうち，異常データ中の数値 d_{ij} については，モデル M を仮定したときにその数値を得る確率が $\mathrm{P}(d_{ij}|M)\approx 0$ となり，丸め誤差などを考えると $\mathrm{P}(d_{ij}|M)$ の値は信頼できない．したがって，式(6)の右辺はこのような数値によって信頼できない値になってしまうと考えられる．さらに，不能データ中の数値 d_{ij} については，$\mathrm{P}(d_{ij}|M)$ の値を計算することはできない．ベイズ的アプローチに基づく評価規準を，例外的データが存在する回帰問題に適用する場合，例外的データについて，モデル M を仮定したときにデータを構成する数値を得る確率 $\mathrm{P}(d_{ij}|M)$ の計算方法を決定しておく必要がある．

2.3 最小記述長原理

本節では，最小記述長原理を，ベイズ的アプローチの拡張という観点から簡単に説明しておく．式(5)に示すベイズの式の両辺において，底が2の対数をとって -1 を乗ずると式(7)が得られる．以下，対数 log の底はすべて2であるとする．

$$-\log\mathrm{P}(M|D) = -\log\mathrm{P}(M) - \log\mathrm{P}(D|M) + \log\mathrm{P}(D) \tag{7}$$

情報理論によれば，データ X を0と1を用いて符号化したときの平均の意味での記述長の下限は，情報量 $-\log\mathrm{P}(X)$ で与えられる．事後確率 $\mathrm{P}(M|D)$ が増加するときに式(7)の右辺は単調に減少し，$\log\mathrm{P}(D)$ は定数であることから，モデルおよびそのモデルを仮定したときのデータ集合を短く符号化してそれぞれの情報量 $-\log\mathrm{P}(M)$，$-\log\mathrm{P}(D|M)$ を求め，両者についての和の小ささでモデルの妥当性を表せることが分かる．

これまで述べたモデルの評価規準はすべて，この符号化の考え方を用いて説明することができる．たとえば最尤法では，モデルの妥当性は，モデルを仮定した場合のデータ集合をある確率分布に従って符号化したときの記述長 $-\log\mathrm{P}(D|M)$ の短さで表される．最小2乗法は最尤法において，残差 $r_{ij}(M)$ の確率分布が正規分布となる場合に相当する．ベイズ的アプロー

チでは，モデルの妥当性は，モデルを仮定したときのデータ集合の記述長 $-\log\mathrm{P}(D|M)$ とモデルの記述長 $-\log\mathrm{P}(M)$ についての和の短さで表される．以上の評価規準では，モデルを仮定したときのデータ集合を符号化するための確率分布とモデルを符号化するための確率分布があらかじめ与えられている必要がある．

これらに対し，最小記述長原理では，モデルを仮定したときのデータ集合 D とモデル M を符号化するための確率分布は，それぞれ記述長 $-\log\mathrm{P}(D|M)$，$-\log\mathrm{P}(M)$ が平均の意味で最も短くなる確率分布であるとする．最小記述長原理に基づくモデルの評価規準 *badness* を式 (8) に示す．

$$\begin{aligned} &badness(M,D) \\ &\quad\equiv -\log\mathrm{P}(M) - \log\mathrm{P}(D|M) \\ &\quad= -\log\mathrm{P}(M) - \sum_{i=1}^{n}\sum_{j=1}^{m}\log\mathrm{P}(d_{ij}|M) \end{aligned} \qquad (8)$$

しかし，例外的データが存在する回帰問題では，不能データについては $-\log\mathrm{P}(d_{ij}|M)$ は計算できず，異常データについては $\mathrm{P}(d_{ij}|M) \approx 0$ であるために $-\log\mathrm{P}(d_{ij}|M) \approx \infty$ となってしまう．したがって，最小記述長原理に基づく評価規準は，そのままの形では例外的データが存在する回帰問題におけるモデルの評価規準として適切でない．

2.4 経験的評価規準

経験的評価規準とは，理論的背景はないが経験によってその有効性が実証されたモデルの評価規準である．多数の例外的データが存在する回帰問題においては，適切に定められた経験的評価規準を用いると正確にモデルを評価できる場合がある．たとえば，4 章で述べる地震波速度構造推定問題において，モデルの妥当性をモデルで説明できる通常的データの個数で表す経験的評価規準が提案されている[6),7)]．この評価規準を用いて推定したモデルは地震学において過去に得られた結果とほぼ一致しており，このことは多数の例外的データが存在する回帰問題に対して適切に定められた経験的評価規準が有効であることを示唆している．ただし，経験的評価規準は，理論的な根拠を持たないためにその妥当性が保証されず，一般には信頼性に欠ける．

3. 例外的データを扱える最小記述長原理

3.1 情報量規準

モデルの評価規準を例外的データが存在する回帰問題に適用する場合，モデルが説明できない例外的データ d_i について，確率 $\mathrm{P}(d_{ij}|M)$ をどのように計算するかが問題であった．カットつき最小 2 乗法に基づく評価規準では例外的データを除外して，それ以外の通常的データに基づいてモデルの妥当性を表す．しかし，例外的データが多数存在する場合においては，通常的データを例外的データと間違えてしまう可能性があり，モデルの妥当性を正しく表すことは困難であると考えられる．その他のロバスト推定法では，通常的データと異常データを区別するための深い知識や経験が必要である．ベイズ的アプローチに基づく評価規準と最小記述長原理に基づく評価規準もそれぞれ 2.2，2.3 節で述べたように，そのままの形では例外的データが存在する回帰問題におけるモデルの評価規準として適切でない．

本論文では，最小記述長原理を例外的データが扱えるように拡張し，この拡張した最小記述長原理に基づく評価規準を提案する．最小記述長原理では，モデルの妥当性を，(1) モデルの記述長と，(2) そのモデルを仮定したときのデータの記述長についての和の短さで表している．本論文では，すべてのデータを単一のモデルを仮定して符号化するのではなく，通常的データを説明する**通常的モデル** M_n と例外的データを説明する**例外的モデル** M_e を仮定し，それぞれの種類のデータをフラグで区別して個別に符号化する方法を提案する．すなわち，**モデルの妥当性を，次の 3 つの記述長についての和の短さで表す方法を提案する．**

(1) 通常的モデルと例外的モデルの記述長，
(2) 通常的モデルを仮定したときの通常的データの記述長，
(3) 例外的モデルを仮定したときの例外的データの記述長．

上述の記述長は情報量を表すため，この評価規準を**情報量規準**と呼ぶことにする．以下，通常的データと例外的データの集合をそれぞれ $n(M)$，$e(M)$ で表す．また，データを構成する数値の個数を ν，通常的データを構成する数値の個数を $\nu_{\mathrm{n}}(M)$，例外的データを構成する数値の個数を $\nu_{\mathrm{e}}(M)$ で表すことにする．

3.2 データの符号化

通常的データと例外的データを個別に符号化する場合，データを構成する数値 d_{ij} の記述長 $-\log\mathrm{P}(d_{ij}|M)$ は，式 (9) に示すように，それぞれの種類における記述長 $-\log\mathrm{P}(d_{ij}|M_{\mathrm{n}}, d_i \in n(M))$ あるいは $-\log\mathrm{P}(d_{ij}|M_{\mathrm{e}}, d_i \in e(M))$ と，$-\log(\nu_{\mathrm{n}}(M)/\nu)$ あるいは $-\log(\nu_{\mathrm{e}}(M)/\nu)$ についての和で与えられる．ここでの $-\log(\nu_{\mathrm{n}}(M)/\nu)$，$-\log(\nu_{\mathrm{e}}(M)/\nu)$ は，式 (9) に示すように，前述の通常的データと例外的データを区別するフラグの記述長を表していると見なすことが

できる．

$$
\begin{aligned}
&-\log \mathrm{P}(d_{ij}|M) \\
&= -\log\{\mathrm{P}(d_{ij}|M, d_i \in n(M)) \cdot \mathrm{P}(d_i \in n(M)|M) \\
&\quad + \mathrm{P}(d_{ij}|M, d_i \in e(M)) \cdot \mathrm{P}(d_i \in e(M)|M)\} \\
&= \begin{cases} -\log \mathrm{P}(d_{ij}|M_{\mathrm{n}}, d_i \in n(M)) - \log\left(\frac{\nu_{\mathrm{n}}(M)}{\nu}\right) \\ \qquad (d_i \in n(M) \text{ のとき}) \\ -\log \mathrm{P}(d_{ij}|M_{\mathrm{e}}, d_i \in e(M)) - \log\left(\frac{\nu_{\mathrm{e}}(M)}{\nu}\right) \\ \qquad (d_i \in e(M) \text{ のとき}) \end{cases}
\end{aligned} \tag{9}
$$

通常的データ $d_i \in n(M)$ については，現実の回帰問題では残差を構成する数値 $r_{ij}(M)$ がほぼ正規分布に従うことが知られているので，通常的モデル M_n としては，残差を構成する数値についての正規分布を仮定する．一方，例外的データ $d_i \in e(M)$ については，残差が計算できないか，たとえ計算できたとしても絶対値が非常に大きい数値を含み信頼できないので，例外的モデル M_e としては，残差ではなくデータを構成する数値 d_{ij} を説明するモデルを仮定する．例外的モデルとして用いる確率分布は問題に依存し，めったに起こらない事象についてはポアッソン分布，整数全体については整数のユニバーサル事前確率[9]というように，データが発生する機構に応じた確率分布を用いる．

ここで，通常的データと例外的データの区別は，次のように決定する．まず，不能データについては，その定義に従って例外的データとする．その他のデータについては，残差がある閾値 θ よりも絶対値が大きい数値を含むデータを例外的データとし，それ以外のデータを通常的データとする．ここで閾値 θ は，最小記述長原理の定義に従い，全体の記述長が最も短くなる値とする．式 (10) に本論文で提案する情報量規準 *badness* を示す．ただし，$d_i \in n(M)$ について $|r_{ij}(M)| < \theta$ である．

$$
\begin{aligned}
&badness(M, D) \\
&\equiv -\log \mathrm{P}(M_{\mathrm{n}}) - \log \mathrm{P}(M_{\mathrm{e}}) \\
&\quad + \sum_{i\ (d_i \in n(M))} \sum_{j=1}^{m} \left\{ -\log\left(\frac{\nu_{\mathrm{n}}(M)}{\nu}\right) \right. \\
&\qquad \left. -\log \mathrm{P}(d_{ij}|M_{\mathrm{n}}, d_i \in n(M)) \right\} \\
&\quad + \sum_{i\ (d_i \in e(M))} \sum_{j=1}^{m} \left\{ -\log\left(\frac{\nu_{\mathrm{e}}(M)}{\nu}\right) \right. \\
&\qquad \left. -\log \mathrm{P}(d_{ij}|M_{\mathrm{e}}, d_i \in e(M)) \right\}
\end{aligned} \tag{10}
$$

4. 地震波速度構造推定問題への適用

4.1 地震波速度構造推定問題

情報量規準の有効性を確認するために，本論文では，地震データから地中の構造を推定する問題である地震波速度構造推定問題を用いることにする．地震波速度構造推定問題においては，データ集合中には推定する構造の範囲外で起こった地震のデータや S/N 比が低いデータなどの例外的データが多数存在する．また，地震の規模や観測計器まわりの振動などの理由により，データ d_i にはいくつかの欠落値が存在する場合が多く，地震波速度構造推定問題は例外的データが多数存在する回帰問題のうちでも困難な問題のひとつとなっている．なお，地震の震源位置を正確に計算するためには，正確な地震波速度構造が必要であるために，地震波速度構造推定問題は，地震学において重要な課題のひとつとなっている．

地震波速度構造推定問題は，地震から得られるデータ集合 D から，構造 M を推定する回帰問題となっている．地層の厚さが緯度方向と経度方向の両方について均一である地震波速度構造を一次元構造と呼ぶ．一次元構造 M は，k 番目の地層の厚さ u_k とその地層において地震波が伝わる速度 v_k から構成される．すなわち，層の数を l とすると，$M = \{(u_1, v_1), (u_2, v_2), \cdots, (u_l, v_l)\}$ となる．対象とする地震が n 回あり，i 番目の地震から得られるデータを d_i で表すと，$D = \{d_1, d_2, \cdots, d_n\}$ となる．データ d_i は，m 個の観測点で観測されたP波とS波についての到着時刻から構成される．データ d_i は，観測点 j でのそれぞれの到着時刻を p_{ij}, s_{ij} で表すと，$d_i = \{p_{i1}, s_{i1}, p_{i2}, s_{i2}, \cdots, p_{im}, s_{im}\}$ となる．地震波速度構造推定問題における順問題 f は，データ d_i と地震波速度構造 M を入力として，残差 $r_i(M)$ および付加情報 $h(d_i, M)$ を出力する問題に相当する．すなわち，$f(d_i, M) = \{r_i(M), h(d_i, M)\}$ である．付加情報 $h(d_i, M)$ は震源の緯度 $lat_i(M)$，経度 $lon_i(M)$，深さ $dep_i(M)$ と地震発生時刻 $time_i(M)$ およびそれらの誤差 $elat_i(M)$, $elon_i(M)$, $edep_i(M)$, $etime_i(M)$ から構成される．P波到着時刻 p_{ij} とその理論値の差をP波到着時刻残差と呼び，$r^{\mathrm{P}}_{ij}(M)$ で表す．同様に，S波到着時刻 s_{ij} とその理論値の差をS波到着時刻残差と呼び，$r^{\mathrm{S}}_{ij}(M)$ で表す．残差 $r_i(M)$ は，$r_i(M) = \{r^{\mathrm{P}}_{i1}(M), r^{\mathrm{S}}_{i1}(M), r^{\mathrm{P}}_{i2}(M), r^{\mathrm{S}}_{i2}(M), \cdots, r^{\mathrm{P}}_{im}(M), r^{\mathrm{S}}_{im}(M)\}$ となる．

4.2 情報量規準の地震波速度構造推定問題への適用

情報量規準を地震波速度構造推定問題に適用するに

あたって，注意すべき点が2つある．最初の点は，P波到着時刻残差 $r_{ij}^{\mathrm{P}}(M)$ の確率分布と，S波到着時刻残差 $r_{ij}^{\mathrm{S}}(M)$ の確率分布が異なることである．したがって，通常的データを符号化する際に，P波到着時刻とS波到着時刻を異なるモデル M_{nP} と M_{nS} を用いてそれぞれ個別に符号化し，その際3章で述べたように，通常的データにおけるP波到着時刻の個数 $\nu_{\mathrm{nP}}(M)$ とS波到着時刻の個数 $\nu_{\mathrm{nS}}(M)$ に応じた長さのフラグを用いる必要がある．2番目の点は，S波到着時刻残差 $r_{ij}^{\mathrm{S}}(M)$ の絶対値は一般に，P波到着時刻残差 $r_{ij}^{\mathrm{P}}(M)$ の絶対値よりも約2倍大きいことである．したがって，データ d_i が通常的データであるか例外的データであるかを決定する際に，異なる閾値を用いなければならない．ここでは，P波到着時刻残差については θ，S波到着時刻残差については 2θ を用いることにした．θ は，3節で述べたように，通常的データの記述長と例外的データの記述長についての和が最も短くなる値とした．

不能データは，到着時刻残差を構成する数値の絶対値の大きさに関係なく例外的データとなる．地震学で用いられている方法に従い，震源位置 $(lat_i(M), lon_i(M), dep_i(M))$ が，推定する地震波速度構造の範囲外にあるか，あるいは震源の緯度，経度，深さの誤差 $elat_i(M), elon_i(M), edep_i(M)$ の絶対値が，あらかじめ決めた値以上である場合は，そのデータを不能データであると見なすことにした．

例外的データについてはデータを構成する数値を符号化するが，3章ではこの符号化に用いる確率分布は問題に依存すると述べた．地震波速度構造推定問題において，データを構成する数値は地震波の到着時刻を表しており，各地震について地震波が早く届く観測点と遅く届く観測点が存在することから，到着時刻の確率分布は何種類かのポアッソン分布を合成した分布であると考えられる．したがって，データを構成する数値すなわち地震波の到着時刻を符号化するために用いる確率分布としては，何種類かのポアッソン分布を合成した分布として導かれるパスカル分布[10]を用いた．式(11)に，地震波速度構造推定問題における情報量規準 *badness* を示す．ただし，$d_i \in n(M)$ について $|r_{ij}^{\mathrm{P}}(M)| < \theta$，$|r_{ij}^{\mathrm{S}}(M)| < 2\theta$ である．

$$
\begin{aligned}
&badness(M, D) \\
&\equiv -\log \mathrm{P}(M_{\mathrm{nP}}) - \log \mathrm{P}(M_{\mathrm{nS}}) - \log \mathrm{P}(M_{\mathrm{e}}) \\
&+ \sum_{i\,(d_i \in n(M))} \left[\sum_{j\,(p_{ij})} \left\{ -\log \left(\frac{\nu_{\mathrm{nP}}(M)}{\nu} \right) \right.\right. \\
&\qquad \left. -\log \mathrm{P}(p_{ij} | M_{\mathrm{nP}}, d_i \in n(M), \mathrm{Pwave}) \right\} \\
&\qquad + \sum_{j\,(s_{ij})} \left\{ -\log \left(\frac{\nu_{\mathrm{nS}}(M)}{\nu} \right) \right. \\
&\qquad \left.\left. -\log \mathrm{P}(s_{ij} | M_{\mathrm{nS}}, d_i \in n(M), \mathrm{Swave}) \right\} \right] \\
&+ \sum_{i\,(d_i \in e(M))} \left[\sum_{j\,(p_{ij})} \left\{ -\log \left(\frac{\nu_{\mathrm{e}}(M)}{\nu} \right) \right.\right. \\
&\qquad \left. -\log \mathrm{P}(p_{ij} | M_{\mathrm{e}}, d_i \in e(M)) \right\} \\
&\qquad + \sum_{j\,(s_{ij})} \left\{ -\log \left(\frac{\nu_{\mathrm{e}}(M)}{\nu} \right) \right. \\
&\qquad \left.\left. -\log \mathrm{P}(s_{ij} | M_{\mathrm{e}}, d_i \in e(M)) \right\} \right] \qquad (11)
\end{aligned}
$$

4.3 記述長の計算例

図1は，ある地震についてのデータ d_1 とその残差 $r_1(M)$ および付加情報 $h(d_1, M)$ を示したものである．ただし，? は欠落値を表す．以下，通常的データの記述長と例外的データの記述長を計算する方法の具体例として，データ d_1 の記述長を求める方法を簡単に示す．ただし，地震学的に見て理論値と付加情報は妥当であるとする．

今，通常的データと例外的データを区別する閾値が $\theta = 0.10$ 秒であるとする．データ d_1 は，$|r_{13}^{\mathrm{P}}(M)| > \theta$，$|r_{1j}^{\mathrm{S}}(M)| > 2\theta$ $(j = 3, 4, 5)$ であるため例外的データとなる．d_1 以外のデータについても，例外的データか通常的データかを判断し，例外的データについて，到着時刻の記述長についての総和が最も短くなるパスカ

	d_1		$r_1(M)$		$h(d_1, M)$
j	p_{1j} (s)	s_{1j} (s)	$r_{1j}^{P}(M)$ (s)	$r_{1j}^{S}(M)$ (s)	$lon_1 = 133.59°$
1	0.50	?	-0.05	—	$lat_1 = 34.91°$
2	0.63	?	0.04	—	$dep_1 = 26.18$ km
3	0.65	1.61	-0.16	-0.39	$time_1 = -6.62$ s
4	0.14	0.69	0.01	-0.38	$elon_1 = 0.35°$
5	0.00	0.45	0.09	-0.25	$elat_1 = 1.47°$
					$edep_1 = 1.35$ km
					$etime_1 = 1.47$ s

図1 地震のデータとその残差，付加情報の例
Fig. 1 Example of the earthquake data, their residuals, and their additional information.

表1 3つの評価規準の比較
Table 1 Comparison of the three criteria.

地震データの個数（%）		100	90	80	70	60	50
正答率（%）	情報量規準	100	100	100	100	85	75
	経験的評価規準	100	90	65	25	10	10
	カットつき最小2乗法	0	0	0	0	0	0

表2 通常的データを構成する数値の全データに占める割合（最小値，平均値，最大値）
Table 2 Ratio of the normal data (minimum, average, maximum).

地震データの個数（%）	100	90	80	70	60	50
情報量規準	(-,70,-)	(69,70,71)	(68,70,71)	(69,70,72)	(67,71,74)	(69,71,73)
経験的評価規準	(-,68,-)	(66,68,69)	(66,68,69)	(67,68,70)	(66,68,72)	(66,69,71)
カットつき最小2乗法	(-,65,-)	(64,65,66)	(64,65,67)	(63,65,68)	(60,65,68)	(62,65,68)

ル分布のパラメータを推定する．このパスカル分布を用いて，データ d_1 を構成する到着時刻の記述長を求め，それらの和をデータの記述長とする．また，例外的データのフラグについては，その記述長をP波到着時刻とS波到着時刻について個数分加えた値，たとえばデータ d_1 については $-5\log(\nu_e(M)/\nu)-3\log(\nu_e(M)/\nu)$ を，フラグの長さとする．データ d_1 の記述長は，データの記述長とフラグの長さの和で与えられる．

一方，閾値が $\theta = 0.20$ 秒である場合，データ d_1 は，$|r_{1j}^{\mathrm{P}}(M)| < \theta$ $(j = 1,2,\cdots,5)$，$|r_{1j}^{\mathrm{S}}(M)| < 2\theta$ $(j = 3,4,5)$ であるため通常的データである．d_1 以外のデータについても，例外的データか通常的データかを判断し，通常的データについて，P波到着時刻残差とS波到着時刻残差についてのそれぞれの総和が最も短くなる正規分布のパラメータを推定する．これらの正規分布を用いて，残差 $r_1(M)$ を構成するP波到着時刻残差とS波到着時刻残差の記述長を求め，それらの和を残差の記述長とする．また，通常的データのフラグ，P波のフラグおよびS波のフラグについては，それぞれの記述長を個数分加えた値，たとえばデータ d_1 については $-5\log(\nu_{\mathrm{nP}}(M)/\nu) - 3\log(\nu_{\mathrm{nS}}(M)/\nu)$ を，フラグの長さとする．データ d_1 の記述長は，残差の記述長とフラグの長さについての和で与えられる．

以上の計算をさまざまな閾値 θ について行い，通常的データの記述長と例外的データの記述長についての和が最も短くなるように θ の値を決める．最終的な d_1 の記述長は，その θ を用いたときの記述長である．

4.4 近畿地方北部・中国地方東部における推定

近畿地方北部・中国地方東部は，過去に数度人工的に地震を起こすことにより，その構造が明らかにされている[11)~13)]．この地方の地震波速度構造を推定する問題において，情報量規準を地震学で従来用いられてきたカットつき最小2乗法に基づく評価規準および経験的評価規準[6),7)]と比較した．ここで用いた経験的評価規準は，地震波速度構造の妥当性を，通常的データの個数で表したものである．データとしては，京都大学の鳥取，阿武山微小地震観測網で1987～1991年の間に観測された665回の地震から得られたものを用いた．

情報量規準はどのような構造についても適用できるが，ここではその妥当性を示すことが目的であるので，一次元構造を対象とした．地震学によれば，西南日本における地震波速度構造は通常4層構造であり，第3層と第4層などについていくつかの条件が成立することが知られている．これらの条件に従い，第3層の深さは50 km以下，速度は下の層ほど速く，第3層までは5.0 km/s以上7.5 km/s以下，第4層では7.0 km/s以上8.0 km/s以下であるとした．モデルの範囲は，観測点の位置を考えて，北緯34.7度から35.6度，東経133.1度から136.1度とした．実験では，各層の速度は0.5 km/s，厚さは5 kmおきに調べた．

実験では665個のデータを用いたが，データが少ないとき，すなわち665個からランダムに選んだ90，80，…，50％のデータを用いた場合についても調べた．表1は，前述の3つの評価規準に基づいて推定を行ったとき，20回の実験中推定結果が正確であった回数をそれぞれ％で示したものである．ただし，すべてすなわち100％のデータについては，1回の実験結果である．また，参考のために，通常的データを構成する数値が全データに占める割合の最小値，平均値，最大値を表2に示す．

表1より，本論文で提案する情報量規準は，他の評価規準に比較して推定結果が正確である場合が多い，すなわち最も良い評価規準となっていることが分かる．経験的評価規準は，特に80％以下の地震数においては推定結果が正確である場合が著しく減少する．したがって，経験的評価規準は，正確な推定を行うためには多数のデータを必要とする評価規準である．実験

で用いた経験的評価規準は，モデルの妥当性を通常的データの個数に基づいて表すので，数値が少ないデータを多数説明するモデルを高く評価する傾向がある．このことは，表2において，経験的評価規準を用いた場合の通常的データを構成する数値の割合が，情報量規準を用いた場合に比較して少ないことからも分かる．以上より，実験で用いた経験的評価規準は，通常的データと例外的データの情報量に基づいて表す情報量規準に比べて単純であり，そのために推定結果が正確である場合が少なかったと考えられる．また，表1は，カットつき最小2乗法に基づく評価規準においては，推定結果がきわめて正確度に劣ることを示している．カットつき最小2乗法に基づく評価規準は説明するデータの個数を考慮しないため，表2に示すように，この規準を用いた場合の通常的データを構成する数値の割合は，他の規準を用いた場合に比較して少ない．本実験で用いたデータは例外的データを多数含むため，説明するデータの個数を考慮しないカットつき最小2乗法に基づく評価規準は，正答率がきわめて低くなったと考えられる．実際には，カットつき最小2乗法に基づく評価規準は，地震学の専門家が領域知識などを用いて例外的データを除外したデータに対して用いられるか，あるいは対象とするモデルの種類をかなり限定して用いられる．

これらの結果は，例外的データが多数存在し，対象とするモデルの種類が限定されていない場合には，カットつき最小2乗法のような例外的データを除外する方法に基づく評価規準や経験的評価規準に比較して，本論文で提案する情報量規準が優れていることを示している．

5. おわりに

多数の例外的データが存在する回帰問題においては，従来用いられてきた評価規準は適切ではないか，対象問題について深い知識や経験が必要であった．ここでは最小記述長原理を，通常的データを説明するモデルと例外的データを説明するモデルを仮定することにより拡張して，この拡張に基づく情報量規準を提案した．多数の例外的データが存在する回帰問題の典型例である地震波速度構造推定問題における実験の結果，地震学で従来用いられてきた評価規準に比較して，情報量規準はきわめて正確であることを確認することができた．

回帰問題として形式化できる問題は現実に数多く存在し，現実の問題を扱うときには例外的データが多数存在する場合があると考えられる．本論文で提案した，例外的データを扱える最小記述長原理に基づく情報量規準は，地震波速度構造推定問題以外の多くの問題についても有効であると考えられる．

謝辞 地震学について教えていただいた京都大学防災研究所の入倉孝次郎教授，片尾浩博士，根本泰雄氏，および根岸弘明氏に感謝します．また，根本氏と根岸氏にはデータの提供についてもお世話になりました．厚くお礼を申し上げます．

参考文献

1) Rissanen, J.: *Stochastic Complexity in Statistical Inquiry*, World Scientific, Singapore (1989).
2) Quinlan, J.R. and Rivest, R.L.: Inferring Decision Trees Using the Minimum Description Length Principle, *Information and Computation*, Vol.80, pp.227–248 (1989).
3) Wallace, C.S.: Coding Decision Trees, *Machine Learning*, Vol.11, No.1, pp.7–22 (1987).
4) 山西健司：MDL入門：計算論的学習理論の立場から，人工知能学会誌，Vol.7, No.3, pp.435–442 (1992).
5) Yamanishi, K.: Probably Almost Discriminative Learning, *Machine Learning*, Vol.18, No.1, pp.23–50 (1995).
6) 根本泰雄，鈴木英之進，望月将志，小平秀一ほか：Azores三重点周辺自然地震記象でのヒューリスティックス探索を用いた地震波速度構造モデルの推定，日本地震学会講演予稿集1994年度秋季大会，p.135 (1994).
7) 根本泰雄，鈴木英之進，根岸弘明：ヒューリスティックス探索を用いた地震波速度構造モデルの推定（西南日本自然地震記象を基に），地球惑星科学関連学会1995年合同学会予稿集，p.647 (1995).
8) 中川 徹，小柳義男：最小二乗法による実験データ解析，東京大学出版会 (1982).
9) Rissanen, J.: A Universal Prior for Integers and Estimation by Minimum Description Length, *The Annals of Statistics*, Vol.11, No.2, pp.416–431 (1983).
10) A.M. ムード，F.A. グレイビル，D.C. ボウズ（著），大石泰彦（訳）：統計学入門（上），マグロウヒル好学社 (1978).
11) 橋爪道朗，川本 整，浅野周三，村松郁栄ほか：第1回，第2回倉吉爆破および花房爆破観測より得られた西部日本の地殻構造，第2部，西部日本の地殻構造，地震，Vol.19, No.2, pp.125–134 (1966).
12) Sasaki, Y., Asano, S., Muramatu, I., Hashizume, M., et al.: Crustal Structure in the Western Part of Japan Derived from the Observation of the First and Second Kurayosi

and the Hanabusa Explosions. Part 2, *Bulletin of the Earthquake Research Institute*, Vol.48, pp.1129–1136 (1970).
13) Yoshii, T., Sasaki, Y., Tada, T., Okada, H., et al.: The Third Kurayosi Explosion and the Crustal structure in the Western part of Japan, *Journal of the Physics of the Earth*, Vol.22, pp.109–121 (1974).

(平成7年8月17日受付)
(平成8年9月12日採録)

鈴木英之進（正会員）

昭和40年生．昭和63年東京大学工学部卒業．平成5年同大学院工学系研究科博士課程修了．工学博士．同年，東京工業大学工学部情報工学科助手を経て，平成8年度より横浜国立大学工学部電子情報工学科講師．データマイニング，モデルの推定など人工知能に関する研究に従事．人工知能学会，AAAI，IEEE Computer Society 各会員．

志村　正道（正会員）

昭和35年東京大学工学部卒業．昭和40年同大学院博士課程修了．工学博士．同年大阪大学基礎工学部助教授，昭和51年東京工業大学工学部助教授を経て，同大学院情報理工学研究科教授，現在に至る．人工知能，学習機械などの研究に従事．著書に「機械知能論」,「人工知能」などがある．人工知能学会，電子情報通信学会，ACM，IEEE，AAAI 各会員．