

12 - 6 確率予測の採点方式「拡張ブライアスコア」とその適用例

The scoring rule "Extended Brier's Score" for a stochastic prediction and an example of application

林 豊 (気象研究所)

Yutaka Hayashi (Meteorological Research Institute)

1. はじめに：予測の採点の方法の重要性

地震の発生の予測において、採点方法の適切さは重要な問題である。地震予知連絡会での予測実験に関する検討の中で、背景の地震活動あるいは平常の活動からみて発生しても当たり前の事象を高い予知率や適中率で予測しても、その予測にあまり意味がない¹⁾、という指摘もある。複数の予測手法の指標値を比較すれば、より高いスコアを得られる手法や最大のスコアを得られる予測パラメータを選択できるが、これは採点環境に最適化する手法を発見することであり、必ずしもより優れた予測手法を選ぶことを意味しない(第1図)。予測の採点に不適切な指標が用いられれば、利用価値が低い予測手法が高いスコアを得ることもあるからである。また、最適化の際も、極値を探索する指標を採点手法だと読み替えられるので、指標が適切でない限り、当該予測手法における最も適切なパラメータを見つけられる保証はない。

本稿では、平常の活動等から容易に推測しうる当たり前の予測との比較によってスコアを求める仕組みを持ち、かつ、予測手法が予測者の信念を曲げて採点方法におもねることができない仕組みも併せ持つような、確率予測のための適切な採点方法の例「拡張ブライアスコア」を提案する。

2. 拡張ブライアスコア

確率予測のための採点方法が満たすべき条件として、「適正な採点ルール」²⁾を拡張して、

- (a) 「適正な採点ルール」：自分の判断通りに予測しなければ、高い点数が得られない、
- (b) 基準予測（平常の活動等から容易に推測できる予測値など基準となる予測）のスコアが一定値である（例、0点）、
- (c) 基準予測に関する対称性、
- (d) 難易度に関する正当な序列

の条件を設けた（付録）。(a)～(d)を同時に満たす採点式の一般解は導ける³⁾が、解の一つに、よく知られた採点式であるブライアスコア⁴⁾を拡張した式のものがある。この「拡張ブライアスコア」とよぶべき採点式は、予測値を f 、基準予測値を c 、予測対象事象の発生の有無を i （事象が発生すれば1、発生しなければ0）、スコアを XBS として、

$$XBS_i(f, c) = \frac{(i-c)^2 - (i-f)^2}{c(1-c)} \quad (1)$$

で表される。基準予測値が1/2の時には、ブライアスコア(BS)の一次関数

$$XBS_i(f, 1/2) = 1 - 4BS_i(f) \quad (2)$$

で表せる。従って、拡張ブライアスコアは、基準予測値というパラメータを追加して、平常の活動等からみた予測対象現象の発生しやすさを加味できるように、ブライアスコアを拡張したものと解釈できる。拡張ブライアスコアは次の特徴を持つ。

- ・尤度などと違い、予測値が0や1であっても安定に計算可能

- ・ブライアスコアなどと違い、任意に定めた基準予測との比較を織り込んだ採点式
- ・基準予測のスコアは0点、完全な予測（予測値が常に0か1で、かつ、すべて適中）のスコアの期待値は1点。基準予測より悪い予測は、スコアは負になる。

3. 適用例：相似地震予測実験の性質の検討

拡張ブライアスコアの有効性を調べるために、日本周辺の相似地震（波形が極めて類似する小繰り返し地震）の予測実験⁵⁾の確率予測値と、予測期間の地震カタログに当該地震が含まれているか否か（これを正解とする）を用いて、採点した（第2図）。この予測実験では、相似地震が系列毎に独立な更新過程で、地震の発生間隔が対数正規分布に従って発生すると仮定し、両確率モデルパラメータの確率分布を系列毎にベイズ推定するモデルに基づいて、相似地震系列毎に、2006～2010年に4回の予測時点から3か月・6か月・1年以内に当該系列の地震が発生する確率がそれぞれ予測されている。本節では、4回とも予測対象となった91系列を抽出し、このうち予測結果の尤度の算出ができない1系列を除く90系列を採点対象とした。

ブライアスコアからでは3か月予測が好成績に見える（第2図左）が、これは必ずしも正しい見方ではない。ここでは、統計期間内の平均地震発生頻度を生起率とするポアソン過程を仮定して求められる単純な予測を基準予測にとったが、3か月予測は基準予測と同程度のスコアしか得ておらず、むしろ、1年予測の方が基準予測からの改善が大きい（第2図左・中）からである。

拡張ブライアスコアには、この例のように基準予測と比較して判明する予測手法の性質が単独のスコア計算だけで把握できるメリットもある（第2図右）。

4. 議論

確率予測の採点に頻繁に用いられるブライアスコアは、2節の(b)の条件を満たさないため、発生しても当たり前の事象の予測に高いスコアを与えうる。3節の例では、ブライアスコアで相似地震予測実験の結果を評価すると、この予測手法が3か月予測よりも1年予測に向いていることを見抜けず、予測の性質に妥当性に疑問のある解釈をもたらすことが判明した。

一方、尤度は基準予測との対数尤度の差をとれば、2節の(b)の条件を満たし、3節の例でも相似地震予測実験の結果に対して拡張ブライアスコアと似た解釈を得られる（第2図中・右）。しかし、予測値が0と1の時にスコアが定義できず、3節の適用例では尤度を算出できない予測例を除外する必要があった。また、尤度では2節の各条件も定義できないことから、尤度または派生する情報量規準等を用いた予測の最適化では、0や1とそれらに近い予測値を必要以上に避ける方法が選択される傾向があると考えられる。

同様に、一般に予測手法はその選択・最適化の過程で用いられた採点手法（指標・統計量等）の性質を反映した性格を持つと推察できる。これは、採点手法が適切でない限り、採点結果（指標値等）の比較に基づく予測手法の良し悪しを議論することや、最適化で予測手法を改善するといった主張は、正当性が担保されないことを示唆しており、予測手法の選択・最適化における根本的な問題を含んでいる。しかしながら、確率予測では、2節の(a)～(d)に挙げた採点手法が満たすべき条件とその解（拡張ブライアスコア等）は、この根本的な問題の解決への指針を示していると期待できる。

参考文献

- 1) 今給黎哲郎 (2016): 平常的な地震活動と「予測」の価値－当たり前の地震について－. 予知連会報, **95**, 425-431.
- 2) Murphy, A. H. and E. S. Epstein (1967): A note on probability forecasts and “Hedging”. *J. Appl. Meteor.*, **6**, 1002-1004.
- 3) 林豊 (1995): 確率予報のための適正な評価法に関する数学的考察. 大阪管区府県気象研究会誌, 平成7年度, 88-89.
- 4) Brier, G. W. (1950): Verification forecasts expressed in terms of probability. *Mon. Weath. Rev.*, **78**, 1-3.
- 5) Okada, M., N. Uchida and S. Aoki (2012): Statistical forecasts and tests for small interplate repeating earthquakes along the Japan Trench. *Earth Planets Space*, **64**, 703-715.

付録

2 節の (a) ~ (d) の各条件は, 次のとおり定式化できる.

- (a) 適正な採点ルール (つまり $p \equiv c$ になる条件)

$$E_s(p, c) > E_s(f, c) \text{ for all } f \neq p \quad (\text{A1})$$

$$E_s \equiv pS_1(f, c) + (1-p)S_0(f, c) \quad (\text{A2})$$

- (b) 基準予測のスコアが一定値である

$$S_i(c, c) = \text{const} \quad (\text{A3})$$

- (c) 基準予測に関する対称性

$$S_1(f, c) \equiv S_0(1-f, 1-c) \quad (\text{A4})$$

- (d) 難易度に関する正当な序列

$$\left. \frac{\partial S_1}{\partial c} \right|_{f=\text{const}} \leq 0 \quad (\text{A5})$$

ここで,

p : 予測者の真の信念に基づく確率 (予測者は, この確率で事象が発生すると考えている)

f : 予測値 (予測者は, 実際にはこの確率を予測する)

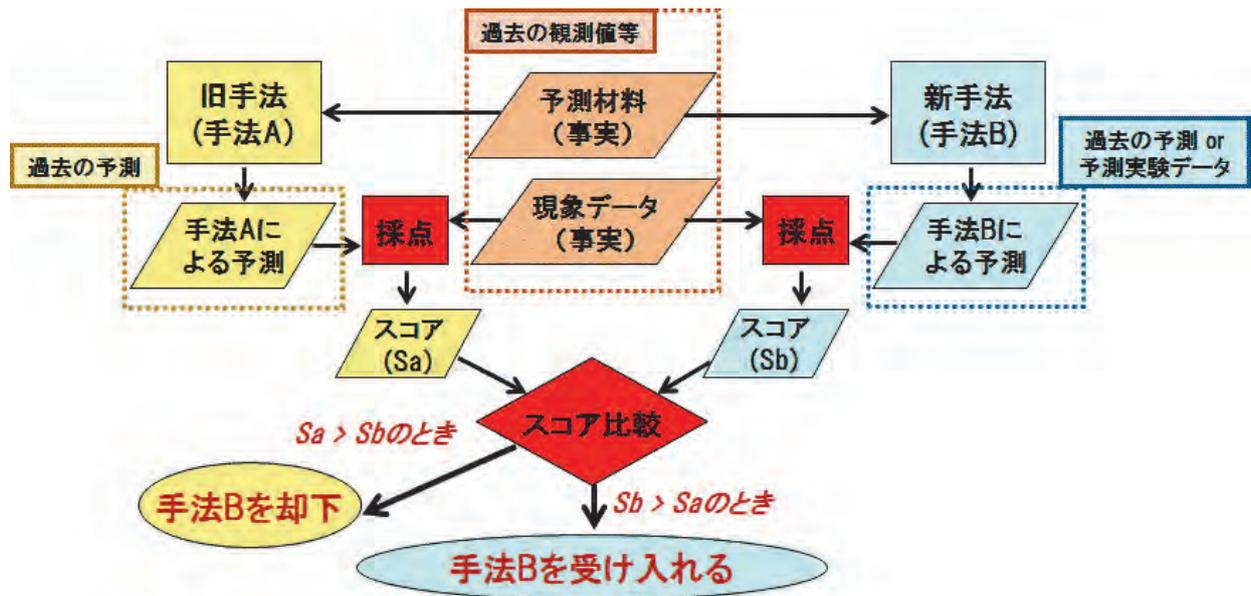
c : 基準予測値

i : 予測対象事象の発生の有無 (事象が発生すれば 1, 発生しなければ 0)

$S_i(f, c)$: スコア

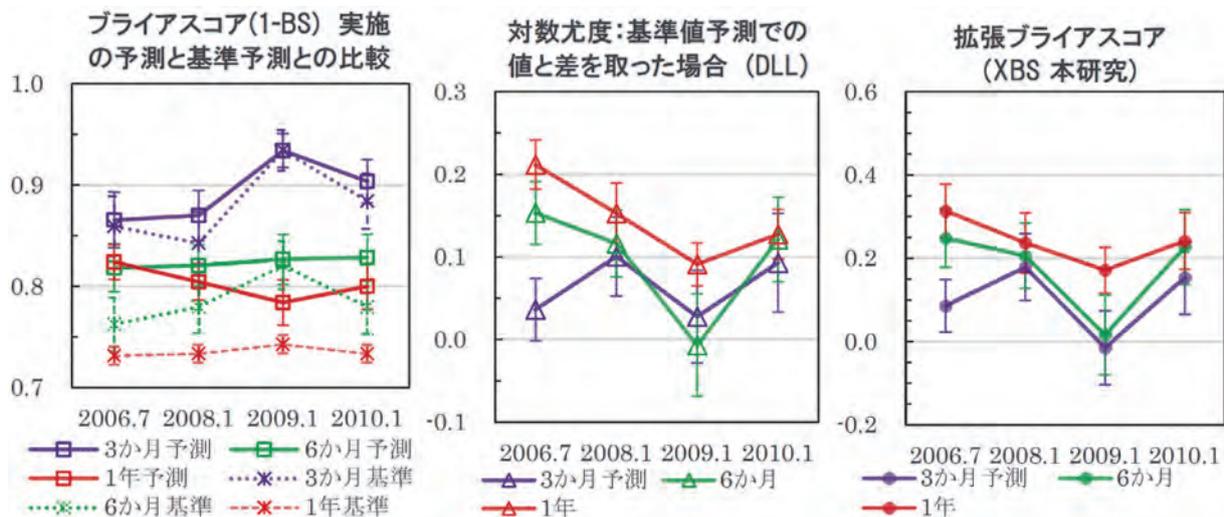
E_s : スコアの期待値

である.



第1図 予測に使う手法を改善する一般的な手順。この手順の繰り返して最大スコアを得る予測手法は見つかるが、採点方法に最も適応した予測手法だからといって、優れた手法とは限らない。

Fig. 1 A general flowchart to improve a prediction method. Repeating application of this chart enables us to find which prediction method get the highest score. However, the best-adjusted prediction method to the scoring rule does not always mean the most excellent one.



第2図 ブライアスコア・対数尤度・拡張ブライアスコアを用いた相似地震の確率予測実験5)の採点結果の比較。2006～2010年の4回、毎回90相似地震系列の「今後3か月・6か月・1年以内に当該相似地震が発生する確率」を予測する計1080回分を採点対象とした。

Fig. 2 Comparison of evaluations of the stochastic prediction on small repeating earthquakes by Brier's score (BS), log-likelihood (LL) and Extended Brier's Score (XBS). Total 1080 prediction values (3 kind of prediction time windows: 3 months, 6 months and 1 year long; at 4 times during 2006 - 2010; 90 series of repeating earthquakes) are evaluated.