

評価手法の説明

地震活動評価に基づく地震発生予測研究グループ（事務局、東京大学地震研究所内同研究グループ）

文責：東京大学地震研究所 楠城一嘉、平田 直、鶴岡 弘

最終更新：2010年7月

1. はじめに

このウェブページでは、地震発生予測検証実験に使用されている評価手法を簡単に日本語で説明する。本実験は原則的に CSEP (Collaboratory for the Study of Earthquake Predictability) プロジェクトのルール、及び先行研究である RELM (Regional Earthquake Likelihood Models) プロジェクトのルールと同じである。従って、評価手法の詳細は、[RELM 特集号](#)、[CSEP ウェブサイト](#)、及び "4. 参考文献" を参照されたい。

2. 評価手法

使用する評価法は、Schorlemmer et al. (2007) と Zecher et al. (2010) が提案した地震の確率予測を評価する統計的検定法である (Schorlemmer et al., 2010)。

予測空間(位置 + 時間 + マグニチュードからなる空間)を事前に多数の小領域に分割する。CSEP ではこの小領域を "bin" と呼び、各 bin ごとに発生率(または発生数)を予測する。その予測と観測結果から、一定の方式で bin ごとのスコアを計算し、それを予測空間全体について合計したものの、すなわち観測スコアを求める。

一方で、予測した発生率が正しいとしても、地震の決定精度等の偶然的な要因でスコアは絶えず変動するので、そのスコアのばらつきを理論的に求める必要がある。そこで CSEP では、シミュレーション(乱数実験)に頼る。予測した条件や発生率を満たす模擬地震を大量に発生させ、多数の地震カタログを作成し、そしてカタログ毎に模擬スコアを求める。多数の模擬スコアの分布から、スコアの理論的分布を推定する。もし実際の地震にもとづく観測スコアが分布の危険域(分布の裾に当たる部分)にあれば、予測は観測と矛盾する検定結果を得る。反対に、危険域になければ、予測と観測は矛盾しないとみなす。同一の観測地震に対して、複数の予測が存在する場合、検定結果の相違からモデル間の優劣が決まる。

CSEP で用いられている N(umber) テスト、M(agnitude) テスト、S(pace) テスト、L(ikelihood) テスト、R(atio) テストは、このようなアイデアをもとにしている。

2.1. N テスト

N テストでは、地震の予測総数と観測総数は整合するかが評価される。

地震総数(観測スコア: N)と、模擬スコアの分布を比較して、少なくとも N 個の地震が起きる確率(1)と、多くても N 個の地震が起きる確率(2)の二種類を求めることができる。もし 1 が非常に小さいならば、予測した地震数は観測した地震数より非常に少ないことを示す。つまり、過少予測を表す。一方、 2 が非常に小さいならば、予測した地震は非常に多いこと(過大

予測)を表す。これまでの研究では 0.025 より小さい値の α が得られたとき、過少予測とみなされている。同様に、 β が 0.025 より小さいとき、予測は過大であるとみなされている。

2.2. L テスト

これは、予測される地震数とマグニチュード、そして空間分布が、観測と整合するかを評価するテストである。

各 bin ごとに対数尤度 (log-likelihood) を計算して、予測空間全体について合計したもの (結合対数尤度: joint log-likelihood) が観測スコアとなる。このスコアは負の値をとるので、0 に近いほどよく合うことを示す。

シミュレーションによる模擬スコアの分布と観測スコアを比較するために、模擬スコアの分布の下位から数えて、観測スコアがどのくらいに位置しているかを表す指標、いわゆる分位数 ($0 \leq \alpha \leq 1$) を使用する。もし $\alpha = 0$ ならば分布の最下位を表し、 $\alpha = 1$ の時は最上位を示す。CSEP では、危険域は下位 2.5% としている。つまり、もし $\alpha < 0.025$ ならば、予測は観測を説明出来ないことになる。一方、 α が危険域に入らない場合は、予測と観測は矛盾していないとみなされる。ちなみに、 α が 0.975 より大きい方の分布の裾に入る場合 (上位 97.5% に入る) 予測が非常に良いことを示す。この場合もモデルは棄却されない。

2.3. S テスト

S テストは、地震の空間分布について予測と観測の整合性を評価する。

あるマグニチュード以上の地震発生率 (または発生数) を個々の bin で足し合わせ、その空間分布に関して予測と観測の整合性を検定する。L テストと同様に、模擬スコアの分布の下位から数えて、観測スコアはどのくらいに位置しているかを表す指標 (分位数: $0 \leq \alpha \leq 1$) を用いる。もし α が非常に小さいならば (0.025 より小さいならば)、観測された空間分布は予測分布と整合しないとみなされる。

2.4. M テスト

M テストは、予測される地震の規模別頻度分布が観測とあっているかを評価する。

S テストでは空間分布を考えたが、ここでは、マグニチュードの分布を考える。つまり、個々のマグニチュードについて、地震の発生率を全ての空間 bin で足し合わせる。S テストと同様に、もし指標 (分位数: $0 \leq \alpha \leq 1$) が非常に小さいならば (0.025 より小さいならば)、予測したマグニチュード分布は観測と合わないことを示す。これまでの研究によると (Zecher et al., 2010)、ほとんどケースで M テストに合格する。これは、グーテンベルグ・リヒター則もしくは、その改良形を用いた予測がほとんどであることによる。

2.5. R テスト

これは、二つの予測がある時に、基準と考える予測に対して、もう一つの予測が有意に良いかを検定する。

このテストでは、L テストを応用した結合対数尤度が観測スコアとなる。模擬スコアの分布は、基準となる予測を用いて得られる。これまでのテストと同様に、指標 (分位数: $0 \leq \alpha \leq 1$) に基づいて、模擬スコアの分布と観測スコアを比較する。 α が危険域 ($\alpha < 0.025$) にある時、基準となる予測より有意に良い予測とは言えないことを示す。これは片側検定なので、基準となる予測を

入れ替えて同様の検定をする。この場合の指標(分位数 : $0 \leq \dots \leq 1$)も と同様な危険域(< 0.025) を考える。

3. 参考文献

- Kagan, Y. Y., and D. D. Jackson (1994), Long-term probabilistic forecasting of earthquakes, *J. Geophys. Res.*, 99(B7), 13,685-13,700, doi:10.1029/94JB00500.
- D. Schorlemmer, M. C. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A. Rhoades (2007), Earthquake likelihood model testing, *Seismological Research Letters*, 78(1), 17-29, doi: 10.1785/gssrl.78.1.17.
- D. Schorlemmer et al. (2010), First results of the regional earthquake likelihood models experiment, *Pure Appl. Geophys.*, doi: 10.1007/s00024-010-0081-5.
- J. D. Zechar, M. C. Gerstenberger, and D. A. Rhoades (2010), Likelihood-based tests for evaluating space-rate-magnitude earthquake forecasts, *Bulletin of the Seismological Society of America*, 100(3), 1184-1195, doi: 10.1785/0120090192.
- 論文のダウンロードは[こちら](#) (要ユーザー名, 要パスワード)。

4. 問合せ先

〒113-0032 東京都文京区弥生1 - 1 - 1 東京大学地震研究所内「地震活動評価に基づく地震発生予測研究グループ」事務局及び、地震発生予測検証実験センター(平田直、鶴岡弘、楠城一嘉)

- e-mail: ZISINyosoku-submit@eri.u-tokyo.ac.jp
- ウェブページ : <http://www.eic.eri.u-tokyo.ac.jp/ZISINyosoku>