

付録B 統計的検証で利用される代表的な指標¹

B.1 平均誤差(Mean Error)

予報の平均的な偏りを示す指標で、系統誤差またはバイアスを示すために使われる。一般的に ME と記されることが多く、次の式で定義される。

$$ME = \frac{1}{N} \sum_{i=1}^N (x_i - a_i)$$

但し、 N は標本の数、 x は予報値、 a は真値である。数値予報では、真値として客観解析値や観測値が用いられることが多い。

ME が 0 に近いほど、予報の平均的な偏りが小さいことを示す。

B.2 平方根平均二乗誤差(Root Mean Square Error)

予報誤差の大きさを示す指標であり、一般的に RMSE と記される。前項で示された表記を用いて、次の式で定義される。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - a_i)^2}$$

RMSE が小さいほど、予報の精度が良いを示す。

B.3 アノマリー相関

気候値(または平年値)からのずれを平年偏差(アノマリー)と定義する。アノマリー相関は、予報と解析の平年偏差の相似の程度を示す指標であり、それぞれの平年偏差の相関係数で定義される。 c を気候値とし、その他の表記は第 B.1 項に従う。

$$\text{アノマリー相関} = \frac{\sum_{i=1}^N (X_i - \bar{X})(A_i - \bar{A})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (A_i - \bar{A})^2}}$$

但し、

$$X_i = x_i - c_i, X = \frac{1}{N} \sum_{i=1}^N X_i, A_i = a_i - c_i, A = \frac{1}{N} \sum_{i=1}^N A_i.$$

である。

アノマリー相関の最小値は-1、最大値は 1 であり、予報と解析の空間パターンが一致している場合には 1 となり、逆に全くパターンが反転している場合には-1となる。

B.4 スプレッド

アンサンブル予報のメンバーの広がり示す指標であり、一回のアンサンブル予報の分散の統計期間中の平均の平方根で定義される。

$$\text{スプレッド} = \sqrt{\frac{1}{N} \sum_{n=1}^N \left(\frac{1}{M} \sum_{m=1}^M (x_m - \bar{x}_n)^2 \right)}$$

但し、 M は一回のアンサンブル予報でのメンバー数、 N は統計期間中に行われたアンサンブル予報の数を表し、 x は予報値、 \bar{x} はアンサンブル平均値である。

アンサンブル予報のスプレッドは、アンサンブル平均の RMSE と同程度であることが好ましいとされている。

アンサンブル予報の検証についての詳細は、高野(2002)を参照されたい。

B.5 カテゴリー検証

B.5.1 分割表

ここでは、最も単純な 2 つの基準を用いた 2×2 の分割表を用いたカテゴリー検証について説明する。基準は、予報と実況のそれぞれに一つずつ設定することができるが、数値予報の検証の場合には同じ基準を用いることがほとんどである。分割表は、予報と実況がそれぞれ基準を満たす(「あり」)頻度と満たさない(「なし」)頻度から作成される(表 B.5.1)。各スコアは、表 B.5.1 の表記を用いて定義される。

表 B.5.1 2×2 の分割表。FO,FX,XO,XX はそれぞれの頻度数を表す

		実況	
		あり	なし
予報	あり	FO	FX
	なし	XO	XX

但し、 $N=FO+FX+XO+XX$ とする。

B.5.1 適中率

予報が適中した確率を示す指標であり、次のように定義される。

$$\text{適中率} = \frac{(FO + XX)}{N}$$

適中率は 0 から 1 の値をとり、1 であるほど予報の精度が良いことを示す。

B.5.2 空振り率

予報では「あり」であったが、実況では「なし」であった空振りの程度を示す指標であり、次のように定義される。

$$\text{空振り率} = \frac{FX}{(FO + FX)}$$

空振り率は 0 から 1 の値をとり、0 に近いほど空振りが少ないことを示す。

B.5.3 見逃し率

実況では「あり」であったが、予報では「なし」であった見逃しの程度を示す指標であり、次のように定義される。

¹ 本田有機

$$\text{見逃し率} = \frac{XO}{(FO + XO)}$$

見逃し率は0から1の値をとり、0に近いほど見逃しが少ないことを示す。

B.5.4 捕捉率

実況で「あり」だったときに、正しく「あり」と予報されていた確率を示す指標であり、次のように定義される。

$$\text{捕捉率} = \frac{FO}{(FO + XO)}$$

捕捉率は0から1の値をとり、1に近いほど予報の精度が良いことを示す。また、捕捉率が1に近いことは見逃しが少ないことも意味するが、空振りの頻度を推定することは出来ない。

B.5.5 バイアスコア

実況の「あり」の頻度に対する予報の「あり」の頻度の比を示す指標であり、次のように定義される。

$$\text{バイアスコア} = \frac{(FO + FX)}{(FO + XO)}$$

最小値は0、最大値は無限大である。予報と実況でそれぞれの基準を満たす頻度が一致する場合1となる。

B.5.6 スレットスコア

基準を満たす適中(FO)にのみ着目して予報精度を評価する指標で、次のように定義される。

$$\text{スレットスコア} = \frac{FO}{(FO + FX + XO)}$$

実況の「あり」の頻度に対して「なし」の頻度が非常に大きい基準について検証する場合、適中率では予報「なし」で実況「なし」の適中の頻度(XX)にスコアが大きく左右されてしまうことがある。スレットスコアを用いることで、(XX)の影響を排除し、上のような基準を用いた検証においても予報の精度を評価することができる。

スレットスコアは0から1の値をとり、1に近いほど予報の精度が良いことを示す。

B.5.7 スキルスコア

スキルスコアとは、ランダム予測による適中を除いて評価する指標を示す一般的な呼称である。ここでは、Heidkeのスキルスコアと呼ばれるものを紹介する。

Heidkeのスキルスコアは最も代表的なスキルスコアであり、ランダム予測による適中数を除いた適中率として定義される。

$$\text{スキルスコア} = \frac{(FO + XX) - Sc}{N - Sc}$$

但し、 S_c はランダム予測による適中数を示し、

$$Sc = \frac{(FO + XO)}{N} \times (FO + FX) + \frac{(FX + XX)}{N} \times (XO + XX)$$

である。

スキルスコアが0よりも小さい場合には、予報精度がランダム予測よりも悪いことを示す。

本研修テキストでは3×3の分割表を用いたスキルスコアを用いており、この場合は次のようになる。分割表は3つのカテゴリから作成される(表B.5.2)。

$$\text{スキルスコア} = \frac{(F_1O_1 + F_2O_2 + F_3O_3) - Sc}{N - Sc}$$

但し、

$$N = \sum_{j=1}^3 \sum_{i=1}^3 F_i O_j$$

$$Sc = \frac{(F_1O_1 + F_2O_1 + F_3O_1)}{N} \times (F_1O_1 + F_1O_2 + F_1O_3) + \frac{(F_1O_2 + F_2O_2 + F_3O_2)}{N} \times (F_2O_1 + F_2O_2 + F_2O_3) + \frac{(F_1O_3 + F_2O_3 + F_3O_3)}{N} \times (F_3O_1 + F_3O_2 + F_3O_3)$$

である。

また、本研修テキストでは使われていないが、スキルスコアにはエクイタブルスレットスコア(Equitable Threat Score)と呼ばれるものもあり、ランダム予測の適中数を除いたスレットスコアとして定義される。両スキルスコアの比較等については、本田(1998)を参照されたい。

表B.5.2 3×3の分割表。F_iO_jはそれぞれの頻度数を表す。

		実況		
		O ₁	O ₂	O ₃
予報	F ₁	F ₁ O ₁	F ₁ O ₂	F ₁ O ₃
	F ₂	F ₂ O ₁	F ₂ O ₂	F ₂ O ₃
	F ₃	F ₃ O ₁	F ₃ O ₂	F ₃ O ₃

参考文献

高野清治, 2002: アンサンブル予報の利用技術. 気象研究ノート第201号, 73-103.

本田有機, 1998: 雨の統計的検証. 平成10年度数値予報研修テキスト, 気象庁予報部, 30-40.