

## 付録A 統計的検証で利用される代表的な指標<sup>1</sup>

### A.1 平均誤差、平方根平均二乗誤差

予報誤差を表す基本的な指標として平均誤差 (Mean Error、一般に ME、バイアスまたは系統誤差と記される) と平方根平均二乗誤差 (Root Mean Square Error、一般に RMSE と記される) がある。これらは次式で定義される。

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

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

ここで、 $N$  は標本数、 $x_i$  は予報値、 $a_i$  は実況値 (真値) である (実況値は客観解析値や観測値で近似されることが多い)。ME は予報値の実況値からの偏りの平均である。また、RMSE は最小値 0 に近いほど予報が実況に近いことを示す。なお、RMSE は ME からの寄与を分離して、

$$RMSE^2 = ME^2 + \sigma_e^2$$

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

と表すことができる。ここで  $\sigma_e$  はランダム誤差の標準偏差と解釈される。

### A.2 アノマリー相関係数

アノマリー相関係数 (Anomaly Correlation Coefficient、ACC、一般にアノマリー相関、アノマリー相関スコア、アノマリー相関値とも記される) とは予報値の基準値からの偏差 (アノマリー) と実況値の基準値からの偏差との相関係数であり、次式で定義される。

$$ACC \equiv \frac{\sum_{i=1}^N (X_i - \bar{X})(A_i - \bar{A})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \sum_{i=1}^N (A_i - \bar{A})^2}}$$

$$(-1 \leq ACC \leq 1)$$

ただし、

$$X_i = x_i - c_i, \quad \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$$

$$A_i = a_i - c_i, \quad \bar{A} = \frac{1}{N} \sum_{i=1}^N A_i$$

である。ここで、 $N$  は標本数、 $x_i$  は予報値、 $a_i$  は実況値、 $c_i$  は基準値である。なお、基準値としては平年値 (気候値) が用いられることが多い。アノマリー相関は予報と実況の基準値からの偏差の相似の度合いを示し、両者の空間パターンが一致している場合には最大値 1 をとり、逆に全くパターンが反転している場合には最小値 -1 をとる。

### A.3 スプレッド

アンサンブル予報のメンバーの広がりを示す指標であり、次式で定義される。

$$\text{スプレッド} \equiv \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{1}{M} \sum_{m=1}^M (x_{mi} - \bar{x}_i)^2 \right)}$$

ここで、 $M$  はアンサンブル予報のメンバー数、 $N$  は標本数、 $x_{mi}$  は  $m$  番目のメンバーの予報値、 $\bar{x}_i$  は

$$\bar{x}_i \equiv \frac{1}{M} \sum_{m=1}^M x_{mi}$$

で定義されるアンサンブル平均である。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 (高野 2002)。

### A.4 カテゴリー検証

カテゴリー検証では、まず、対象となる現象の「あり」、「なし」を判定する基準に基づいて予報と実況それぞれにおける現象の有無を判定し、その結果により標本を分類する。そして、それぞれのカテゴリーに分類された頻度数をもとに予報の特性を検証する。

#### A.4.1 分割表

分割表はカテゴリー検証においてそれぞれのカテゴリーに分類された頻度数を示す表である (表 A.4.1)。各スコアは、表 A.4.1 に示される各区分の頻度数を用いて定義される。

また、以下では全事例数を  $N = FO + FX + XO + XX$  実況「現象あり」の頻度数を  $M = FO + XO$  で表す。

#### A.4.2 適中率

$$\text{適中率} \equiv \frac{FO + XX}{N} \quad (0 \leq \text{適中率} \leq 1)$$

表 A.4.1 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

<sup>1</sup> 藤田 匡

適中率 (Percent Correct、Proportion Correct) は予報が適中した割合である。最大値 1 に近いほど予報の精度が高いことを示す。

#### A.4.3 空振り率

$$\text{空振り率} \equiv \frac{FX}{FO+FX} \quad (0 \leq \text{空振り率} \leq 1)$$

空振り率 (False Alarm Ratio) は、予報「現象あり」の頻度数に対する空振り(予報「現象あり」、実況「現象なし」)の割合である。最小値 0 に近いほど空振りが少ないことを示す。

#### A.4.4 見逃し率

$$\text{見逃し率} \equiv \frac{XO}{FO+XO} \quad (0 \leq \text{見逃し率} \leq 1)$$

見逃し率 (Miss Rate、Frequency of Misses) は、実況「現象あり」の頻度数 ( $M = FO + XO$ ) に対する見逃し(実況「現象あり」、予報「現象なし」)の割合である。最小値 0 に近いほど見逃しが少ないことを示す。

#### A.4.5 捕捉率

$$\text{捕捉率} \equiv \frac{FO}{FO+XO} \quad (0 \leq \text{捕捉率} \leq 1)$$

捕捉率 (Probability of Detection、Prefigurance、適中率と訳されることもある) は、実況「現象あり」であったときに予報が適中した割合である。最大値 1 に近いほど見逃しが少なく予報の精度が高いことを示す。ただし、この指標から空振りの頻度 ( $FX$ ) を推定することは出来ない。ROC 曲線(第 A.5.5 項)のプロットに用いられ、この場合一般に Hit Rate と記される。

#### A.4.6 False Alarm Rate

False Alarm Rate (Probability of False Detection) と呼ばれる、空振り率と訳されることもある) は実況「現象なし」であったときに予報が外れた割合である。

$$Fr \equiv \frac{FX}{FX+XX} \quad (0 \leq Fr \leq 1)$$

最小値 0 に近いほど空振りの予報が少なく予報の精度が高いことを示す。ROC 曲線(第 A.5.5 項)のプロットに用いられる。第 A.4.3 項の空振り率とは分母が異なる。

#### A.4.7 バイアスコア

バイアスコア(Bias、Frequency Bias) は実況「現象あり」の頻度数に対する予報「現象あり」の頻度数の比である。バイアスコア  $B$  は次式で定義される。

$$B \equiv \frac{FO+FX}{FO+XO} \quad (B \geq 0)$$

予報と実況で「現象あり」の頻度数が一致する場合 1 となる。1 より大きいほど予報の「現象あり」の頻度過多、1 より小さいほど予報の「現象あり」の頻度過少である。

#### A.4.8 気候学的出現率

現象の気候学的出現率  $P_c$  (一般に、(単に)現象の出現率、現象の出現相対頻度、Sample Climatology、Sample Climate、Climatological Probability、Sample Relative Frequency、Event Frequency、Base Rate などと呼ばれる) は、標本から見積もられる現象の平均的な出現確率であり、次式で定義される。

$$P_c \equiv \frac{M}{N}$$

この量は実況のみから決まり、予報の精度にはよらない。予報の精度を評価する基準を設定する際にしばしば用いられる。

#### A.4.9 スレットスコア

スレットスコア (Threat Score、TS、Critical Success Index) と呼ばれる) は「現象あり」の場合の予報適中頻度数( $FO$ ) に着目して予報精度を評価する指標であり、次式で定義される。

$$TS \equiv \frac{FO}{FO+FX+XO} \quad (0 \leq TS \leq 1)$$

出現頻度の小さい現象 ( $XX \gg FO, FX, XO$ ) について  $XX$  の影響を除いて検証するのに有効である。最大値 1 に近いほど予報の精度が高いことを示す。なお、スレットスコアは現象の気候学的出現率の影響を受けやすく、例えば異なる環境下で行われた予報の比較には適さない。この問題を緩和するため次項のエクイタブルスレットスコアなどが考案されている。

#### A.4.10 エクイタブルスレットスコア

エクイタブルスレットスコア (Equitable Threat Score、ETS、Gilbert Skill Score) と呼ばれる) は気候学的な確率で「現象あり」が適中した頻度を除いて予報精度を評価する指標であり、次式で定義される(Schaefer 1990)。

$$ETS \equiv \frac{FO - S_f}{FO + FX + XO - S_f} \quad \left(-\frac{1}{3} \leq ETS \leq 1\right)$$

ただし、

$$S_f = P_c(FO + FX), \quad P_c = \frac{M}{N}$$

である。ここで、 $P_c$  は現象の気候学的出現率(第 A.4.8 項)、 $S_f$  は現象の「あり」、「なし」をランダムに  $FO + FX$  回予報した場合(ランダム予報)の「現象あり」の適中頻度数である。最大値 1 に近いほど予報の精度が高いこ

とを示す。ランダム予報で 0 となる。また、 $FO = XX = 0$ 、 $FX = XO = N/2$  の場合に最小値  $-1/3$  をとる。

## A.5 確率予報に関する指標

### A.5.1 ブライアスコア

ブライアスコア(Brier Score、BS) は確率予報の統計検証の基本的指標である。ある現象の出現確率を対象とする予報について、次式で定義される。

$$BS \equiv \frac{1}{N} \sum_{i=1}^N (p_i - a_i)^2 \quad (0 \leq BS \leq 1)$$

ここで、 $p_i$  は確率予報値(0 から 1)、 $a_i$  は実況値(現象ありで 1、なしで 0)、 $N$  は標本数である。BS は完全に適中する決定論的な( $p_i=0$  または 1 の)予報(一般に完全予報と呼ばれる)で最小値 0 をとり、0 に近いほど予報の精度が高いことを示す。また、現象の気候学的出現率  $P_c = M/N$  ( $M$  は実況「現象あり」の頻度数、第 A.4.8 項参照)を常に確率予報値とする予報(一般に気候値予報と呼ばれる)のブライアスコア  $BS_c$  は

$$BS_c = P_c(1 - P_c)$$

となる。ブライアスコアは現象の気候学的出現率の影響を受けるため、異なる標本や出現率の異なる現象に対する予報の精度を比較するのには適さない。例えば上記  $BS_c$  は  $P_c$  依存性を持ち、同じ予報手法(ここでは気候値予報)に対しても  $P_c$  の値に応じて異なる値をとる(Stanski et al. (1989) など)。次項のブライアスキルスコアはこの問題を緩和するため気候値予報を基準にとり、そこからのブライアスコアの変化によって予報精度を評価する。

### A.5.2 ブライアスキルスコア

ブライアスキルスコア (Brier Skill Score、BSS) はブライアスコアに基づいた指標であり、気候値予報を基準とした予報の改善の度合いを示す。ブライアスコア BS、気候値予報によるブライアスコア  $BS_c$  を用いて

$$BSS \equiv \frac{BS_c - BS}{BS_c} \quad (BSS \leq 1)$$

で定義される。完全予報で 1、気候値予報で 0、気候値予報より誤差が大きいと負となる。

### A.5.3 Murphy の分解

Murphy (1973) は、ブライアスコアと予報の特性との関連を理解しやすくするため、ブライアスコアを信頼度(reliability)、分離度(resolution)、不確実性

(uncertainty) の 3 つの項に分解した。これを Murphy の分解と呼ぶ(高野 (2002) などに詳しい)。

確率予報において、確率予報値を  $L$  個の区間に分け、標本を確率予報値の属する区間に応じて分類することを考える。確率予報値が  $l$  番目の区間に属する標本数を  $N_l$  ( $N = \sum_{l=1}^L N_l$ )、このうち実況が「現象あり」であった頻度数を  $M_l$  ( $M = \sum_{l=1}^L M_l$ ) とすると、Murphy の分解によりブライアスコアは以下のように表される(確率予報値の  $l$  番目の区間の区間代表値を  $p_l$  とする)。

$$BS = \text{信頼度} - \text{分離度} + \text{不確実性}$$

$$\text{信頼度} = \sum_{l=1}^L \left( p_l - \frac{M_l}{N_l} \right)^2 \frac{N_l}{N}$$

$$\text{分離度} = \sum_{l=1}^L \left( \frac{M_l}{N} - \frac{M_l}{N_l} \right)^2 \frac{N_l}{N}$$

$$\text{不確実性} = \frac{M}{N} \left( 1 - \frac{M}{N} \right)$$

信頼度は確率予報値( $p_l$ )と実況での現象出現相対頻度( $M_l/N_l$ )が一致すれば最小値 0 となる。分離度は確率予報値に対応する実況での現象の出現相対頻度( $M_l/N_l$ )が気候学的出現率( $P_c = M/N$ )から離れているほど大きい値をとる。不確実性は現象の気候学出現率が  $P_c = 0.5$  の場合に最大値 0.25 をとる。この項は実況のみによって決まり、予報の手法にはよらない。また、不確実性 =  $BS_c$  が成り立つ。これらを用いてブライアスキルスコアを次のように書くことができる。

$$BSS = \frac{\text{分離度} - \text{信頼度}}{\text{不確実性}}$$

### A.5.4 確率値別出現率図

確率値別出現率図 (Reliability Diagram、Attributes Diagram と呼ばれる)は、予報された現象出現確率  $P_{fcst}$  を横軸に、実況で現象が出現した相対頻度  $P_{obs}$  を縦軸にとり、確率予報の特性を示した図である(図 A.5.1 参照、Wilks (1995) などに詳しい)。一般に、確率予報の特性は確率値別出現率図上で曲線として表される。この曲線を信頼度曲線 (Reliability curve) と呼ぶ。

信頼度曲線の特性は、Murphy の分解(第 A.5.3 項)の信頼度、分離度と関連付けることができる。横軸  $P_{fcst}$  の各値について、信頼度(あるいは分離度)への寄与は、信頼度曲線上の点から対角線  $P_{obs} = P_{fcst}$  上の点(あるいは直線  $P_{obs} = P_c$  上の点)までの距離の二乗として表現される。 $P_{fcst}$  の各値でのこれらの寄与を、標本数

に比例する重みで平均して信頼度(あるいは分離度)が得られる。例えば、no-skill line (直線  $P_{obs} = (P_{fcst} + P_c)/2$ ) 上の点では、信頼度と分離度への寄与は等しい大きさを持ち、ブライアスキルスコアへの寄与が 0 となる。また no-skill line と直線  $P_{fcst} = P_c$  との間の領域(分離度への寄与 > 信頼度への寄与、図 A.5.1 灰色の領域)内に位置する点は、ブライアスキルスコアに正の寄与を持つ。

特別な場合として、気候値予報(第 A.5.1 項参照)では 1 点  $(P_{fcst}, P_{obs}) = (P_c, P_c)$  が信頼度曲線に対応する。また、次の 2 つの特性を示す確率予報は精度が高い。

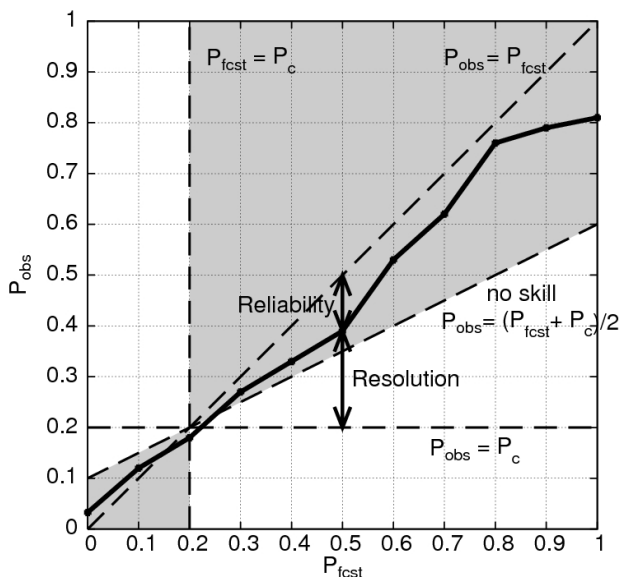


図 A.5.1 確率値別出現率図の模式図。横軸は予報現象出現確率、縦軸は実況現象出現相対頻度、実線が信頼度曲線である。対角線、直線  $P_{obs} = P_c$  からの距離の二乗がそれぞれ信頼度 (Reliability)、分離度 (Resolution) への寄与に対応している。灰色の領域内の点はブライアスキルスコアに正の寄与を持つ。

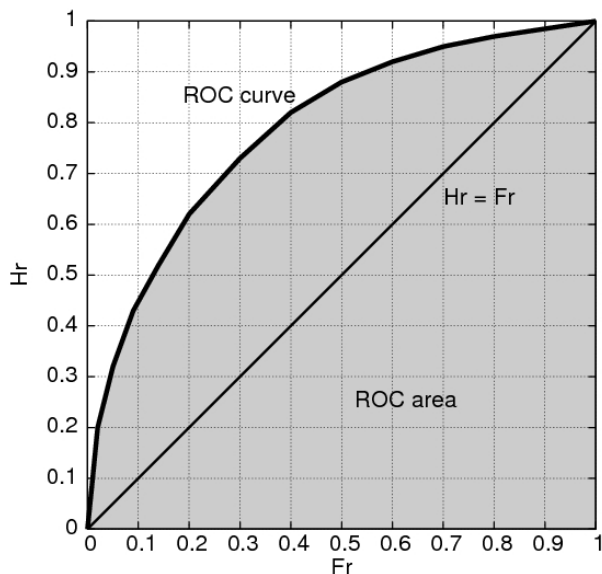


図 A.5.2 ROC 曲線の模式図。横軸は Fr、縦軸は Hr である。灰色の領域の面積が ROC 面積である。

- 信頼度曲線が対角線に(信頼度が最小値 0 に)近い。
- 信頼度曲線上の大きい標本数に対応する点が点  $(P_{fcst}, P_{obs}) = (P_c, P_c)$  (気候値予報)から離れた位置(確率値別出現率図の左下または右上寄り)に分布する(分離度が大きい)。

### A.5.5 ROC 面積スキルスコア

確率予報では、現象の予報出現確率にある閾値を設定し、これを予報の「現象あり」「現象なし」を判定する基準とすることが可能である。さまざまな閾値それぞれについて作成した分割表をもとに、閾値が変化したときの  $Fr - Hr$  平面(ここで  $Fr$  は False Alarm Rate(第 A.4.6 項)、 $Hr$  は Hit Rate(第 A.4.5 項))上の軌跡をプロットしたものが ROC 曲線(相対作用特性曲線、Relative Operating Characteristic curve、ROC curve)である(図 A.5.2 参照、高野(2002)などに詳しい)。平面内の左上方の領域では  $Hr > Fr$  であり、平面の左上側に膨らんだ ROC 曲線特性を持つ確率予報ほど精度が高いと言える。従って、ROC 曲線から下の領域(図 A.5.2 灰色の領域)の面積(ROC 面積、ROC area、ROCA)は情報価値の高い確率予報ほど大きくなる。ROC 面積スキルスコア(ROC Area Skill Score、ROCASS)は情報価値のない予報( $Hr = Fr$ )を基準として ROC 面積を評価するものであり、次式で定義される。

$$ROCASS \equiv 2(ROCA - 0.5) \quad (-1 \leq ROCASS \leq 1)$$

完全予報で最大値 1 をとる。また、情報価値のない予報(例えば、区間 [0,1] から一様ランダムに抽出した値を確率予報値とする予報など)で 0 となる。

### 参考文献

- 高野清治, 2002: アンサンブル予報の利用技術. アンサンブル予報, 気象研究ノート, **201**, 73-103.
- Murphy, A. H., 1973: A new vector partition of the probability score. *J. Appl. Met.*, **12**, 595-600.
- Schaefer, J. T., 1990: The critical success index as an indicator of warning skill. *Wea. Forecasting*, **5**, 570-575.
- Stanski, H. R., L. J. Wilson, and W. R. Burrows, 1989: Survey of common verification methods in meteorology. *Research Report No. (MSRB) 89-5*, Forecast Research Division, Atmospheric Environment Service, Environment Canada.
- Wilks, D. S., 1995: *Statistical Methods in the Atmospheric Science; an introduction, International Geophysical Series vol. 59*. Academic Press, 464pp.