

付録B 数値予報研修テキストで用いた表記と統計的な指標¹

本テキストで使用した表記と統計的な指標などについて以下に説明する。

B.1 研修テキストで用いた表記について

B.1.1 時刻の表記について

研修テキストでは、時刻を表記する際に、通常国内で用いられている日本標準時(Japan Standard Time、JST)の他に、協定世界時(Coordinated Universal Time、UTC)を用いている。

数値予報では国際的な観測データの交換やプロダクトの利用等の利便を考慮して、時刻はUTCで表記されることが多い。JSTはUTCに対して9時間進んでいる。

B.1.2 予報時間について

数値予報では、統計的な検証や事例検証の結果を示す際に、予報対象時刻の他に、初期時刻からの経過時間を予報時間(Forecast Time-period、FT)として表記している。

このテキストでは、予報時間を

「予報時間」=「予報対象時刻」-「初期時刻」

で定義し、例えば、6時間予報の場合、FT=6と表記しており、時間の単位(h)を省略している。

B.2 検証に用いた基本的な指標

B.2.1 平均誤差、平方根平均二乗誤差、誤差の標準偏差 (テキスト内で利用多数)

予報誤差を表す基本的な指標として平均誤差(Mean Error、ME、バイアスと表記した節もある)と平方根平均二乗誤差(Root Mean Square Error、RMSE)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 は実況値である(実況値は客観解析値、初期値や観測値が利用されることが多い)。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の寄与と

それ以外を分離して、

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

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

と表すことができる。 σ_e はランダム誤差である。

B.2.2 アノマリー相関係数 (第1.4節で利用)

アノマリー相関係数(Anomaly Correlation Coefficient、ACC)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。

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

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。第1.4節では基準値として気候値(複数年のデータを使った平均値)を用いている。アノマリー相関は予報と実況の基準値からの偏差の相関を示し、基準値からの偏差の増減のパターンが一致している場合には最大値の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の-1をとる。

B.2.3 スプレッド (第1.4節、第1.5節で利用)

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、次式で定義する。

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

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。

¹ 坂本 雅巳(平成20年度数値予報研修テキスト編集担当)

B.3 カテゴリー検証で用いた指標など

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事例数をもとに予報の特性を検証する。

B.3.1 分割表(第 1.1 節で利用)

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事例数を示す表である(表 B.3.1)。各スコアは、表 B.3.1 に示される各区分の事例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」の事例数を $M = FO + XO$ 、実況「現象なし」の事例数を $X = FX + XX$ で表す。

表 B.3.1 分割表。FO、FX、XO、XX はそれぞれの頻度数を表す。

		実況		計
		あり	なし	
予報	あり	FO	FX	FO+FX
	なし	XO	XX	XO+XX
計		M	X	N

B.3.2 適中率 (第 3.1 節、第 3.2 節で利用)

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

適中率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。

B.3.3 空振り率 (第 2.1 節、第 3.1 節で利用)

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

空振り率は、予報「現象あり」の事例数に対する空振り(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。このテキストでは分母を $FO + FX$ としているが、代わりに N として定義する場合もある。

B.3.4 見逃し率 (第 2.1 節で利用)

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

見逃し率は、実況「現象あり」の事例数 ($M = FO + XO$) に対する見逃し(実況「現象あり」、予報「現象なし」)の割合²である。最小値 0 に近いほど見逃しが少ないことを示す。

² 分母を M とする代わりに、 N として定義する場合もある。

B.3.5 捕捉率 (第 3.1 節、第 3.3 節で利用)

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

捕捉率は、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いことを示す。ROC 曲線(第 B.4.5 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。

B.3.6 誤検出率 (第 3.7 節で利用)

誤検出率(False Alarm Rate)は実況「現象なし」であったときに予報が外れた割合であり、第 B.3.3 項の空振り率とは分母が異なる。

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

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 B.4.5 項)のプロットに用いられる。

B.3.7 バイアスコア (利用多数)

バイアスコア(Bias Score, BI)は実況「現象あり」の事例数に対する予報「現象あり」の事例数の比であり、次式で定義される。

$$BI \equiv \frac{FO + FX}{M} \quad (0 \leq BI)$$

予報と実況で「現象あり」の事例数が一致する場合 1 となる。1 より大きいほど予報の「現象あり」の頻度過大、1 より小さいほど予報の「現象あり」の頻度過小である。

B.3.8 体積率 (第 3.3 節で利用)

体積率は全事例数に対する予報の「現象あり」の割合である。

$$\text{体積率} \equiv \frac{FO + FX}{N}$$

体積率が等しければ捕捉率(第 B.3.5 項)が大きいほど良い予報(見逃しが少ない)であり、捕捉率が等しければ体積率が小さいほど良い予報(空振りが少ない)である。

B.3.9 気候学的出現率 (第 3.3 節で利用)

現象の気候学的出現率 P_c 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。

B.3.10 スレットスコア (利用多数)

スレットスコア (Threat Score, TS) は予報、または、実況で「現象あり」の場合の予報適中事例数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。

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

出現頻度の低い現象 ($N \gg M$ 、従って、 $XX \gg FO, FX, XO$ となって、予報「現象なし」による寄与だけで適中率が1になる現象) 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精度比較には適さない。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。

B.3.11 エクイタブルスレットスコア (利用多数)

エクイタブルスレットスコア (Equitable Threat Score, ETS) は気候学的な確率で「現象あり」が適中した頻度を除いて求めたスレットスコアであり、次式で定義される (Schaefer 1990)。

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

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 B.3.9 項)、 S_f は「現象あり」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」の適中事例数である。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。

B.3.12 スキルスコア (第 3.3 節で利用)

スキルスコア (Skill Score, Heidke Skill Score) は気候学的な確率で「現象あり」および「現象なし」が適中した頻度を除いて求める適中率であり、次式で定義される。

$$Skill = \frac{FO + XX - S}{N - S} \quad (-1 \leq Skill \leq 1)$$

ただし、

$$S = Pm_c(FO + FX) + Px_c(XO + XX),$$

$$Pm_c = \frac{M}{N}, \quad Px_c = \frac{X}{N}$$

である。ここで、 Pm_c は「現象あり」、 Px_c は「現象なし」の気候学的出現率 (第 B.3.9 項)、 S は現象の「あり」を

$FO + FX$ 回 (すなわち、「なし」を残りの $XO + XX$ 回) ランダムに予報した場合 (ランダム予報) の適中事例数である。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 をとる。

B.3.13 $n \times n$ 分割表とスコア (第 3.1 節で利用)

表 B.3.1 の分割表では、事象を 3 個以上のカテゴリに分類する場合に対応できない。このとき、カテゴリの個数が n であれば、表 B.3.1 を $n \times n$ の分割表に拡張したものを使用する (表 B.3.2)。例えば、天気予報を「晴れ」、「曇り」、「降水あり」の 3 つのカテゴリに分類して評価する場合は、 3×3 分割表 ($n=3$) を用いる。全事例数を

$$N_n = \sum_{j=1}^n \sum_{i=1}^n F_i O_j$$

とすると、第 3.1.4 項で利用されている適中率 (第 B.3.2 項) は、 $n \times n$ 分割表の場合以下のようになる。

$$\text{適中率}(n \times n) = \frac{\sum_{i=1}^n F_i O_i}{N_n}$$

表B.3.2 $n \times n$ 分割表。 $F_i O_j$ はそれぞれの事例数を表す。

		実況			
		O_1	O_2	...	O_n
予報	F_1	$F_1 O_1$	$F_1 O_2$...	$F_1 O_n$
	F_2	$F_2 O_1$	$F_2 O_2$...	$F_2 O_n$

	F_n	$F_n O_1$	$F_n O_2$...	$F_n O_n$

B.4 確率予報に関する指標など

B.4.1 ブライアスコア (第 3.1 節で利用)

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。

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

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 (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

$P_c = M/N$ (第 B.3.9 項) を常に確率予報値とする予報 (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には適さない。例えば上の 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)。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。

B.4.2 ブライアスキルスコア (第 1.4 節、第 3.1 節、第 3.2 節で利用)

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

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

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。

B.4.3 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」であった事例数を 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)。

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

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

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

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

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。

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

B.4.4 確率値別出現率図 (第 3.1 節、第 3.2 節で利用)

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 B.4.1 参照、Wilks (2006) 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 B.4.3 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

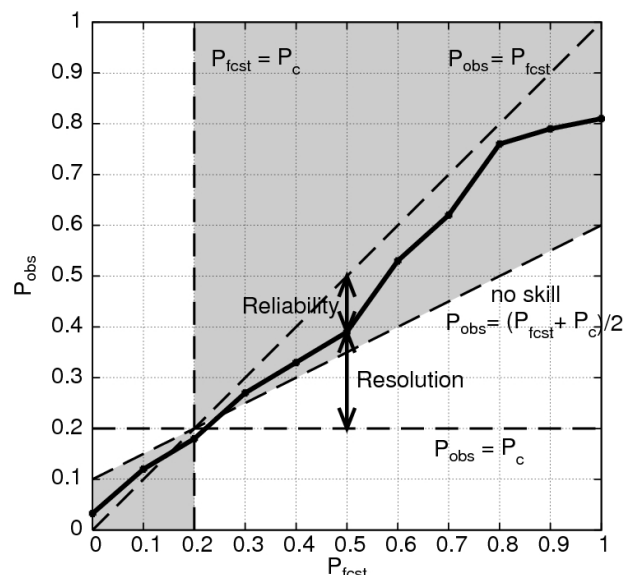


図 B.4.1 確率値別出現率図の模式図。横軸は予報現象出現確率、縦軸は実況現象出現相対頻度、実線が信頼度曲線である。対角線、直線 $P_{obs} = P_c$ との差の二乗がそれぞれ信頼度 (Reliability)、分離度 (Resolution) への寄与に対応している。灰色の領域内の点はブライアスキルスコアに正の寄与を持つ。

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 B.4.1 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 B.4.1 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。

- ・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）。

B.4.5 ROC 面積スキルスコア（第 1.4 節、第 2.2 節で利用）

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 は第 B.3.6 項の False Alarm Rate、 Hr は第 B.3.5 項の Hit Rate）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 B.4.2 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 B.4.2 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。

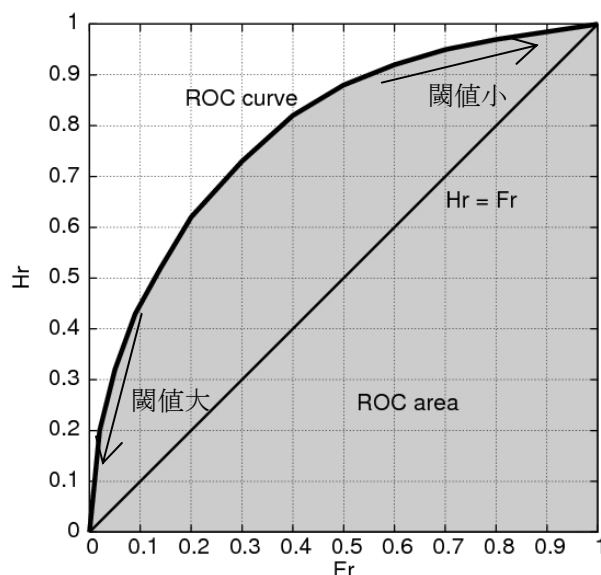


図 B.4.2 ROC 曲線の模式図。横軸は Fr 、縦軸は Hr である。灰色の領域の面積が ROC 面積である。

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., 2006: Statistical Methods in the Atmospheric Sciences Second Edition. *International Geophysical Series vol.91*. Academic Press, 648pp.