

確率的気象予測の問題点*

阿部克也**

要旨

近年、「予測は本来確率的である」との主張が強まり、気象予測においても「確率予報」がほとんど無批判のうちにこなわれている。しかし、気象現象の確率的予測は重大な実践的・科学的問題を含んでいる。ここでは次元の意味の分析により、「確率予測の検証不能性」が示され、同時に「予測モデルが宿命的に重要な誤差をもつこと」が明らかにされる。

1. はじめに

近年、電子計算機の発達に関連しつつ、気象の「確率予報」が大きな流れを作ろうとしている。米国では、力学的予測モデルとの組合せ(Model Output Statistics)の形で公共予報業務に組み込まれて久しく、わが国でも気象庁の業務として降水確率予報が一部の地域に対して実施されている。一般市民は「新たな気象サービス」として軽く受け取ると同時に「もう一つすっきりしない」と感じているようである(「朝日」1980年6月30日付夕刊、「読売」1981年3月15日付)。気象学・気象事業に関する人々の反応は、検証法などの曖昧さの故に、さらに複雑である。

「予測は本来確率的である」との主張は、数学的には完全であるが、量的予測を確率予測の思想圏に置くものであり、科学的に重要な問題をはらむとともに実践的にも種々の複雑な問題を引き起こすと思われる。ここでは確率予報が気象事業・気象学に及ぼす影響の大きさを考慮して、次元の意味の吟味に基いて、確率予報の基本的な問題点である「確率予測の検証不能性」および「確率予測モデルの宿命的誤差」を明らかにする試みを行いたい。

2. 確率予測と検証

最初に、ことさら新味はないかも知れないが、本稿で用いる術語「検証」と「評価」の意味を確認しておこう。

- (1) 検証は予測と実況の比較によってなされる。
- (2) 評価は利用者により、予測と実況および彼の利害や感情に基いてなされる。

「予測」と「予報」については、特に必要のない限り「予測」を用いる。

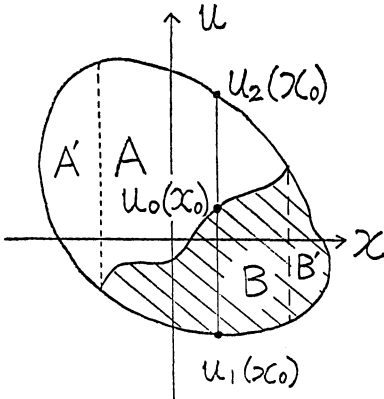
さて、ある現象の発生確率 p ($0 < p < 1$)、に対して現象が発生した場合、「確率 p の方が偶然出現した」と理解するのが妥当であり、 p と「現象の発生」の比較は無意味であるが、Brier (1950) の提案による「Brier Score (BS)」が確率予測の有効な検証法であるとの誤った「理解」が広がっている。BS は平均自乗誤差の確率予測への形式的適用であるが、その後異った検証法の提案もある(鈴木, 1981参照)。本章では次元の意味の確認を基に、確率予測の検証が不可能であることを示そう。

簡明のため降雨現象を例にとり「現象の有無」と「確率」の次元について考察する。

「降雨確率100%」は「雨が降る」と同義のように見られがちであるが「次元」を異にしていることに注目したい。すなわち、前者は「100回が100回とも(必ず)雨が降る」であり、「100回が100回とも」で「回」が帳消しになるように見えるが、「(必ず)」は消えない。「要するに雨が降る」という根拠により「降雨確率100%」と「雨が降る」を同義とすることは、「結果が同一なら経過も同一である」と言えないことと類似の理由により誤りで

* Basic Problems of Probabilistic Weather Prediction.

** Katsuya Abe, 高層気象台。
 —1981年6月25日受領—
 —1982年8月9日受理—



第1図 統計的確率モデルの概念図。部分空間Aで現象が発生し、Bでは現象は発生しない。AUBの外部は現象に無関係である。

ある。同様に「降雨確率0%」と「雨が降らない」も異なる「次元」を有する。さらに「降雨確率70%」が「雨が降る」とも「雨が降らないとも次元を異にすることは理解しやすい。そして、その「次元」は「降雨確率100%」や「降雨確率0%」と同一である。降雨確率は「特定の条件下で過去に降雨の発生した割合」に他ならないのである。したがってそれは「雨が降る」とも「雨が降らない」とも異質である。にもかかわらず、「降雨確率90%」という表現は「雨が降る」をわれわれに強く想起させることも否定できない。「明日が降らない方の10%の中に含まれている」可能性が不当に排除される「感覚」を筆者はうまく説明できない。

第1図に確率予測モデルの概念を図示するが、同図にしたがって上の「次元」の相違を一見容易に確認することができる。第4章と一部説明が重複するが、 x と u は「予測因子 (predictor) 空間」を作っている。そして x の周辺分布として(x を具体的予測因子とする) 確率予測モデルが定式化されているものとする。領域Aにおいて現象が発生し、領域Bでは現象が発生しない。 x の変化に伴い現象発生確率が変化し、 A' (点線から左) で100%、 B' (破線から右) で0%となっている。先の「雨が降る」は x, u 空間のAに対応し、「降雨確率100%」は同じく A' に対応している。明らかに $A' \subset A$ であり、また $B' \subset B$ である。さらに「確率70%」に対する雨なのか「確率30%」に対する雨であるかは降雨現象の観測によって定まるものではないことも明瞭である。降雨確率は予測因子空間の x 座標に対応し、「雨が降る」と

「雨が降らない」は2次元領域A, Bにそれぞれ対応している。

上述のように「次元」の異なる「確率」と「現象の発生」の比較の不合理性により、「確率予測」と「検証」は無縁なのである。検証は本来「予測」と「実況」以外の「余分な要素」を考慮しない。しかし「評価」となると事情が異なり、上の「余分な要素」に相当する「利害」等を含む。確率予測の検証 (verification) は不可能であるが、その評価 (valuation) は利用者によってなされる。しかし、評価は利用者の利害や感情にかかわり、必ずしも一定の方式ではなされない。評価という言葉は日常語としては「予報が当たったかどうか」、「予報の根拠である予報技術」および「予報が社会生活に役立ったかどうか」の三者に対して用いられる。この三様の「評価」の前二者の区別は簡単にはできないが、本稿の「検証」を現象予測の根拠となる気象要素の予測に適用することと予測手順の物理的検討の結合が「予報技術の評価」に近いであろう。しかし、日常語「評価」の第一と第三は明確に区別される。すなわち、本稿の術語の「検証」と「評価」がそれらである。

確率予測の検証法を模索する人々は「検証」と「評価」の厳密な区別をしていないようである。また、彼らは「雨が降る」⇔「降雨確率100%」⇔「降雨確率70%」における「次元」の変化を見落としているようである。例えば Brier (前出) は「従来の検証基準で好成績を得る予報が必ずしも高い価値をもたない」と言い、Murphy (1972) は「精度と価値の基準の関係」を論じて、「予報 r と実況 d により精度基準 $V(r, d)$ と価値基準 $U(r, d)$ を『適正』に定めるとき、 U が V の増加関数になる」としている。

次にBSの不合理性を具体的に見よう。

3. Brier Score の不合理性

BSは提案者 Brier の名にちなむ通称であり、原論文では次式の形で提案された。

$$P = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^m (f_{ij} - E_{ij})^2 \quad (1)$$

ここで j は予報日を示し、 i は現象 (予測対象) の階級を示す。また、 f_{ij} は予測確率であり、 E_{ij} は「 i 番目の階級」の発生と非発生に対応してそれぞれ「1」、「0」の値をとる。階級が2個 (例えば、現象の発生と非発生) のとき(1)は、 $f_{ij} + f_{2j} = 1$ 、 $E_{ij} + E_{2j} = 1$ から、

$$P = \frac{1}{n} \sum_{j=1}^n \{(f_{1j} - E_{1j})^2 + (f_{2j} - E_{2j})^2\}$$

$$= \frac{1}{n} \sum_{j=1}^n 2(f_{1j} - E_{1j})^2 \quad (2)$$

となり、添字「1」「2」が不要になる。 $p_j = f_{1j}$, $y_j = E_{1j}$ とおけば、

$$P/2 = \frac{1}{n} \sum_{j=1}^n (p_j - y_j)^2 \equiv \text{BS} \quad (3)$$

となり、(3) が half P-Score あるいは Brier Score と呼ばれている。

(3) の $(p_j - y_j)^2$ の代わりに $|p_j - y_j|$ が用いられたこともあるが、現在は元の $(p_j - y_j)^2$ が良いとされている。

$p_1 = p_2 = \dots = p_n = p_0$ のとき、 $y_j = 1$ の「 j 」の個数を n_0 とするとき、

$$\text{BS} = \frac{1}{n} \{(n - n_0)(p_0 - 0)^2 + n_0(p_0 - 1)^2\} \quad (4)$$

であり、 $n_0 = p_0 n$ ならば、さらに

$$\text{BS} = (1 - n_0/n)p_0^2 + (n_0/n)(1 - p_0)^2 = p_0(1 - p_0) \quad (5)$$

となる。(5) の $p_0(1 - p_0)$ は上の n 個の y_j の分散に等しく、一定の条件の下に、Brier 自身言っているように「 n 回の予報 p_j の幾つかが p_0 と異なる場合の BS は $p_0(1 - p_0)$ より大きい」が、これら 2 つの事は「検証法」としての BS を正当化するものではない。

ところで、 E_{ij} あるいは y_j に対して「1」、「0」を当てる必然性は全くなく、「確率が 0~1 の数で表される」から便宜上「1」、「0」が用いられるに過ぎない。ここに、「雨が降る」と「降雨確率 100%」が「次元」の相違を無視して安易に同義とされている不合理性が指摘される。

BS の延長として

$$\text{BS} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (p_{ij} - y_{ij})^2 \quad (6)$$

(j は予測日、 i は東京、千葉などの地点を示す) が現れたり、(6) に近いものとして、 y_j の代わりに「予報対象領域 A 中の現象発生部分 a_j と現象非発生部分 $(A - a_j)$ 」に対してそれぞれ $y_{1j} = 1$, $y_{2j} = 0$ を当てて、

$$\text{BS} = \frac{1}{\{a_j + (A - a_j)\}n} \sum_{i=1}^n \{a_j(p_j - y_{1j})^2 + (A - a_j)(p_j - y_{2j})^2\}$$

$$= \frac{1}{An} \sum_{j=1}^n \{a_j(p_j - 1)^2 + (A - a_j)(p_j - 0)^2\}$$

$$= \frac{1}{n} \sum_{j=1}^n \left\{ (p_j - a_j/A)^2 + \frac{a_j}{A}(1 - a_j/A) \right\} \quad (7)$$

が用いられている。しかし、これらも本質的に BS であり「次元」の相違は解消されていない。

4. 統計的確率モデルの誤差

確率モデルには、統計的に定められるものと物理法測によって定められるものがあるが、両者の区別は予測因子 (predictor) と予測対象 (predictand) を関係づける手段による。予測因子が観測される物理量であるか、あるいは物理法測に基いて計算される物理量であるか等は上の分類には無関係である。例えば、MOS 方式の確率予測モデルは統計的確率モデルである。本稿で扱う「確率予報」の「確率」が「数学的確率」とは若干意味が異なることは、確率予報を支持する人々も認めているが「どのように異なるのか」については語られていない。この章ではその相違を示し、同時に統計的確率モデルの誤差の性質を明らかにしたい。

第 2 章では直接には触れなかったが、「現象の発生」と「現象の非発生」は「次元」を異にしており、確率モデルの構成にはこの「次元」の差異を解消する(思考的)手続きが必要である。硬貨を自由落下させる際の「表」と「裏」の出現確率を考える場合は、初めからそれらは同「次元」であり、人為的に「Hitt」「Fault」と捉える段階で「次元」の差異が生まれる「硬貨の落下距離」と「硬貨を手から放すときの鉛直線との角度」の 2 者を変化させるとき、理想的な「条件付確率」のモデルが得られるが、「現象発生確率」も以下のように考えることにより「硬貨の自由落下」と同様に解釈される。

まず、単一(連続)予測因子に従うモデルを考えよう。ある現象の発生が予測因子 x と、気象学的には選定できないが「存在すると考えられる」未知の因子 u の結合により完全に説明されると仮定する。このとき、第 1 図に従って、現象発生と現象非発生に対応する点 (x, u) はそれぞれ領域 A, B に属し、領域 $A \cup B$ の外部はこの現象には無縁の領域である。また、 A, B は x の関数 $u_0(x), u_1(x), u_2(x)$ によって定まれているとする。 $u_0(x)$ などは陰関数あるいは図的表示を考えれば不連続(A, B が非連結)であってもかまわないが、ここでは第 1 図に従う。さて、標本に対応する点 (x, u) の分布(確率密度)を $\rho(x, u)$ とすれば ρ は気候の要素である。

$$\pi = \frac{\int_A \rho dx du}{\int_{A \cup B} \rho dx du}, \quad q(x) = \frac{\int_{u_0(x)}^{u_2(x)} \rho du}{\int_A \rho dx du},$$

$$r(x) = \frac{\int_{u_1(x)}^{u_0(x)} \rho du}{\int_B \rho dx du} \quad (8)$$

とおけば、 π は気候的現象発生率であり、 $q(x)$ 、 $r(x)$ はそれぞれ現象の発生・非発生に対する x の出現率である。これらにより、 x に対する現象発生確率 $p(x)$ は

$$p(x) = \frac{\pi q(x)}{\pi q(x) + (1-\pi)r(x)} \quad (9)$$

と表される（ベイズの定理）。

以上の設定は複数予測因子の場合にも、表現の煩雑さを別にすれば通用し、さらに u の代わりに「諸々の要因の複合としての u 」を設定することも可能である。この仮想的モデルでは「次元」の差異は解消されており、「硬貨の自由落下」との比較が可能である。

「硬貨の自由落下」のモデルの場合は望み次第に多くの実験を行うことが可能であり、したがって「一定の誤差」の範囲でモデルの現実的設定が可能である。しかし、気象現象の確率の場合は「実験（試行）」を行うことができない。予測因子空間を 100 個の部分空間に分割するとしても、各部分空間のおおのに対して 40 個の標本を得るには少なくとも 4,000 個の標本が必要であるが、1 日に 1 個の標本が得られるとき約 11 年でようやく 4,000 個の標本が得られるのである。そして、11 年の標本から定まる確率モデルの「誤差」は分布関数 ρ の気候的変動に大きく依存する。第 1 表は 1961 年～1980 年の高層気象台の地上資料による 2 年間雨日数の変動を示しているが、同表によれば (8)、(9) の π は平均で 0.262、最大で 0.294、最小で 0.228 となっている。この変動は分布 ρ に由来しており、 ρ の変動は当然ながら $q(x)$ 、 $r(x)$ をも変化させる。現象の発生・非発生に対する x の値から $q(x)$ 、 $r(x)$ を定める限りそれは避けられない。「実験」のきかない統計的モデルの宿命的弱点である。「誤差の大きさ」は「真の値」の存在なしには無意味であり、気象現象の発生確率には「真の値」は存在しない。先の π の値の変動の実際例にみられるように、相当多くの標本がある場合でさえ「真の値」ではなく、「変動幅の目安」が存在するに過ぎない。本稿での「宿命的誤差」は上の意味で用いられている。

5. 精度の尺度について

ところで、Brier（前出）らは「確率予報の精度の評価

第 1 表 館野における雨日数の変化。数字は 1 ミリ以上の雨（あるいは雪）の日数。

年 代	雨 日 数
1961～62	189
1963～64	193
1965～66	198
1967～68	210
1969～70	177
1971～72	193
1973～74	183
1975～76	193
1977～78	167
1979～80	214

の尺度として BS が妥当である」としているが、先の「次元」の矛盾の故に BS は精度の目安としても不合理である。小河原（1959）も言うように、「確率モデル」の精度の検証は相当数の独立標本により可能であるが、その方法は予測値ごとの現象発生率を調べるのが自然であろう。この観点から、BS の不合理さを示す例を挙げる。

モデル予測値が 20%、60% の場合がそれぞれ 10 回づつあったとしよう。これに対して、

$$BS = \frac{1}{20} \{2(1-0.2)^2 + 8(0-0.2)^2 + 6(1-0.6)^2 + 4(0-0.6)^2\} = 0.20 \quad (10)$$

が期待される。20%、60% の予報に対してそれぞれ 2 回と 6 回現象が発生したとすると、この予報は妥当であったと言えよう。つまり、「妥当であった予報」に「BS = 0.20」が対応する。ところで「現象なし」の 12 回、「現象あり」の 8 回に対してそれぞれ 0% と 50% の予報がなされる場合には 12 回に対して誤差なし、8 回に対して「8/8 - 0.5」の誤差ありということで「BS = 0.10」となる。さらに「現象なし」の 12 回と「現象あり」の 8 回に対してそれぞれ 20% と 60% を予報するときは 12 回に対して「0/12 - 0.2」、8 回に対して「8/8 - 0.6」の誤差が見られる一方「BS = 0.088」となる。すなわち、妥当であった予報の BS よりも小さな BS を示す予報が存在する。

上の例についても一つの観点から BS の不合理さが確認される。すなわち、20% の予報が 10 回なされる中で 2 回現象が発生する確率は ${}_{10}C_2 0.2^2 (1-0.2)^8 = 0.3020$ 、同様に 60% の予報 10 回に対して 6 回現象が発生する確率

は ${}_{10}C_6 0.6^6(1-0.6)^4=0.2508$ である。両者は独立であるから、先の第1の予報例に対して例のような実況が実現する確率は両者の積として、

$$\begin{aligned} &({}_{10}C_2 0.2^2(1-0.2)^8) \times ({}_{10}C_6 0.6^6(1-0.6)^4) \\ &= 0.0757 \end{aligned} \quad (11)$$

で与えられる。同様に第2、第3の予報例については

$$\begin{aligned} &({}_{12}C_{12} 0^{12}(1-0)^0) \times ({}_8C_8 0.5^8(1-0.5)^0) \\ &= 0.0039 \end{aligned} \quad (12)$$

$$\begin{aligned} &({}_{12}C_{12} 0.2^0(1-0.2)^{12}) \times ({}_8C_8 0.6^8(1-0.6)^0) \\ &= 0.0012 \end{aligned} \quad (13)$$

が(11)に相当する確率である。これら3つの数値はBSと同じ傾向で減少しているが、その意味は全く逆である。すなわち、「BSが良くなること」と「理論的により起こりにくい実況が実現すること」が対応している。実況は絶対的であるから、「実況の実現確率が小さいこと」は「予報が妥当でないこと」を表すことになる。サイコロに例えるならば、12回ふって「1の目」が2回出ることが実況で、サイコロが「6面体である」と言ったり「4面体である」と言ったりすることが予報に対応する。

$$\begin{aligned} &{}_{12}C_2(1/6)^2(5/6)^{10}=0.2961, \quad {}_{12}C_2(1/4)^2(3/4)^{10} \\ &= 0.2323 \end{aligned} \quad (14)$$

が(11)等に対応するが、「4面体」と「予想する」ことの不自然さが(14)からうかがわれる。ちなみに、商品の品質管理にも上の論法が用いられている。

(11)に代表される尺度は「予測値毎の現象発生率を調べる」方法の組合せである。順序立てて説明するなら、一連の予報から予報値(p_1, p_2, \dots, p_k)の同じものを一定数(n)ずつ取り出し、おのおのに対する実況(現象あり： $s(l)$ 回、現象なし： $(n-s(l))$ 回)とつき合わせて、

$$I = \left\{ \prod_{l=1}^k {}_n C_{s(l)} p_l^{s(l)} (1-p_l)^{n-s(l)} \right\}^{1/kn} \quad (15)$$

を計算し、この I によって予報の妥当性を比較するのである。(11)~(13)を I に直してもそれらの大きさの順序は変わらない。また、先の例に「20回とも8/20の予報」をする例を追加するとき、(11)に対応して

$${}_{20}C_8(8/20)^8(12/20)^{12}=0.1797 \quad (16)$$

が計算され「不都合」に見えるが、(11)と(16)を I になおすならば、

$$\begin{aligned} &[\{ {}_{20}C_4 0.2^4(1-0.2)^{16} \} \{ {}_{20}C_{12} 0.6^{12}(1-0.6)^8 \}]^{1/40} \\ &= 0.9222 \end{aligned} \quad (11)'$$

$$\{ {}_{20}C_8(8/20)^8(12/20)^{12} \}^{1/20}=0.9178 \quad (16)'$$

となり、「気候値を予報しておけば良い」との批判に応えることができる。その理論的根拠は

$$\begin{aligned} &{}_n C_r \left(\frac{r}{n} \right)^r \left(\frac{n-r}{n} \right)^{n-r} > {}_n C_{r'} \left(\frac{r'}{n} \right)^{r'} \left(\frac{n-r'}{n} \right)^{n-r'}, \\ &\left(r < r' \leq \frac{n}{2} \right) \end{aligned} \quad (17)$$

が成立することである。Stirlingの公式により、

$$\begin{aligned} &{}_n C_r \left(\frac{r}{n} \right)^r \left(\frac{n-r}{n} \right)^{n-r} \sim \sqrt{\frac{n}{2\pi r(n-r)}} \rightarrow 0 \\ &(n, r \rightarrow \infty) \end{aligned} \quad (18)$$

$$\begin{aligned} &\left\{ {}_n C_r \left(\frac{r}{n} \right)^r \left(\frac{n-r}{n} \right)^{n-r} \right\}^{1/n} \sim \left\{ \frac{n}{2\pi r(n-r)} \right\}^{1/2n} \rightarrow 1 \\ &(n, r \rightarrow \infty) \end{aligned} \quad (18)'$$

であり、

$$\begin{aligned} &({}_{10}C_2 0.2^2 \times 0.8^8)^{1/10}, \quad ({}_{10}C_6 0.6^6 \times 0.4^4)^{1/10}, \\ &({}_{20}C_8 0.4^8 \times 0.6^{12})^{1/20} \end{aligned}$$

などはすべて1に近づくのであるが、(17)はその収束の速さに差があることを示している。そして、 I は「妥当」に順位づけをするのである。ただし、「0%」と「100%」は I ではカバーできない。これは確率の性質上やむを得ない。

さらに大切なことは、 I はあくまでも「モデル」の妥当性の尺度であり、個々の「予報と実況の差」なるものは全く考慮しないことである(第2章参照)。

6. あとがき

本稿の主旨は確率についての基本的な理解が曖昧なままに確率予報が行われていることに対する反省である。確率予報は気象学の成果ではなく、計算機の発達の結果である。元来無意味な「個々の予報の誤差」にこだわるならば、計算確率の「修正」によって、特定の利用者(個々の予報ではなく「予報全体」を利用する者)には有益なはずの確率表示の利点を殺してしまうであろう。確率予報は従来予報の「付録」として、計算確率どおりに発表するのが最も自然で、かつ、有効なのである。また、利用者が確率予報の意味を正しく理解することも不可欠である。こんなことを敢えて書くのは、筆者自身にしても「80%」と聞くと「100%」のように感じ、「残余の20%」をほとんど意識しない「不合理な感性」を捨て切れなからである。そして、「50%」には困惑するのである。

おわりに、本稿をまとめるに当たり貴重なご意見をいただいた丸山健人、増田善信、生田保夫の各氏、また、暖かい援助をいただいた職場の方々に感謝の意を表した

い.

文献

- (*) Berkofsky, L., 1950: An objective determination of probability of fog formation, *Bull. Amer. Met. Soc.*, 31, 5-10.
- (*) Brelford, W.M. and R.H. Jones, 1967: Estimating probabilities, *Mon. Wea. Rev.*, 95, 570-576.
- Brier, G.W., 1950: Verification of forecasts expressed in terms of probability, *Mon. Wea. Rev.*, 78, 1-3.
- (*) Jones, R.H., 1968: A nonlinear model for estimating probabilities of K events, *Mon. Wea. Rev.*, 96, 383-384.
- (*) Lund, I.A., 1955: Estimating the probability of a future event from dichotomously classified predictors, *Bull. Amer. Met. Soc.*, 36, 325-328.
- (*) 村田 全, 1966: 数学の思想, NHK ブックス, 42, NHK.
- (*) 丸山 儀四郎, 1956: 確率および統計, 共立出版, 1-68, 133-155.
- Murphy, A.H., 1972: Ordinal relationship between measures of the "accuracy" and "value" of probability forecasts; preliminary results, *Tellus*, XXIV, 531-542.
- 小河原正巳, 1959: 予報の情報量と O.R., 気象と統計, 10, 2, 1-3.
- (*) 小沢 正, 1977: MOS について, 大気, 17, 18, 3-54.
- (*) Russel, B., 1920: 数理哲学序説 (平野智治訳, 1954, 岩波書店).
- 鈴木栄一, 1981: 天気予報における確率・統計的方法の WMO シンポジウム, 天気, 28, 121-128.
- (*) 高木貞治, 1949: 数の概念, 岩波書店.
- (*) 遠山 啓・長妻克亘, 1962: 量の理論, 明治図書.
- 上の文献中 (*) 印のものは本稿の参考にしたが、本文には引用されていない。