天気予報の価値をどう測るか

山田　真吾

本稿では，天気予報を「経済的な価値」という観点から評価する試みを紹介する．問題を単純化するために，ある現象が発生するか，発生しないかによって異なる費用が発生するが，ある対策を取れば費用を削減することができるという簡単な「コスト・ロス・モデル」を導入する．これを用いて，天気予報を用いた場合と用いない場合の費用の期待値を比較する．また，完全予報を用いた場合の費用を1，予報を用いない場合の費用を0に規格化することにより，「利益スキルスコア」を導出する．最後に，簡単な実例を示し，天気予報を有効に利用するためには，その平均的な精度を知ることが不可欠であり，そのような情報が公開されることが重要であることを述べる．

キーワード：天気予報の価値，コスト・ロス・モデル，利益スキルスコア

1.　はじめに

ひと昔前に比べると，天気予報（最近では気象情報とも言われることも多いが，本稿では政策判断・意思決定に用いられる気象に関する情報全般を「天気予報」と総称することにする）はよく当たるようになったと言われる．コンピュータ技術や気象衛星をはじめとするリモートセンシング技術の進展に伴い，「数値天気予報モデル（大気中で起きる様々な物理現象を微分方程式で表し，それを積分して大気の将来の状態を予測するもの）」による予報の精度は日進月歩で向上し，天気予報の精度向上に貢献してきた．

気象庁が一般に発表している天気予報については，気象庁ホームページ（http://www.jma.go.jp）にいくつかの客観的な精度評価指標が公表されている（例えば，前日の夕方に発表した「翌日の降水の有無」予報の適中率や「翌日の朝の最低気温」予報の根2乗平均誤差）．これらの指標は，確かに客観的な予報精度を示したものではあるが，「天気予報の社会的な価値」という観点から見たものではない．Murphy（1993）は，天気予報の評価に関して，「一貫性」，「品質」，「価値」という三つの観点があると論じた．このうち，「一貫性」は予報者が正しいと信じる予報と実際に出された予報との一貫性を測るものであり，政策的あるいは作為的な偏りがないほうがよいとするものである．また，「品質」は予報とそれに対応する観測（実況）の一致を測るものであり，前述した適中率や根2乗平均誤差はこれを測ったものと言える．一方で，「価値」は天気予報を何らかの意思決定・政策判断に用いた場合に，それを用いなかった場合に比べてどれだけの経済的な利益（損失の回避あるいは費用の減少）が得られるかを測るものである．天気予報は，何らかの判断（例えば，今日傘を持っていくかどうか）に用いられなければ意味がないが，その判断に伴う経済的な利益や損失（例えば，常に傘を持ち歩くことの経費や傘を持たないで雨に降られた場合の被害額）は，個人によっても，状況によっても違うから，客観的に求めるのが難しい．このため，これまではあまり「価値」による評価は行われてこなかった．しかし，天気予報を民間気象事業者から対価を払って買うことができるようになり，また国民の税金を気象業務に支出することに対して相応の「経済価値」を求めるようになった今日，上に述べた「価値」の基準による天気予報の評価は，避けて通れない問題となっている．立平（1999）は，コスト・ロス・モデルを用いて，気象情報を意思決定に使う方法について解説を行っている．本稿は，立平（1999）にならって，「価値」の基準による天気予報の評価方法を定式化し，いくつかの実例を示すとともに，最大限の「価値」を引き出すための注意点を解説したものである．

やまだ　しんご
気象庁 予報部数値予報課
〒100-8122 千代田区大手町1-3-4

2.　コスト・ロス・モデルを用いた定式化

先に述べたように，天気予報を用いた意思決定過程は，現実には非常に複雑な分岐を持っている．しかし，ここでは最も単純なケースのみを考える．それは，ある気象状態（以後単に「現象」と呼ぶ）が起こるか否かによって，ある行為（以後単に「対策」と呼ぶ）を

表1　発生する費用の分割表

費用	対策有	対策無
現象有	C	L
現象無	C	0

取るかどうかを判断する，という二者択一のケースである．表1は，現象の有無および対策の有無によって2×2分割表を作成し，それぞれの場合に発生する費用を示したものである．表1は，対策を取らなかったにもかかわらず現象が起こった場合にはLという被害（費用）が生じること，逆に，対策を取っていて現象が起こらなかった場合には，対策に要するCという費用が生じることを示している．対策を取らず現象も起こらなかった場合には，被害も対策費も0なので費用は0，対策を取っていて現象が起こった場合には，被害は0だが対策費Cが発生するので費用はCと考える．現実には，対策を取っても防げない被害も存在するが，それは天気予報の利用の有無にかかわらず生じる被害であるから，天気予報を利用した場合の費用と利用しない場合の費用の差だけを問題にする場合には，Lからあらかじめ差し引いておいても違いは生じない．言い換えると，費用Cをかけて対策を取ることによって軽減可能な被害額をLと考えればよい．当然ながら，CよりもLのほうが大きくなければ対策を取る意味がないし，Cが大きいほどLも大きくなると考えられる．以後，Cをコスト，Lをロスと呼ぶことにする．

天気予報の誤りには，(1)現象ありと予報して実際には現象が起こらなかった場合（いわゆる「空振り」予報）と(2)現象なしと予報したのに実際には現象が起こった場合（いわゆる「見逃し」予報）の2種類がある．前節で述べた適中率を評価指標に用いる場合には，これら2種類の誤りに対しては同じ大きさのペナルティが課せられる．しかし，「価値」による評価においては，コストCとロスLの比の大きさによって，2種類の誤りに対して異なる大きさのペナルティが課せられることを次に示す．

「価値」による評価においては，天気予報を用いた場合と用いない場合の費用の差を評価の対象とするため，まず，天気予報を用いない場合の費用の最小値を求める．天気予報を用いない場合の意思決定方法としてはいくつかの方法が考えられるが，一番単純な方法として，次の三つのケースを考える．

(1) 常に，対策を取る．

(2) 常に，対策を取らない．

(3) 現象の発生確率と同じ確率でランダムに対策を取る．

現象が確率$p\,(0<p<1)$で発生するとき，上記の場合に，生じる費用の期待値を求める．(1)の場合には，毎回対策費Cがかかるが，被害は0で済むので，費用H_1は，

$$H_1=C$$

(2)の場合には，コストは0であるが，確率pで被害Lが発生するから，費用H_2は，

$$H_2=pL$$

(3)の場合には，対策費Cの発生する確率はp，被害Lの発生する確率は$p(1-p)$であるから，費用H_3は，

$$H_3=pC+p(1-p)L$$

$H_1<H_2$となる条件は，$C<pL$である．つまり$C/L<p$の場合には，(1)の意思決定方法のほうが(2)に比べて費用の期待値が小さくなると言える．$H_1<H_3$となる条件を求めると，$C<pC+p(1-p)L$から，$C<pL$となる．この条件は，$H_1<H_2$の条件とも同じであるから，$C/L<p$ならば，上記3種類の意思決定方法の中では，(1)が最も費用を小さくする方法であることがわかる．言い換えると，C/L（以後，コスト・ロス比と呼び，γと表記することにする）が現象の発生確率pよりも小さければ，常に対策を取ることが，常に対策を取らないことやランダムに対策を取るよりもよい判断である．

次に，$H_2<H_3$となる条件を求めると，$pL<pC+p(1-p)L$から，$C>pL$となる．この条件は，$H_2>H_1$の条件とも同じであるから，$\gamma=C/L>p$ならば，上記3種類の意思決定方法の中では，(2)が最も費用を小さくする方法であることがわかる．

以上をまとめると，予測情報が得られない場合，コスト・ロス比γが現象の発生確率pよりも小さければ常に対策を取り，逆であれば常に対策を取らないと判断するのが，最善の意思決定法であると言える．言い換えると，現象の発生頻度がある程度大きい場合，対策を取るための費用が非常に少なくてすむ場合，あるいは，現象が起こった場合の被害額（正確には対策費Cがかかる対策を取ることにより軽減可能な被害額L）が非常に大きい場合には，常に対策を取るのがよく，それらの逆の場合には，常に対策を取らないのがよい．このような判断は，「無技術最良予測」と呼ばれる（菊池原，1988）．このような単純なモデルにおいては，現象の発生確率pとγ（絶対値ではな

表2　現象発生/非発生予測の検証分割表

割合	予測有	予測無	合計
現象有	W	$P-W$	P
現象無	$Q-W$	$N-P-Q+W$	$N-P$
合計	Q	$N-Q$	N

表3　表2を相対頻度を用いて書き直したもの

割合	予測有	予測無	合計
現象有	w	$p-w$	p
現象無	$q-w$	$1-p-q+w$	$1-p$
合計	q	$1-q$	1

い）だけが，意思決定においては必要である[注：天気現象の場合，現象の発生相対頻度pは理論的には決められず，経験的に決めるしかない．未来の特定期間内の現象の発生頻度は，一つの予測情報と考えられる．予測情報を利用しない場合には，pとして過去のある期間における現象の発生相対頻度ρを利用することになる．この場合の「無技術最良予測」は「気候値最良予測」と呼ぶことができる．ρと特定期間の現象発生相対頻度pが異なる場合を考慮する必要があるが，ここでは十分長い評価期間を取ることを想定し，pとρの差は非常に小さいとする]．

さて，次に，現象の発生・非発生に関する予測情報が存在する場合について考える．その検証結果が表2のような分割表の形で与えられたとする．表2の各項目のうち独立なものは4個であるが，各項を予測の総回数Nで割り相対頻度にすることにより，3個に減る．ここで，現象発生の相対頻度$p=P/N$，現象あり予報の相対頻度$q=Q/N$，現象あり予報の適中した相対頻度$w=W/N$を独立な変数とすると，表3のようになる．ただし，$0\leqq w\leqq p\leqq 1$，$0\leqq w\leqq q\leqq 1$である．先に述べた2種類の誤りのうち，「空振り」の相対頻度は$q-w$，「見逃し」の相対頻度は$p-w$と表されることに注意しておく．

現象ありと予測された場合には対策を取り，現象なしと予測された場合には対策を取らないという意思決定を行った場合，費用の期待値H_fは，

$$H_f=qC+(p-w)L$$

である．特に，予測が完全な場合，すなわち，「空振り」も「見逃し」も0の場合には，$p=q=w$だから，費用の期待値H_pは，

$$H_p=pC$$

となる．

表3で表される精度を持つ予報を利用した場合の費用の期待値H_fが「無技術最良予測」による費用の期待値より小さくなるのは，どんなときであろうか．

$\gamma=C/L>p$のときには，「無技術最良予測」による費用の期待値はH_2だから，$H_2-H_f>0$より，

$$H_2-H_f=pL-qC-(p-w)L=wL-qC>0$$

したがって，$\gamma<w/q$となる．ここで，w/qは，現象ありという予測が発表されたケースのうち，実際に現象が起こった割合（この割合をDoswell et al. (1990)に従ってFOH (Frequency Of Hits) と呼ぶことにする）である．γがpよりも大きい利用者にとっては，FOHがγよりも大きいことが，無技術最良予測より費用が小さくなる条件であることがわかる．

一方，$\gamma=C/L<p$のときには，「無技術最良予測」による費用の期待値はH_1だから，$H_1-H_f>0$より，

$$H_1-H_f=C-qC-(p-w)L$$
$$=(1-q)C-(p-w)L>0$$

したがって，$\gamma>(p-w)/(1-q)$となる．ここで，$(p-w)/(1-q)$は，現象なしという予測が発表されたケースのうち，実際には現象が起こった割合（同じくDoswell et al. (1990) に従ってDFR (Detection failure Ratio) と呼ぶことにする）を表している．γがpよりも小さな利用者にとっては，DFRがγよりも小さいことが，無技術最良予測より費用が小さくなる条件であると言える．

以上をまとめると，コスト・ロス比γが現象の発生確率pよりも大きな利用者にとっては，「現象の発生」が予測された場合に限った適中・不適中の割合が関心事であり，適中の割合FOHがγより大きくなければ，予報を利用する価値はない．逆に，γがpよりも小さな利用者にとっては，「現象の非発生」が予測された場合に限った適中・不適中の割合が関心事であり，不適中の割合DFRがγより小さくなければ，予報を利用する価値はないと言える．注意すべきは，予報が完全（「見逃し」も「空振り」も0）でない限り，予報を利用することによって，費用を節減できる利用者は限られているということである．つまり，γがFOHよりも大きい利用者やDFRよりも小さい利用者は，予報を利用すべきではなく，前者は常に対策を取り，後者は常に対策を取らないほうが費用が少ない．予報を利用して効果があるのは，DFR$<\gamma<$FOHの利用者であり，その範囲の広さ，すなわち，

$$\alpha_{\max}=\frac{w}{q}-\frac{p-w}{1-q} \tag{A}$$

は，予報の価値を表す一つの指標となると考えられる．(コスト・ロス比に対する利用者の分布が一様であると仮定すれば，$a_{\max}$ は予報によって利益を得られる利用者の数を表す)．このスコアは，Claytonのスキルスコアと呼ばれている（Clayton, 1934)．

これまで述べた定式化によれば，たとえ完璧な予報を得たとしても，$H_p = pC$ の費用が発生することになる．このことを考慮するため，完全予報を用いた場合の利益が1，「無技術予報」の利益が0となるように規格化を行い，「利益スキルスコア（VSS: Value Skill Score)」と呼ぶことにする．

$\gamma < p$ のときには，

$$\begin{aligned} VSS &= (H_1 - H_f)/(H_1 - H_p) \\ &= ((1-q)C - (p-w)L)/(1-p)C \\ &= ((1-q)\gamma - (p-w))/(1-p)\gamma \end{aligned} \tag{B}$$

$\gamma > p$ のときには，

$$\begin{aligned} VSS &= (H_2 - H_f) > (H_2 - H_p) \\ &= (wL - qC)/p(L-C) \\ &= (w - q\gamma)/p(1-\gamma) \end{aligned} \tag{C}$$

VSS を γ で微分すると，

$$\begin{aligned} d(VSS)/d\gamma &= (p-w)/(1-p)\gamma^2 > 0 \quad \text{if} \quad \gamma < p \\ &= -(q-w)/p(1-\gamma)^2 < 0 \quad \text{if} \quad \gamma > p \end{aligned}$$

したがって，$\gamma < p$ では γ について単調増加，$p < \gamma$ では γ について単調減少であるから，$\gamma = p$ のときに VSS が最大となり，最大値は，

$$VSS_{\max} = \frac{w - pq}{p(1-p)} \tag{D}$$

である．正の VSS が得られる利用者が存在する条件は，$VSS_{\max} > 0$ から，$w - pq > 0$ となり，これは，現象ありの適中数がランダム予報による適中数よりも多いことを表している．$VSS_{\max}$ が0であれば，ランダム予報と同じであり，1に近いほど完全予報に近いので，予報精度の一つの指標として用いることができる．

(D)式を書き換えると，

$$VSS_{\max} = \frac{w}{p} - \frac{q - w}{1 - p} \tag{D$'$}$$

となる．(D)$'$ の第1項は，実際に現象が起こった事例のうち正しく予測されていた割合（同じくDoswell et al. (1990) に従いPOD (Probability Of Detection) と呼ぶ）であり，第2項は，実際には現象が起こらなかった事例のうち現象ありの予測が行われていた割合（同じくDoswell et al. (1990) に従いPOFD (Porbability Of False Detection) と呼ぶ）を表している．PODとPOFDの差による評価指標は，Peirceのスキルスコアと呼ばれている（Peirce, 1884）［このスコアは，TSS（True Skill Statistics）やHansen-KuiperのVスコアと呼ばれることもある］．

先に述べたClaytonのスキルスコア（(A)式）とPeirceのスキルスコア（(D)$'$式）を比較すると，各項の分母に含まれる q が p に，第2項の分子に含まれる p が q に，それぞれ置き換わっていることがわかる．したがって，$p = q$ のとき，すなわち現象ありを予報した回数と現象の発生回数が一致していれば，ClaytonのスキルスコアとPeirceのスキルスコアは一致する．しかし，$p \neq q$ のときには，二つのスキルスコアは一致しない．

さらに，(D)$'$ 式を次のように変形する．

$$VSS_{\max} = 1 - \frac{p - w}{p} - \frac{q - w}{1 - p} \tag{D$''$}$$

見逃し $p - w$ または空振り $q - w$ が同じ回数増加した場合のスコアの減少幅を見ると，見逃しに対しては $1/p$ に比例し，空振りに対しては $1/(1-p)$ に比例する．$p \neq 1/2$ の場合には，1回の見逃しと空振りに対するペナルティの大きさが異なることがわかる．

3. 利用者集団全体の「利益」の期待値に基づく評価指標の算出

前節の議論から，Claytonのスキルスコアは正の「利益スキルスコア（VSS)」を得られるコスト・ロス比の範囲の広さを，Peirceのスキルスコアは期待される VSS の最大値を表す指標であることがわかった．

個々の利用者（コスト・ロス比 γ が確定）に対しては，(B)式または(C)式を用いて，VSS を算出できる．ところで，γ が異なる複数の利用者（利用者集団と呼ぶ）に対して，予報を発表している気象庁のような場合，予報の価値をどのように測ればよいだろうか．

前節の結果をおさらいすると，予報を利用した場合に得られる利益 B は，次式で表せる．

$$\begin{aligned} B &= H_1 - H_f = (w - q\gamma)L \quad \text{if} \quad p < \gamma \\ &= H_2 - H_f = ((1-q)\gamma - (p-w))L \quad \text{if} \quad p > \gamma \end{aligned}$$

コスト・ロス比 γ に対する利用者の数密度 $n(\gamma)$ がわかれば，B に数密度をかけて積分することにより，利用者集団全体の「利益」の平均値 $\langle B \rangle$ が求まる．しかし，現実の多数の利用者に対して，それぞれの γ を知ることは困難である．そこで，簡単な数密度分布を仮定して，$\langle B \rangle$ を求めてみることにする．ただし，予報精度に関する正しい情報が利用者に伝えられており，利益が負になる利用者は，予報を利用しないと考

える．したがって，利益が正となる $w/q<\gamma<(p-w)/(1-q)$ の範囲で利益スキルスコアを積分する．

まず，$n(\gamma)L\equiv 1$ の場合を考える．このとき，$n(\gamma)=\gamma/C$ であるから，利用者のコストがほぼ一定の場合には，数密度分布は γ に比例，すなわち L に反比例して減るとことになる．積分を実行すると，

$$\begin{aligned}\langle B\rangle &= \int_{(p-w)/(1-q)}^{p}[(1-q)\gamma-(p-w)]d\gamma \\ &\quad +\int_{p}^{w/q}[w-q\gamma]d\gamma \\ &= \frac{1-q}{2}\left[p^2-\left(\frac{p-w}{1-q}\right)^2\right]-(p-w)\left(p-\frac{p-w}{1-q}\right) \\ &\quad +w\left(\frac{w}{q}-p\right)-\frac{q}{2}\left[\left(\frac{w}{q}\right)^2-p^2\right] \\ &= \frac{1}{2}\left[\frac{(w-pq)^2}{1-q}+\frac{(w-pq)^2}{q}\right]=\frac{(w-pq)^2}{2q(1-q)} \qquad \text{(E)}\end{aligned}$$

完全予報の利益は，(E)式で $w=q=p$ とおいて，

$$\langle B\rangle_p=\frac{p(1-p)}{2} \qquad \text{(F)}$$

$\langle B\rangle$ を $\langle B\rangle_p$ で規格化すると，集団全体としての「利益スキルスコア」$\langle BSS\rangle$ が得られ，次のように表せる．

$$\langle BSS\rangle=\frac{(w-pq)^2}{pq(1-p)(1-q)} \qquad \text{(G)}$$

この場合，$\langle BSS\rangle$ は，Clayton のスキルスコアと Peirce のスキルスコアの積となっている．

次に，$n(\gamma)L\equiv 1/\gamma$ の場合を考える．このとき，$n(\gamma)=1/C$ となり，利用者の数密度が C に反比例することになる．積分を実行すると，

$$\begin{aligned}\langle B\rangle &= \int_{p-w/1-q}^{p}\left[(1-q)-\frac{(p-w)}{\gamma}\right]d\gamma \\ &\quad +\int_{p}^{w/q}\left[\frac{w}{\gamma}-q\right]d\gamma \\ &= (p-w)\ln\left(\frac{p-w}{1-q}\right)+w\ln\left(\frac{w}{q}\right)-p\ln p \qquad \text{(H)}\end{aligned}$$

$w=pq$ のときに，$\langle B\rangle=0$ となることが確かめられる．また，完全予報の場合には $w=q=p$ より

$$\langle B\rangle_p=-p\ln p \qquad \text{(I)}$$

となる．ここで，$\lim_{x\to 0}x\ln x=0$ を用いた．

$w-pq=\alpha$ とおくと，

$$\begin{aligned}\langle B\rangle &= (p-w)\ln p\left[1-\frac{\alpha}{p(1-q)}\right] \\ &\quad +w\ln p\left[1+\frac{\alpha}{pq}\right]-p\ln p \\ &= (p-w)\ln\left[1-\frac{\alpha}{p(1-q)}\right]+w\ln\left[1+\frac{\alpha}{pq}\right]\end{aligned}$$

$\alpha\ll pq$ かつ $\alpha\ll p(1-q)$ の場合には，

$$\begin{aligned}\langle B\rangle &\approx -\frac{(p-w)\alpha}{p(1-q)}+\frac{w\alpha}{pq} \\ &= -\frac{(p-pq-\alpha)\alpha}{p(1-q)}+\frac{(pq+\alpha)\alpha}{pq}=\frac{\alpha^2}{pq(1-q)}\end{aligned}$$

$w-pq$ が小さいうちは，その2乗に比例して利益が増加することがわかる．

この場合も，$\langle B\rangle/\langle B\rangle_p$ を考えることにより，スキルスコア $\langle BSS\rangle$ を求めることができる[注：現実には，予報の成績（w と q）は事前にはわからないので，上の仮定は成り立たない．予報を利用するか否かのしきい値を変えると，BSS の積分値は必ず小さくなるから，ここで算出した $\langle BSS\rangle$ は，その上限値であると言える]．

4. 確率形式による予報の評価

前節までの議論では，「現象あり」または「現象なし」という予報（「断定予報」と呼ばれる）が行われ，それとは別に予報精度に関する情報が（分割表あるいはFOH，DFR，POD，POFDという形で）与えられると考えた．

これに対して，現象発生確率による予報が行われることがある．例えば，6時間に1 mm以上の降水があるかどうかを予報する「降水確率予報」は，その一例である．この場合の確率は，例えば降水確率30%の予報が多数回出された場合，そのうち約3割の事例で雨が降るであろうということを意味している．

断定予報の場合，予報精度に関する情報は別に取得する必要があるが，確率形式の予報は，予報自体に精度に関する情報を含んでいると考えてよい．ある確率値を予報した多数の予報例を集めて，そのうち実際に現象が起こった事例の割合を求めた場合，それが予報された確率値と等しいならば，「信頼度が完全である」と呼ばれる．

この場合，前節で見たように，予報の利用者は，自らのコスト・ロス比 γ と予報された確率値 ρ を比較することによって，予報を利用するかどうかを判断することができる．例えば，γ が現象の発生確率 p よりも小さい利用者は，ρ が γ よりも小さいときに対策を取らないという判断をすれば，常に対策を取るという「無技術最良予測」よりも費用が少なくなる．逆に，γ が現象の発生確率 p よりも大きい利用者は，ρ が γ よりも大きいときに対策を取るという判断をすれば，常に対策を取らないという「無技術最良予測」よりも費用が少なくなる．

もし，信頼度が完全であり，$p<\gamma<\rho$ または $p>\gamma>\rho$ である確率値の予報が行われる回数が0でなければ，予報を利用することによって費用を減少させることができる[注：仮に，信頼度が完全でなくても，ある確率値が予報された場合に，実際に現象が起こる条件付確率がわかっていれば，確率値を条件付確率値に読み替えることにより，損失を防ぐことができる].

コスト・ロス比が γ である利用者が確率予報を利用した場合の費用を算出してみる．費用の期待値 H_f は，

$$H_f = pL + \int_\gamma^1 (Cq(\rho) - Lw(\rho))d\rho$$
$$= C + \int_0^\gamma (Lw(\rho) - Cq(\rho))d\rho$$

ここで，$q(\rho)$ は確率値 ρ の予報が発表された相対頻度，$w(\rho)$ は確率値 ρ の予報が発表され，かつ現象が発生した相対頻度を表している．また，

$$\int_0^1 w(\rho)d\rho = p, \int_0^1 q(\rho)d\rho = 1$$

の関係を用いた．

信頼度が完全であれば，$w(\rho) = \rho q(\rho)$ であるから，

$$H_f = pL + L\int_\gamma^1 (\gamma - \rho)q(\rho)d\rho$$
$$= C - L\int_0^\gamma (\gamma - \rho)q(\rho)d\rho$$

となる．

$\gamma<p$ の場合には，$H_1 = C$ との差

$$H_1 - H_f = L\int_0^\gamma (\gamma - \rho)q(\rho)d\rho \tag{J}$$

が予報を利用したことによる費用の減少額となる．また，$\gamma>p$ の場合には，$H_2 = pL$ との差

$$H_2 - H_f = L\int_\gamma^1 (\rho - \gamma)q(\rho)d\rho \tag{K}$$

が予報を利用したことによる費用の減少額となる．

(H), (I)式ともに γ から離れた予報の回数が多いほど費用の減少額が大きくなることを意味している．

利用者全体の利益を計算する場合には，(J), (K)式をそれぞれ $0<\gamma<p$ あるいは $p<\gamma<1$ の範囲で積分して加えればよい．

以上のことから，確率表現による予報では確率の信頼度が重要であり，また，大きな利益を生み出すためにはできるだけコスト・ロス比 γ から離れた確率値の予報頻度が多い（このことは確率予報の分解度が高いとも表現される）ことが必要であることがわかる．

実際には，確率予報は任意の実数値ではなく，10%刻みといった確率階級値 $p_i (i=0, \cdots, N)$ として発表されることが多い．この場合には，積分は各階級の値の和として表せる．すなわち，

$\gamma<p$ の場合には，

$$H_1 - H_f = L\sum_{i=0}^{J} (\gamma - \rho_i) \cdot q_i \tag{J$'$}$$

ただし，q_i は i 番目の階級が予報された事例数の相対頻度を，ρ_i は i 番目の階級が予報された事例のうち，実際に現象が発生した事例数の割合を示し，J は確率階級値が γ を超えない最大の階級番号を表すものとする．また，$\gamma>p$ の場合には，

$$H_2 - H_f = L\sum_{i=K}^{N} (\rho_i - \gamma) \cdot q_i \tag{K$'$}$$

ただし，K は確率階級値が γ を超える最小の階級番号を表すものとする．

$n(\gamma)L \equiv 1$ の場合に，利用者集団全体に対する費用の減少額を計算すると，次のようになる．

$$\langle B \rangle = \sum_{i=0}^{N} \left[\left(\frac{p+p_i}{2} - \rho_i \right)(p - p_i)q_i \right] \tag{L}$$

ただし，i 番目の階級の確率値を p_i と表した．

完全予報の利益は，断定予報における完全予報の利益と同じであるから，利益スキルスコア $\langle BSS \rangle = \langle B \rangle / \langle B \rangle_p$ は

$$\langle BSS \rangle = \frac{2}{p(1-p)} \sum_{i=0}^{N} \left[\left(\frac{p+p_i}{2} - \rho_i \right)(p - p_i)q_i \right] \tag{M}$$

となる．

5. 利益スキルスコア算出の実例

表4は，気象庁ホームページの「予報検証」のページに掲載されている年集計のスコアから，明日の降水の有無予報の2×2分割表を作成したものである[注：気象庁ホームページには分割表そのものは掲載されていないので，掲載されている検証スコアに近いものが得られるように，各コマに入る事例数の比率を作成した．したがって，近似的な値であることにご注意いただきたい]．表には，節2で定義した各種スコア(FOH $=w/q$, DFR $=(p-w)/(1-q)$, POD $=w/p$, POFD $=(q-w)/(1-p)$)の値も掲載してある．

表4の(a), (b)は，いずれも全国の予報区について明日の降水の有無予報についてまとめた分割表であるが，(a)は朝5時に発表したものであり，(b)は夕方5時に発表したものである．後者の方がリードタイムが短いので予報精度が高いと考えられるが，確かに全てのスコアについて，(b)の精度が(a)のそれを上回っている．

(G)式で定義した利益スキルスコア $\langle BSS \rangle$ を求めてみると，(a)は，0.293, (b)は，0.341となる．また，

表4　明日の降水の有無予報の分割表

(a)　2002年全国平均（朝5時発表）

割合	予測有	予測無	合計
降水有	0.213	0.128	0.341
降水無	0.072	0.587	0.659
合計	0.285	0.715	1.000

FOH=0.75　DFR=0.18　POD=0.62　POFD=0.11

(b)　2002年全国平均（夕方5時発表）

割合	予測有	予測無	合計
降水有	0.224	0.117	0.341
降水無	0.065	0.594	0.659
合計	0.289	0.711	1.000

FOH=0.78　DFR=0.16　POD=0.66　POFD=0.10

(c)　2002年東海地方（夕方5時発表）

割合	予測有	予測無	合計
降水有	0.22	0.11	0.33
降水無	0.05	0.62	0.67
合計	0.27	0.73	1.00

FOH=0.81　DFR=0.15　POD=0.67　POFD=0.07

(d)　2002年九州南部地方（夕方5時発表）

割合	予測有	予測無	合計
降水有	0.28	0.10	0.38
降水無	0.06	0.56	0.62
合計	0.34	0.66	1.00

FOH=0.82　DFR=0.15　POD=0.74　POFD=0.10

(H)式と(I)式で求めた利益から求めたスキルスコアは，(a)で0.231，(b)で0.269となった．後者については，完全予報による利益 $-p\ln p$ の最大値は $p=e^{-1}\fallingdotseq 0.368$ のときに得られるが，降水の発生確率0.341がこれに近いため，相対的なスキルスコアが低くなったものと考えられる．

表4の(c)，(d)は，それぞれ東海地方と九州南部地方についての分割表であるが，全体的な適中数は等しく，見逃し数は(d)が空振り数は(c)が少なくなっている．現象の発生相対頻度も0.33と0.38で5%も異なっている．このような場合に，どちらの予報の方が精度が高いと考えるかは，見方・考え方によって変わってくる．ちなみに，降水予報の検証においてしばしば用いられる不偏スレットスコア $(=(w-pq)/(p+q-w-pq))$ を求めると，(c)は0.450，(d)は0.481である．

これらについて，利益スキルスコア⟨BSS⟩を求めると，(c)は0.393，(d)は0.430となる（後者の計算方法では，0.308と0.339）．いずれの地方も全国平均よりも精度が高く，特に，警報誤り率（POFD）をあまり増やすことなく，警報適中率（POD）を上げている九州南部地方の方が東海地方よりも予報の価値が大きいといえそうである．

表5　1ヶ月平均気温の「高い」（あるいは「低い」）確率の予報についての検証表

確率階級値	発表相対頻度	実現相対頻度
0	0	0
0.1	0.06	0.07
0.2	0.22	0.16
0.3	0.35	0.35
0.4	0.10	0.34
0.5	0.25	0.48
0.6	0.02	0.86
0.7	0	0
0.8	0	0

降水確率予報についても，実例を示したいところだが，残念ながら確率階級ごとの発表回数や適中率は公表されていない．そこで，確率表現による予報が行われている1ヶ月平均気温の予報の検証結果を用いて，スコアを計算してみる．表5は，1ヶ月平均気温が平年に比べて「高い」（あるいは「低い」）となる確率の予報に対して，実際に「高い」（あるいは「低い」）となった割合を示したものである．統計期間は1996年3月から1998年2月までの3年間で，値は気象庁ホームページの「季節予報の上手な利用」—「季節予報とは？」に掲載されているグラフから読み取った（0.9以上の確率値の発表回数は0なので省略した）．

「高い」（あるいは「低い」）の気候学的出現確率が33%であることを利用すると，節4の(F)式を用いて，利益スキルスコアを計算することができる．その結果は，0.092であった．降水確率予報と異なり10%以下や60%以上の確率値が計算される頻度が非常に少ないことや30%のように予報では気候学的確率より小さいにもかかわらず，実際の出現率が気候学的確率よりも大きくなっている階級があることが，利益を少なくしている原因となっていると考えられる．しかし，このような精度を勘案して，意思決定に正しく利用すれば，もっと大きな利益が得られる可能性がある．例えば，階級確率値を検証で得られた事後確率値と等しいと考えて，利益スキルスコアを計算すると，0.099となる．

6.　まとめと議論

簡単なコスト・ロス・モデルに基づいて，あるコスト・ロス比を持つ予報利用者に対して，「無技術最良

予測」による費用と予測を利用した場合の費用の差（利益）を求め，完全予測を利用した場合の利益で正規化したスコアを定式化した．さらに，その予報が不特定多数の利用者集団に対して発表されることを想定して，コスト・ロスに対する利用者数について2通りの分布を仮定して，集団全体に対する利益の大きさを求め，利益スキルスコアを導いた．

このスコアを求めるためには，現象の発生確率（事前確率）と発生予報の有無（あるいは発生確率予報値）に対して現象発生の有無を調べた分割表が必要である．ある特定の期間を対象とすると，現象の発生頻度は事前確率とは異なる可能性が高いが，ここでは十分長い検証期間を取ることによりその違いはごく小さいと考えた．あるいは，ここで求めたスコアは，完全な発生確率情報が与えられた場合の利益の最大値を示すと考えてもよい．ここで強調しておきたいのは，この最大の利益を実現するためには，予報の精度に関する正しい情報が利用者に提供される必要があるということである．これまで，予報精度に関する情報が予報を行っている組織から公表されることは少なかった．あったとしても，全体的な予報の適中率や部分的なスコアであって，利益の計算に必要とされる分割表の形では公表されてこなかった．

コスト・ロス比が決定できる特定利用者にのみ予報が提供される場合には，予報を利用することによって利益がもたらされるかどうかが，需給者双方にとって最大の関心事であり，否が応でも評価されることになる．しかし，不特定多数の利用者に公表される予報の場合には，利用者のコスト・ロス比も多様であり，ある利用者には有益な予報でも，別の利用者には全く価値がないという状況が起こり得る．本来は予報を利用すべきでない利用者が予報を利用してしまうことによって不利益を被り，結果的に予報の価値を下げることを防ぐためには，予報を利用して判断すべきかどうかを事前に判断できるのに十分な予報精度についての情報を知らせる必要がある．したがって，特に不特定多数の利用者に予報を提供する機関においては，予報の精度に関する詳細な情報を積極的に公表することが重要である．

もちろん，社会全体で見た利益の総額は，どのような利用者集団を想定するかによって変化する．利用者集団におけるコスト・ロス比に対する数密度分布だけではなく，利益の絶対値を決めるロスLの大きさも知る必要がある（もし，完全予報に対する相対的な利益率でよいのであれば，Lは必要ない）．これを，実社会において調査するのは不可能に近い．したがって，本稿で行ったようなある程度単純なモデルを仮定することは，やむをえないと考える．しかし，もし利用者集団をある程度限定することができるのであれば，それに対応した利用者のコスト・ロス比分布を用いることにより，実際に近い利益スコアを求めることができるであろう．

スキルスコアを求める場合，予報を用いないときにかかる費用を設定する必要がある．本稿では，常に対策をとるあるいは常に対策を取らないという判断を「無技術最良予測」法にて決定した．しかし，現実には，今日の状況が明日も続く（あるいは，現在の状況が1時間後も続く）という予測法（「持続予報」と呼ばれる）を用いることも多い．もし，現象の発生確率が過去の履歴と独立であれば（現象の起こり方がランダムであれば），「持続予報」は「不偏ランダム予測」と同じスキルを持つことになる．しかし，現実には，同じような気象状況がある程度の期間継続することは，よく経験されることであり，そのような場合には，「持続予報」が「無技術最良予報」よりも高い精度を示す可能性もある．ただし，発生確率の小さい現象については，現象が持続する確率も小さいので，持続予報が気候値最良予測を上回る精度を示す可能性は小さいと考えられる．

参考文献

[1] Clayton, H. H., 1934: Rating weather forecasts. *Bull. Am. Meteorol. Soc.*, 15, 279-283.

[2] Doswell, C. A., 1990: On Summary Measures of Skill in Rare Event Forecasting Based on Contingency Tables. *Weather and Forecasting*, 5, 576-585.

[3] 菊地原英和, 1988: 気象予測の検証と評価, 気象研究ノート, (161).

[4] Murphy, A. H., 1993: What Is a Good Forecast? An Essay on the Nature of Goodness in Weather Forecasting. *Weather and Forecasting*, 8, 281-293.

[5] Peirce, C. S., 1884: The numerical measure of the success of predictions. *Science*, 4, 453-454.

[6] 立平良三, 1999: 気象予報による意思決定—不確実情報の経済価値—, 東京堂出版.

[7] Wandishin, M. S., 2002: On the relationship between Clayton's skill score and expected value for forecasts of binary events. *Meteorol. Appl.*, 9, 455-459.