



# 连续型随机变量的预报评分

马 慰 曾

(湖南湘潭地区气象台)

## 一、引 言

气象工作的主要任务之一是对某些随机天气现象进行预报。为了考核、比较不同地区、时期、人员、工具和方法的预报水平,分析情况,总结经验,必须对这些预报进行质量评定。为此目的,人们设计了各种不同的评分方法。

对一种评分方法,最主要的要求是,评分结果有代表性和比较性。也就是说,评分应能够真实地反映预报技术水平,而不受预报对象分布特点和预报员主观愿望的影响,同时应较简便并能逐次进行,便于应用。自从近代气象学诞生以来,预报评分方法的研究和改进一直受到气象科学工作者的重视。本文提出一种连续型随机变量的预报评分方法,和大家讨论。

## 二、评分公式及使用举例

设 $\xi$ 是连续型随机变量,分布函数为 $F(x)$ 。若预报 $\xi = a$ , 出现 $\xi = b$ , 则称

$$S = |F(b) - F(a)| \quad (1)$$

为这次预报的概率误差, 且评分为

$$Q = [M(S/A) - S] \times 100 + 50 \quad (2)$$

式中的 $A$ 表示事件“预报 $\xi = a$ ”,  $M(S/A)$ 是当实况和预报相互独立时(即非技术预报时) $S$ 对 $A$ 的条件期望。(2)式的实际意义就是:用实际预报的概率误差较非技术预报的平均概率误差的减小量来衡量预报技术水平。可以证明:对一定的预报对象 $\xi$ ,  $M(S/A)$ 仅是 $a$ 的函数。下面我们推导 $M(S/A) = f(a)$ 的表达式:

设 $\xi_a$ 是在条件 $A$ 下 $\xi$ 的观测值,是个随机变量。当实况和预报相互独立时, $\xi_a$ 和 $\xi$ 有相同的分布,这时有

$$\begin{aligned} M(S/A) &= M|F(\xi_a) - F(a)| \\ &= M|F(\xi_a) - F(a)| \\ &= \int_0^1 |x - F(a)| d[\phi(x)] \end{aligned}$$

这里 $\phi(x)$ 是随机变量 $\eta = F(\xi)$ 的分布函数。因 $F(x)$ 是 $\xi$ 的分布函数,故有

$$\phi(x) = \begin{cases} 0 & x < 0 \\ x & 0 \leq x \leq 1 \\ 1 & 1 < x \end{cases} \quad (3)$$

所以

$$\begin{aligned} M(S/A) &= \int_0^1 |x - F(a)| dx \\ &= \int_0^{F(a)} [F(a) - x] dx + \int_{F(a)}^1 [x - F(a)] dx \\ &= \int_0^{F(a)} F(a) dx - \int_0^{F(a)} x dx + \int_{F(a)}^1 x dx - \int_{F(a)}^1 F(a) dx \end{aligned}$$

$$= \frac{1}{2} - F(a)[1 - F(a)] \quad (4)$$

把(1)、(4)代入(2),得

$$Q = \{1 - F(a)[1 - F(a)] - |F(b) - F(a)|\} \times 100 \quad (5)$$

这就是连续型随机变量的预报评分公式。 $Q$ 的计算很简便,因为

$$P \{ \sup_{n \rightarrow \infty} |F_n(x) - F(x)| \rightarrow 0 \} = 1,$$

即当样本足够大时,可以用经验分布函数代替 $F(x)$ ,故有

$$Q \approx \left[ 1 - \frac{\alpha}{N} \left( 1 - \frac{\alpha}{N} \right) - \frac{|\beta - \alpha|}{N} \right] \times 100 \quad (6)$$

式中 $N$ 是样本数, $\alpha$ 是样本中小于预报值的个例数, $\beta$ 是样本中小于实况值的个例数。一般只要 $N > 20$ ,就可满足评分需要。

例:某台预报6月份降水量为190毫米,实况为156.7毫米。在25年的历史资料中,<190毫米的有10年,<156.7毫米的有7年。则评分为

$$\begin{aligned} Q &= \left[ 1 - \frac{10}{25} \left( 1 - \frac{10}{25} \right) - \frac{10 - 7}{25} \right] \times 100 \\ &= 64. \end{aligned}$$

## 三、Q的性质

下面我们讨论 $Q$ 的性质:

(一)当实况和预报相互独立时,有 $M(Q/A) = [M(S/A) - M(S/A)] \times 100 + 50 = 50$ ,即非技术预报得分的期望值是个常数,和 $|F(x)$ 与 $a$ 无关。这表明:

1.不存在较易得分的预报对象。无论预报对象是离差大的、偏态的旱季雨量,或是离差小的、正态的月平均气温,非技术预报得分的期望值都是一样的。 $Q$ 的这种性质使得不同预报项目评分具有比较性。

2.不存在最易得分的预报值域。不能用气候预报提高得分机会。所谓气候预报,就是每次都把众数作为预报值。 $Q$ 的这种性质很重要,它不会促使预报员为了易于得分而尽量使预报值靠近众数,或为了怕失分而不敢报极值。

3.非技术预报得分的期望值是个常数。我们把这个常数定为50分是由于考虑到通常以100为满分,以0为最低分的习惯,并避免出现负分。这就是(2)式最后一项的来源。

(二)如果我们的预报是根据某些确实的客观规律做出的,即属技术预报,在这种情况下,实况和预报就不相互独立,这时

$$M^*(Q/A) = [M(S/A) - M^*(S/A)] \times 100 + 50$$

$M^*(Q/A)$ 和 $M^*(S/A)$ 是实况和预报不相互独立时Q和S对A的条件期望, 移项得

$$\frac{M^*(Q/A) - 50}{100} = M(S/A) - M^*(S/A)。$$

上式对任意的A都成立。即平均得分减去50再除100, 就是实际预报较非技术预报平均概率误差的减小量。预报技术愈高, 平均概率误差就愈小, 平均得分就愈高, 故平均得分可代表预报技术水平, 而平均50分则是技术预报的得分下界, 愈接近50分就表示愈接近非技术预报。

(三) 当预报正确,  $F(b) = F(a)$  时, 得分为

$$Q_{正} = \{1 - F(a)[1 - F(a)]\} \times 100$$

即最高可能得分和预报值有关。当预报的是极值时,  $F(a) = 0$ 或 $1$ , 这时 $Q_{正}$ 有极大值100分。当预报的是中位数(正态时的平均数和众

数) 时,  $F(a) = \frac{1}{2}$ , 这时 $Q_{正}$ 有极小值75分。我们

知道, 愈是接近极值, 预报难度愈大, 这就是说: 得分还和预报难度有关, 预报难度愈大, 报对时得分愈高。

(四) Q的最小可能值是0分。当预报极值而出现了相反的极值时, 有

$$F(a)[1 - F(a)] = 0$$

$$|F(b) - F(a)| = 1$$

并且仅当这时, 方有 $Q = 0$ 。

(五) 通过上述分析可知: 这种评分方法具有代表性和比较性。对这种评分方法, 要想使平均得分高, 唯一途径就是尽量使预报和实况一致, 除此而外, 别无它途。

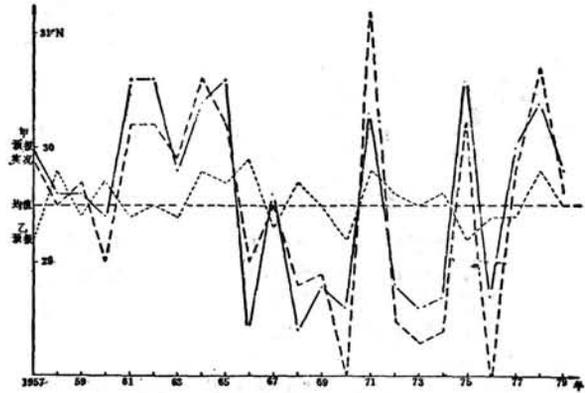
#### 四、与现行评分方法\*的比较

下面我们通过具体例子对这种评分方法和现行评分方法进行一下比较。设有甲、乙两个预报员共同预报湘潭7月份的平均气温, 历年实况和他们的预报值见附表和图。我们分别用两种评分方法去评定他们的水平。

乙的预报每次都是在平均值附近, 和实况也无明显相关, 基本属于气候预报。但按现行评分方法得分平均为78.7分, 达到了受奖标准。甲的预报水平很高, 这从图上可明显看出, 但按现行评分方法却只能得65.7分。显然, 用这样的评分去判断甲和乙的预报水

\* 湖南省现行评分方法规定, 月平均气温(°C)按表评分。

| 预<br>报 | 得<br>分 | 实<br>况 |              |             |         |      |
|--------|--------|--------|--------------|-------------|---------|------|
|        |        | 特低     | 偏低           | 正常          | 偏高      | 特高   |
|        |        | -2.1   | -2.0<br>-1.1 | -1.0<br>1.0 | 1.1-2.0 | ≥2.1 |
| 特低     |        | 100    | 80           | 30          | 0       | 0    |
| 偏低     |        | 80     | 100          | 50          | 0       | 0    |
| 正常     |        | 0      | 30           | 100         | 30      | 0    |
| 偏高     |        | 0      | 0            | 50          | 100     | 80   |
| 特高     |        | 0      | 0            | 30          | 80      | 100  |



附 图

附表 两种评分方法的比较

| 年 份  | 实测值  | 甲    |      | 乙    |      |      |      |
|------|------|------|------|------|------|------|------|
|      |      | 预报值  | 现行评分 | 本文评分 | 预报值  | 现行评分 | 本文评分 |
| 1957 | 29.9 | 30.0 | 100  | 70.1 | 29.2 | 100  | 54.4 |
| 1958 | 29.5 | 29.6 | 100  | 66.4 | 29.8 | 100  | 58.0 |
| 1959 | 29.7 | 29.6 | 100  | 70.7 | 29.4 | 100  | 63.1 |
| 1960 | 29.0 | 29.4 | 100  | 67.5 | 29.7 | 100  | 53.3 |
| 1961 | 30.2 | 30.6 | 50   | 71.3 | 29.4 | 100  | 45.8 |
| 1962 | 30.2 | 30.6 | 50   | 71.3 | 29.5 | 100  | 45.8 |
| 1963 | 29.9 | 29.8 | 100  | 71.1 | 29.4 | 100  | 54.4 |
| 1964 | 30.6 | 30.4 | 30   | 88.7 | 29.8 | 30   | 45.0 |
| 1965 | 30.2 | 30.6 | 50   | 71.3 | 29.7 | 100  | 57.7 |
| 1966 | 29.0 | 28.4 | 50   | 71.3 | 29.9 | 100  | 45.8 |
| 1967 | 29.5 | 29.6 | 100  | 66.4 | 29.3 | 100  | 76.2 |
| 1968 | 28.8 | 28.4 | 50   | 80.0 | 29.7 | 100  | 44.6 |
| 1969 | 28.9 | 28.8 | 100  | 78.6 | 29.5 | 100  | 63.1 |
| 1970 | 28.0 | 28.6 | 30   | 65.6 | 29.2 | 30   | 41.4 |
| 1971 | 31.2 | 30.3 | 30   | 80.0 | 29.8 | 30   | 36.3 |
| 1972 | 28.5 | 28.8 | 100  | 78.6 | 29.6 | 100  | 44.6 |
| 1973 | 28.3 | 28.6 | 30   | 70.0 | 29.5 | 30   | 45.8 |
| 1974 | 28.4 | 28.7 | 30   | 74.3 | 29.6 | 30   | 40.3 |
| 1975 | 30.2 | 30.6 | 50   | 71.3 | 29.2 | 100  | 45.8 |
| 1976 | 27.9 | 28.7 | 30   | 61.3 | 29.4 | 30   | 37.1 |
| 1977 | 29.8 | 30.0 | 100  | 65.8 | 29.4 | 100  | 58.8 |
| 1978 | 30.7 | 30.4 | 30   | 84.3 | 29.8 | 30   | 40.7 |
| 1979 | 29.6 | 29.8 | 100  | 66.7 | 29.5 | 100  | 67.5 |
| 平均   | 29.5 |      | 65.7 | 72.3 |      | 78.7 | 50.7 |

平, 就会得出完全错误的结论。

若用本文提出的评分方法, 则甲平均为72.3分, 乙平均为50.7分。从这个评分我们立即可以得出以下结论: 甲的平均概率误差较非技术预报减小了0.223, 有一定的预报水平。乙则为非技术预报, 没有预报水平。显然, 这个结论是正确的。

现行评分方法所以会得出错误结论, 是由于存在(下转第28页)

(上接第10页)

着最易得分预报区间。另外，它还有较易得分的预报对象。也就是说，若有两个人属于同一预报水平，譬如说都是气候预报，但由于预报对象不同，平均得分就会有很大差异。

由上述讨论可看出，本文所述方法可以显示技术

水平，具有代表性和比较性。

预报员制作天气预报时，除了对客观规律的考虑外，往往还会考虑到评分，而对预报结论作某些修改。从这个意义来说，评分方法就不单是一个检查预报质量的问题，它还可能影响预报质量。因此，预报评分办法问题应引起充分的注意。