

基于信息分配的气象数据分析法

陈瑾文

(贵州省气候中心 贵阳 550002)

提 要

从信息的角度出发,给出了一种基于信息分配的气象数据分析法。该方法对气象数据的客观规律有较好的估计。

关键词: 气象数据 过渡性 模糊划分 信息分配

1 问题的引入

我们不论是作气候分析还是天气预报,都离不开对以往观测到的各种气象数据进行统计分析,进而找出它们的内在客观规律。然而,传统上常用的一些分析方法在进行分析统计时,往往不能客观地反映所要描述对象的变化规律,这必然给我们所作的气候分析和天气预报带来不确定的因素。

我们来看这样一个简单的例子。设某地某周各天的平均云量分布如表1(注:为叙述方便,我们以一周的数据来说明,对更多的数据亦可作类似的讨论)。

表1 云量分布

星期	编号	日平均云量
一	m_1	9.0
二	m_2	7.9
三	m_3	7.7
四	m_4	7.8
五	m_5	6.0
六	m_6	1.5
七	m_7	1.0

根据气象部门的现行规定,日平均云量 ≤ 2.0 为晴天,日平均云量 ≥ 8.0 为阴天,也即是将云量按成数分为三个部分,用区间表

示为 $[0,2]$, $(2,8)$, $[8,10]$ 。现在如果要分析该地该周的云量总体分布情况,按照前面规定,在 $[0,2]$ 中的有2天,在 $(2,8)$ 中的有4天,在 $[8,10]$ 中的有1天,化为百分比后作直方图如图1。

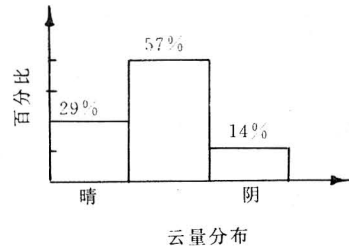


图1 云量分布直方图

从图1我们可以看出,在该周中晴天的比例远高于阴天的比例。然而,从表中不难看出,尽管日平均云量 ≥ 8 的天数仅为1天,但在 $(2,8)$ 中日平均云量接近8成的天数却有3天,而日平均云量 ≤ 2 的天数为2天,在 $(2,8)$ 中日平均云量接近2成的天数却为0天。从直观上来说,应该是阴天的比例高于晴天。特别是由于人为(或仪器)的原因,在观测到的数据中,避免不了会有较小的误差(此种误

差,我们在此称为数据的漂移)。那么,在例中那些云量接近 8 的数据实际上可能大于或等于 8,只是由于观测误差才小于 8。这样,就造成了分析的结果与客观现实有较大的差异。

2 从信息的角度讨论气象数据的特性

从前面的讨论可以看出,传统分析法得到的结论之所以与客观现实有较大的差异,主要是在统计划分时,只是把气象数据作为纯粹的数值看待,而没有考虑到数据的漂移性。实质上,我们可以这样认为,样本数据是一种信息的载体,或者更确切地说,样本点本身就是信息^[1]。因此,样本点既然是一种信息,就不能把它当作纯粹的数值来看待,而要考虑到它的漂移性和过渡性。所谓过渡性就是指每一个样本信息对于我们所作的一个划分来说,其对划分中的各类区间贡献大小不一样。如表 1 中编号为 m_2 和 m_5 的两个样本,其对区间(2,8)的贡献大小就不一样,就 m_2 而言,稍往右偏移一点就落入了阴天的范围中,从而对图 1 中的直方图有影响,而 m_5 则不然,其即使有漂移,也不会落入阴天的范围中,对图 1 中的直方图没有影响。因此,数据作为信息来说,其背后可能隐藏着各种各样的漂移,对于我们希望从这些样本中得到较好的总体估计来说,就要考虑漂移的可能,将已有的样本信息放到动态模型中去考虑。

3 动态模型的数学描述

3.1 传统方法的静态模型描述

以表 1 的数据为例,设定论域为 $[0,10]$ 。

令 $C_1 = [0,2], C_2 = (2,8), C_3 = [8,10]$

$C = \{C_1, C_2, C_3\}$

样本空间为:

$R = \{m_1, m_2, m_3, m_4, m_5, m_6, m_7\}$

则 C 与 R 的关系可用下列特征函数来表示:

示:

$$X_i(m_j) = \begin{cases} 1 & m_j \in C_i \\ 0 & m_j \notin C_i \end{cases} \quad (i = 1, 2, 3; \quad j = 1, 2, \dots, 7) \quad (1)$$

令

$$q_i = \sum_{j=1}^7 X_i(m_j) \quad (i = 1, 2, 3)$$

$$S = \sum_{i=1}^3 q_i$$

$$f_i = q_i/S \quad (i = 1, 2, 3)$$

将各值计算后得表 2。

表 2 传统方法的静态描述数据

编号	X_1	X_2	X_3
m_1	0	0	1
m_2	0	1	0
m_3	0	1	0
m_4	0	1	0
m_5	0	1	0
m_6	1	0	0
m_7	1	0	0
合计 q_i	2	4	1
频率 f_i	0.29	0.57	0.14

可以看出图 1 就是由表 2 中 f_i 构成。

3.2 动态模型描述

传统的分析方法没有考虑到样本数据对划分的贡献大小,主要在于式(1)是一个二值函数,没有涉及到从非 C_i 过渡到 C_i 的中介状态。其特征函数图象如图 2。

那么,如何描述这种过渡性呢?通常较好的办法是用模糊集合来描述。从前面我们讨论信息的过渡性中可以看出,影响最大的是划分点附近的值。如,当 m_i 的值为 0 时,一般情况下漂移到(2,8)段的可能性很小,当 m_i 的值为 2 时,漂移到(2,8)的可能性较大。所以,对每一个样本来说,我们可以认为其信息总量为 1。然后,将图 2 中的特征函数改造如图 3,得到了新的非二值函数 u_i ,该函数称为分类函数。

其实它就是任一云量属于 C_i 的隶属函数,或者说由 u_i 定义了论域上的一个模糊集合。

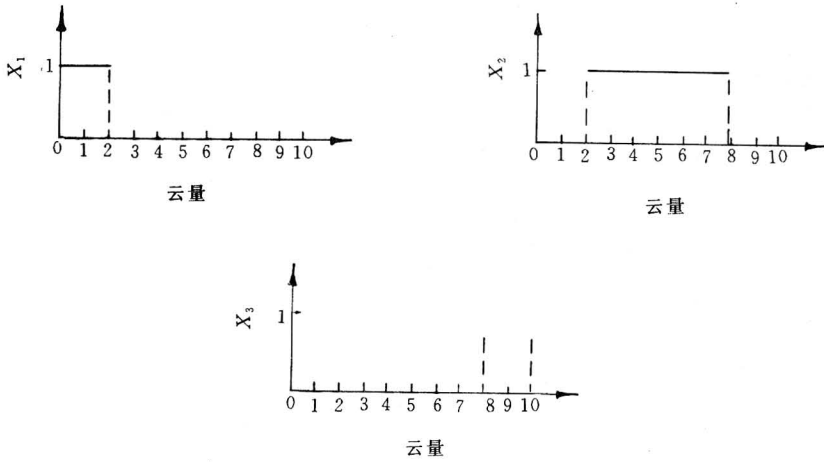


图2 静态模型的特征函数图象

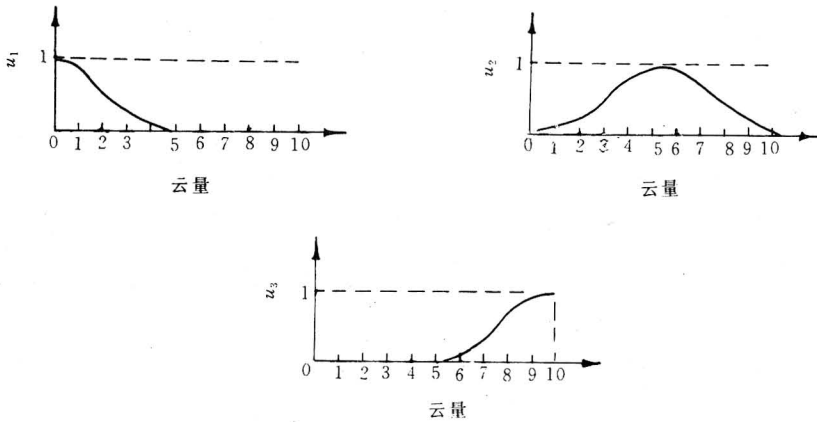


图3 动态模型的特征函数图象

对照图2和图3可以看出,对任一样本而言,其取值不再是0和1,而是在0和1之间取值。注意,改造后的函数要求对任一样本 m_i 来说, $u_1(m_i) + u_2(m_i) + u_3(m_i) = 1$, 即信息总量为1。

3.3 动态模型分析

现在对图3作分析说明如下:

① 当云量取值为0时, $u_1 = 1$,这说明属于 $[0, 2]$ 类的可能性为100%,而 $u_2 = u_3 = 0$,即属于 $(2, 8)$ 和 $[8, 10]$ 类的可能性为0;

② 当云量取值为 1 时, u_1 取值接近 1, 但不等于 1, 是接近 0 的一个小数, 这说明此时属于 $[0, 2]$ 类的可能性较大, 而 u_2 是一个较小的数, $u_3 = 0$, 即属于 $(2, 8)$ 类的可能性很小, 属于 $[8, 10]$ 类的可能性为 0。

③ 当云量取值为 2 时, u_1 和 u_2 取值接近 0.5, $u_3 = 0$, 这说明此时属于 $[0, 2]$ 及 $(2, 8)$ 类的可能性差不多, 而属于 $[8, 10]$ 类的可能性为 0。

类似可讨论其它情形。

通过上述分析可知, 对任一样本数据而言, 其信息就分配到了不同的类中, 根据取值的大小来衡量其对各类贡献的大小, 值大的贡献大, 值小的贡献小, 此亦符合云量的客观变化规律, 如云量为 0 时, 可认为其只属 $[0, 2]$ 类, 而云量为 2 时, 由于存在数据漂移性, 其属 $[0, 2]$ 类的可能性就不是很大了, 因而有

部分信息分配到了 $(2, 8)$ 类。

4 信息分配的具体实现例子

前面我们从理论上讨论了解决云量分析中数据漂移的方法。其实现核心是分类函数的确定, 不同分类函数将产生不同的划分方案, 下面给出一个分段线性的分类函数, 对表 1 例子的云量数据作分析。

分类函数取值如下:

$$u_1(x) = \begin{cases} -0.25x + 1 & , 0 \leq x < 4 \\ 0 & , 4 \leq x \leq 10 \end{cases}$$

$$u_2(x) = \begin{cases} 0.25x & , 0 \leq x < 4 \\ 1 & , 4 \leq x \leq 6 \\ -0.25x + 2.5 & , 6 < x \leq 10 \end{cases}$$

$$u_3(x) = \begin{cases} 0.25x - 1.5 & , 6 < x \leq 10 \\ 0 & , 0 \leq x \leq 6 \end{cases}$$

对应图象如图 4。

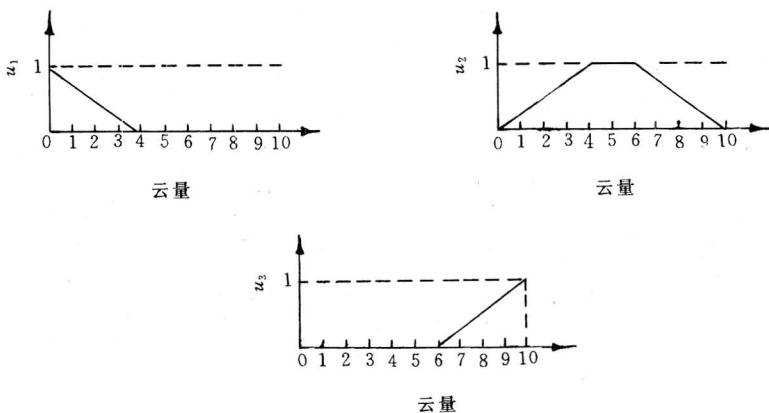


图 4 分段线性的分类函数图象

将表 1 中的数据, 按照分类函数进行计算分配, 得到表 3。

其中:

$$q_i = \sum_{j=1}^7 u_i(m_j) \quad (i = 1, 2, 3)$$

$$S = \sum_{i=1}^3 q_i$$

$$f_i = q_i/S \quad (i = 1, 2, 3)$$

f_i 即为模糊划分下总体的频率分布估计, 作成直方图为图 5。

该方法除用于云量分布分析外, 还可用于诸如雨量、温度、气压等数据分析。

表 3 云量信息分配表

星期	编号	u_1	u_2	u_3
一	m_1	0.00	0.25	0.75
二	m_2	0.00	0.53	0.47
三	m_3	0.00	0.58	0.42
四	m_4	0.00	0.55	0.45
五	m_5	0.00	1.00	0.00
六	m_6	0.63	0.37	0.00
七	m_7	0.75	0.25	0.00
	合计 q_i	1.38	3.53	2.09
	频率 f_i	0.20	0.50	0.30

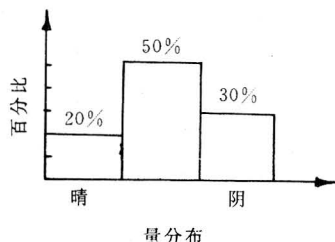


图 5 信息分配后的云量分布直方图

对比图 1 和图 5, 显然图 5 规律性比图 1 强, 从直觉上感到图 5 更可信一些。因为, 从表 1 的数据看, 阴天的比例应比晴天的大, 分析的结果基本上符合实际客观规律。

参考文献

- 1 黄崇福. 模糊信息优化处理技术及其应用. 北京: 北京航空航天大学出版社, 1994: 24~30.

The Method for Analyzing Meteorological Data Based on Information Distribution

Chen Jinwen

(The Climatic Center of the Guizhou Province)

Abstract

The method for analyzing meteorological data based on information distribution is given. The experimental results show that objective law of meteorological data can be easily obtained by the method.

Key Words: meteorological data transition information distribution