气象资料 → $\boxed{\text{预告方法}}$ → 气象预告

输入信息 → $\boxed{\text{信息变换机构}}$ → 输入信息

气象预告问题的信息分析

张学文 著

科学出版社

气象预告问题的信息分析

张学文 著

科 学 出 版 社

1 9 8 1

内 容 简 介

本书对目前统计气象预告中一些问题从信息角度作了分析和回答。它指出了测度预告因子、预告方法、预告结果的优劣和研究预告限度问题的统一的科学计值系统是计算其有关的熵和信息值。书中阐明了预告因子和预告方法在预告过程中的不同地位和作用，普查气象要素(场)的信息的时空分布的重要性等。也指出使预告结论中所含有的信息比预告因子中的信息还多的预告方法是不存在的。

本书可供气象预告人员、气象研究人员和数学工作者等参考。

气象预告问题的信息分析

张学文 著

责任编辑 杨玉梅

科学出版社出版

北京朝阳门内大街 137 号

中国科学院印刷厂印刷

新华书店北京发行所发行 各地新华书店经售

*

1981年7月第一版 开本：787×1092 1/32

1981年7月第一次印刷 印张：5 7/8

印数：0001—2,800 字数：129,000

统一书号：13031·1626

本社书号：2227·13—15

定价： 1.00 元

前　　言

信息科学是本世纪才发展起来的一门科学。但它已被有些人视为与能量和材料科学并列的现代科学技术的三大支柱之一。当代信息科学向各方面扩展的趋势与上世纪能量科学渗入到很多学科的景象是十分相似的。

气象预告问题，从统计的角度看，本质上是如何取得未来时刻的信息的问题。所以从信息理论出发研究一下预告问题是一件十分自然的事。本书就是我们在这方面的初步工作。

广大气象人员在统计预告实践活动中积累了很丰富的实践经验。为了概括这些经验和指导这些实践以尽量避免人力物力浪费，更好地揭示客观的气象统计规律，也需要有一种理论来适应它。我们发现信息理论恰好是指导分析统计预告中种种普遍性问题的一个重要理论基础。

从这种理论分析气象预告过程时，可以把预告因子看成是一个信息源。预告方法则是一个变换、传递信息的机构。预告因子(信息源)经预告方法(信息变换机构)变换成预告结论而输送出来。输出的预告结论中就包含有关于预告对象（未来的天气状况)的信息。

从信息理论分析我们可以为衡量预告任务的大小、为衡量每种预告因子最多有多大的预告本领、为衡量每种预告因子对每个预告对象有多大的预告能力、为衡量预告方法的好坏和预告质量的优劣指出客观的科学的统一的计值系统来。有了科学标准也就可以把信息理论中制约着信息传输过程的规律定量地用到分析气象预告过程中来。这对我们研究预告

问题有着普遍地指导作用。从这种分析中我们将看清预告因子和预告方法在预告过程中的不同地位和作用。我们还将看到使输出的信息(含于预告结论中)比输入的预告因子中含有的信息还要多的预告方法是不存在的；有一定预告能力的预告因子并没有无限多个;类似地说,一定个数的预告因子的预告能力也不是无限多。

依此分析我们可以看到当前在统计预告工作上对基本信息源到底含有多少信息了解的不全面、不系统也不够科学。即对基本气象要素含有多少关于预告对象的信息调查的不够。应当全面系统地完成这一基本信息源的“勘探”工作。这与在能量问题上人们努力探明地下的石油、煤、天然气的储量是同等重要的。

另一方面我们看到输出的信息比输入还多的预告方法是不存在的。这与输出的能量比输入的还多的热机不存在是类似的(即永动机不存在)。而我们现行的一些简单的预告方法在传递信息上已经是高效率的了。在此情况下气象预告质量不高很可能并不是由于预告方法不好，而是预告因子能力有限。

对种种预告问题作具体的信息分析计算，则可以经有限的步骤查明预告质量不佳,哪些是由于预告因子不佳造成的;哪些是由于预告方法不当造成的。依据这种分析才能有的放矢决定从何处着手提高预告质量,从而避免人力、物力浪费,收事半功倍之效。

预告限度问题近年来引起了广泛注意。从一定意义上说本书就是从统计角度较全面地研究了预告限度问题的科学提法和解法。

由于我国统计预告工作者对信息论的理论基础——概率论已有不少了解，所以本书中引用概率论知识之前未对之再

作介绍。仅在一、四章对信息论和随机过程分别用气象工作者易于理解的语言作了扼要知识介绍。第二、三、五章主要介绍我们的研究成果。其中第二章介绍各种气象要素(预告因子)和预告任务大小的熵值计算结果。第三章是重点,它侧重于用信息概念和规律分析气象预告过程中得出的一些重要关系的介绍。最后一章讨论气象信息的时间、空间分布、信息弥散和预告限度等问题。书后附有为计算某些信息的常用表。

本书的这些分析对其他预告问题,例如地震、水文预告等原则上也是适用的。

在完成本书过程中承新疆气象台、新疆石油研究所计算室两单位和江剑民、董忆宁、沈世镒、王为德、王跃山、刘瑞平等同志给以多方协助,在此致以谢意。

目　　录

第一章　熵与信息 ………………………………………… 1

1．对事物不肯定程度的度量——熵 ………………………… 1

2．熵函数的性质 ……………………………………………… 6

3．连续的随机变量的熵 ……………………………………… 8

4．随机矢量的熵 ……………………………………………… 13

5．条件熵 ……………………………………………………… 15

6．信息 ………………………………………………………… 18

7．正态分布下的信息收获 …………………………………… 21

8．信息在变换中的保守性 …………………………………… 24

第二章　预告因子和预告任务的熵 ………………………… 28

1．物理意义 …………………………………………………… 28

2．气象要素的熵 ……………………………………………… 30

3．气象要素场的熵 …………………………………………… 43

4．气象预告任务的熵 ………………………………………… 48

第三章　信息性质在天气预告中的某些应用 ……………… 57

1．用信息量度量天气预告质量的优劣 ……………………… 58

2．预告准确率与预告信息量的一种关系 …………………… 64

3．正态等相关因子的信息总贡献分析 ……………………… 68

4．正态互不相关因子的信息总贡献分析 …………………… 72

5．信息的可加性在解决预告问题上的应用 ………………… 74

6．预告过程分析 ……………………………………………… 75

7．预告因子问题 ……………………………………………… 80

8．预告方法问题 ……………………………………………… 84

第四章　随机过程 …………………………………………… 91

1．随机过程的概念及其分布律 ……………………………… 91

2. 随机过程的数学期望及相关函数，两个随机过程的互相关函数 …… 94
3. 随机过程的运算 …… 97
4. 马尔科夫过程 …… 98
5. 平稳随机过程 …… 109
6. 线性动力系统对随机过程的变换 …… 115
7. 时间序列的自回归 …… 123
8. 非平稳随机过程 …… 128
第五章 气象信息的时空分布 …… 133
1. 气象信息的时空分布 …… 133
2. 马尔科夫链的熵 …… 141
3. 马尔科夫链的信息弥散 …… 143
4. 平稳序列的熵 …… 147
5. 平稳序列的信息弥散 …… 152
6. 预告时效与预告质量的关系问题 …… 157
7. 预告限度问题 …… 159
附录 正态等相关因子的总信息贡献公式（3.14）的证明 …… 163
附表 1 $-p\log_2 p$ 表 …… 166
附表 2 $\log_2 N$ 表 …… 169
附表 3 正态分布时标准差(σ)与熵值(H)的关系 …… 173
附表 4 正态分布时方差(D)与熵值(H)的关系 …… 174
附表 5 正态分布时信息与相关系数 r 的关系 …… 175
参考文献 …… 176

第一章 熵与信息

很多自然科学的发展历史都说明了当一门学科逐步成熟时，经常把它的一些重要概念从定性走向定量化。事实证明，这些定量化了的东西对于探讨它们之间的逻辑、数学、物理关系有莫大好处。显然，如果至今我们对机械的、电的、光的、热的……等种种形态的能量如何定量地表示出来都搞不清，恐怕我们也难以有现代的科学技术水平。类似地讲，假如我们对气象观测和气象预告任务的大小没有一个定量地科学地测度办法；对预告因子的预告本领的大小和预告方法的优劣没有一个定量地、科学地测度办法，那么也难以对制约着气象预告过程中的种种定量关系了解清楚。我们认为信息论中的熵与信息的概念恰好就是测度这些量的一个计值系统。

这一章中我们即对熵与信息的概念、计算和它们的一些内在关系作一简要介绍。

1. 对事物不肯定程度的度量——熵

在日常生活和科学领域，人们经常看到一些事件，它们的结局是事先不能完全肯定的。这常被称为随机事件。掷一枚硬币，正面还是反面向上？掷一个骰子那一点向上？这些都事先不能肯定。观测一次气温其数值是多少？也是事先不能肯定。只要认真思考一下我们周围事物中这种事先不肯定的事件确实不少。研究这种事物固然比研究必然性的事物困难得多，但人们还是找到了从某些侧面研究它的办法。为要恰

当定量地把事物的不肯定程度的大小表示出来，信息论中就引入了熵[1]的概念。

在前例中显然掷硬币结局的不肯定性要比掷骰子结局的不肯定性要小些。现把这两个事件和它们的概率分布（分别为 1/2 和 1/6）表示如下

$$
\begin{array}{lllllll}
\text{掷硬币事件 } A: & [A_1, & A_2] & & & & \\
 & \downarrow & \downarrow & & & & \\
 & [\text{正面}, & \text{反面}] & & & & \\
 & \downarrow & \downarrow & & & & \\
\text{概\quad 率 } p(A): & [1/2, & 1/2] & & & & \\
\text{掷骰子事件 } B: & [B_1, & B_2, & B_3, & B_4, & B_5, & B_6] \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & [1\text{ 点}, & 2\text{ 点}, & 3\text{ 点}, & 4\text{ 点}, & 5\text{ 点}, & 6\text{ 点}] \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
\text{概\quad 率 } p(B): & [1/6, & 1/6, & 1/6, & 1/6, & 1/6, & 1/6]
\end{array}
$$

这里事件 A（掷硬币）有两个概率都为 1/2 的结局。事件 B（掷骰子）有六个概率都为 1/6 的结局。从直观上看 B 的结局多，不肯定性比 A 大。所以凡是概率相等的事件的个数（A 为 2，B 为 6）愈多，则不肯定性愈大，我们就说熵愈大。以 H 表示熵则有

$$H = f(k) \tag{1.1}$$

$$H(A) = f(2)$$

$$H(B) = f(6)$$

$$H(B) > H(A) \tag{1.2}$$

这里 $H(A)$ 和 $H(B)$ 分别表示 A 和 B 两事件的熵。f 是某种函数。k 是概率相等的事件的总个数。

为了决定函数 f 的形式，我们再考虑由各掷一次硬币和骰子组成的复合事件。当 A 为 A_1 时 B 有六种可能情况。A

1）信息论中这个重要概念由于与物理上热力学中熵的概念有类似性，因而信息论创始人即借用这一热力学名词来表示信息论中的概念。

为 A_2 时 B 又有六种可能情况。故复合事件由 12 个可能结局组成。依概率论中独立事件的乘法定理，每个结局出现的概率都是 $1/2 \times 1/6 = 1/12$。即仍为概率相等的事件，仅是事件个数 $k = 12$ 而已。故有

$$H(A, B) = f(12) \tag{1.3}$$

这时我们希望这种由复合事件构成的事件的熵为各个事件的熵的和(两事件独立时)。即有

$$H(A, B) = H(A) + H(B) \tag{1.4}$$

或

$$f(12) = f(2) + f(6) \tag{1.5}$$

再则当某一事件只有一个结局时，它也就不再是不肯定事件，而是出现概率为 1 的必然事件。这时希望不肯定程度变成零，即有

$$f(1) = 0 \tag{1.6}$$

不难看出，如 f 取对数（log）形式，则恰可满足(1.2)，(1.5)和(1.6)式关系。实际上信息论创造者 Shannon 等人推得的不肯定性的度量 H 恰好就是这种形式，即

$$H = \log k \tag{1.7}$$

(1.7) 式就是由 k 个概率相等的互不相容事件组成一个完备事件时每一个结局出现的不肯定性的大小的熵值。这时共有 k 个可能结局。故每一结局出现的概率 p 为 $1/k$。即(1.7)式也可以改写成

$$H = \log k = -\log(k)^{-1} = -\log(1/k)$$

$$H = -\log p \tag{1.8}$$

实际上 (1.8) 式不仅可以用于每一结局的出现概率完全相等的场合，而且可以推广到各结局出现概率不相等的场合上去。现设某地刮风与否问题 C 分为不出现大风 C_1 和出现大风 C_2 两种情况。其对应的出现概率分别为 0.9 和 0.1。这

时间出现和不出现大风的熵各是多少。显然由于不出现大风的肯定性大，故它的熵——不肯定程度就要小。而出现大风的熵就要大些。这时应有 $H(C_1) < H(C_2)$。实际如用 (1.8) 式则有 $-\log 0.9 < -\log 0.1$。这说明用(1.8)式时对各事件是否概率相等并无关系。这就是说某事件的出现概率如是 p 则它出现的不肯定程度——熵就是 $-\log p$。依对数性质不难知在 p 仅能出现于 0 到 1 之间时（概率性质）$-\log p$ 都是正值。

上例中大风出现和不出现的概率不相等熵也就不相等。我们可否进而问刮大风与否 (C) 的熵是多少？当然可以，为表示这一综合的情况最好是用出现大风与不出现大风的熵的平均值代表它。不过大风出现和不出现的概率不相等，取平均时地位也不应相等。一个妥当办法就是依其出现概率作加权平均。即出现机会多的在平均值中的地位相应增加。这样刮大风与否的熵 $H(C)$ 为

$$\begin{aligned} H(C) &= p_1 H(C_1) + p_2 H(C_2) \\ &= -p_1 \log p_1 - p_2 \log p_2 \end{aligned}$$

或

$$H(C) = -\sum_{i=1}^{2} p_i \log p_i \tag{1.9}$$

这样我们就得到各事出现概率不等时总的平均熵值。这个由各事件的熵作加权平均的算法也可以推广到有任意个结局的场合去。即有

$$H = -\sum_{i=1}^{n} p_i \log p_i \tag{1.10}$$

这里事件共有且仅有 n 个结局。这 n 个结局是互不相容事件，且各事件概率之和为 1。即有

$$\sum_{i=1}^{n} p_i = 1 \tag{1.11}$$

(1.10)式概括了掷硬币、掷骰子和刮大风的例子。它是离散的随机事件的总的平均熵的普遍表达式。对离散的随机变量的熵也用此式计算。

(1.8)式是某一特定事件的熵的表达式。(1.10)式则是 n 个互不相容的且恰好组合成一个完备事件的 n 个事件的平均熵。在信息论和本书中绝大多数都是研究这个平均熵。故(1.10)式今后要广泛应用。这里只要知道了概率分布即可算出熵值。

计算熵值时对数以什么为底都可以，只要前后统一即可。某些计算中以自然对数为底较方便。这时求得的熵叫奈特(nat)[1]。有时以 10 为底方便。不过广泛应用的是以 2 为底。这时算得之熵叫比特(bit)，即熵以比特为单位。本书后附有一个 $-p\log_2 p$ 表，可供直接从概率查得对应的熵的比特数值。利用对数换底关系

$$\left.\begin{aligned} \log_2 x &= 3.321928\log_{10} x \\ \log_2 x &= 1.442695\log_e x = 1.442695\ln x \end{aligned}\right\} \tag{1.12}$$

不难把不同单位的熵换算过去。今后计算连续的随机变量的熵时，常出现以自然对数为底的算式。这时求得之熵即为奈特。如计算时把ln改成 $\log_2$，则即成为以比特为单位的熵值。

利用(1.7)式知掷一硬币时 $H = \log_2 2 = 1$ 比特。即一个比特的熵对应的不肯定性与掷硬币时的二择一的情况相当。如某事件有 2^n 个等概率的结局，依(1.7)式熵 $H = \log_2 2^n = n$ 比特。

---

1) 奈特 (nat) 系 natural unit 的缩写。

2. 熵函数的性质

从(1.10)式看熵由 n 个 $-p\log p$ 相加而得。因 p 介于0—1之间故除了必然事件时 $p=1$ 和不可能事件时 $p=0$ 的场合下 $-p\log p$ 为0外，其他场合 $-p\log p$ 都大于零。故熵在离散场合不会为负值。

如研究某地天气状况的熵时我们把天气分成晴、多云、阴和有雨四种状态。则不难从气候资料中统计它们的出现频率以代表概率求得天气状况的熵。假设我们把有雨再进而分成微雨、小雨、中雨、大雨、暴雨五种情况。则总共实有八种情况，依此又可求得另一熵值。显然后一划分办法下结局的不肯定性大，故熵也大。但大多少可否有一般的公式表示它?！这就引出了对事物结局采用不同粗细划分时，它们间的熵的关系的问题。在我们研究不同类型和不同服务需要的气象预告问题时就常碰到这类问题。所以信息论上如给出一个一般换算公式会为我们计算省了不少事。

熵的可加性质恰好回答了这类问题。下面即用更完整的说法提清问题，并给出算式。其推导过程从略。

如最初把事物结局划分成几个，每一结局 E_i 有一确定的概率 p_i。依(1.10)可以算得一个熵值我们称之为 H_1。现设我们对其中某一结局 E_n（不一定仅是一个，也不一定恰是第 n 个）又划分成 m 个互不相容的事件 $F_1, F_2, \cdots, F_m$（参看图1.1）则可有如下事件代数的等式

$$E_n = F_1 + F_2 + \cdots + F_m$$

以 $q_1, q_2, \cdots, q_m$ 分别表示 $F_1, F_2, \cdots, F_m$ 各事件的出现概率，则应有

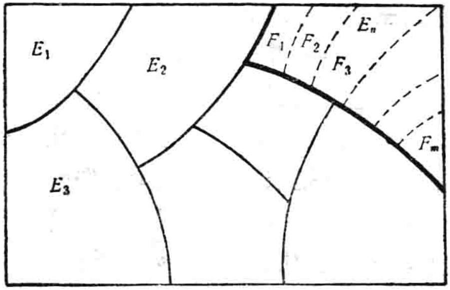

图 1.1　事件 E_n 又被分成 m 个 F 事件之和

$$p_n = q_1 + q_2 + \cdots + q_m = \sum_{k=1}^{m} q_k$$

依概率乘法定理，在 E_n 已经出现的条件下 F_k 出现的概率 $p(F_k|E_n)$ 应有

$$q_k = p_n p(F_k|E_n)$$

或

$$p(F_k|E_n) = \frac{q_k}{p_n}$$

在 E_n 已经出现的条件下，由 $F_1, F_2, \cdots\cdots, F_n$ 又组成了一个完备事件。其中第 k 个事件的出现概率即为 q_k/p_n。这样我们即有三个完备事件，而且有三个熵

$$H_1 = -\sum_{i=1}^{n} p_i \log p_i$$

$$H_2 = -\sum_{i=1}^{n-1} p_i \log p_i - \sum_{k=1}^{m} q_k \log q_k$$

$$H_3 = -\sum_{k=1}^{m} \frac{q_k}{p_n} \log \frac{q_k}{p_n}$$

它们分别代表由各个 E、各个 E 和 F，以及 E_n 已经出现时由各个 F 组成的事件的熵。

不难推得这三个熵有如下简单关系

$$H_2 = H_1 + p_n H_3 \tag{1.13}$$

这个结果很简单。它说明了对原事件中某一特定事件再进行划分时新的熵(H_2)为原熵(H_1)加上新划分事件的熵(H_3)的加权和。加权系数就是被划分的这个事件出现的概率。

当研究有 n 个结局的事件的熵时，这个熵的大小是由这 n 个结局的出现概率决定的。当这 n 个概率恰好相等时，也就是每个的出现概率都是 $1/n$ 时熵达到极大值。依(1.10)式此极大值应为

$$H = -\sum_{i=1}^{n} p_i \log p_i = -n \frac{1}{n} \log \frac{1}{n}$$

$$H = \log n \tag{1.14}$$

例如 $n=2$ 时H为 1 比特;当风按八个方位计算时它的熵最多不超过 3 比特。一般说来有 n 个结局的事件的熵小于或等于 $\log n$, 即有

$$H \leqslant \log n \tag{1.15}$$

由于此式常用,本书后附有一个 $\log_2 n$ 表。

熵还有些有趣的性质我们不讨论了。

3. 连续的随机变量的熵

(1) 连续的随机变量的熵

在前面讨论离散的随机变量的熵时，我们看到只要知道了各事件的出现概率即可用一个求和的公式(1.10)完全确定熵值。当随机变量为连续的情况下，我们对其概率的了解就常表现为已知一个概率分布密度函数。这时连续随机变量的熵就由一个与求和公式(1.10)类似的积分公式所定义，即称

$$H(X) = -\int_{-\infty}^{\infty} f(x)\log f(x)dx \tag{1.16}$$

为连续随机变量 X 的熵。这里 $f(x)$ 就是 X 的概率分布密度。

另外，我们知道在概率论中常以 $E(X)$ 表示随机变量 X 的数学期望值。即有

$$E(X) = \sum_{i=1}^{n} p(x_i)x_i \quad (\text{离散时})$$

$$E(X) = \int_{-\infty}^{\infty} f(x)xdx \quad (\text{连续时})$$

现在我们把 $-p(x_i)$ [或 $-f(x)$] 看成随机变量 X 的函数，即它也是一个随机变量（与 X 有确定关系并与 X 一起都是随机变量)，这样就可将 X 的熵看成是 $-p(x_i)$ [或 $-f(x)$] 的数学期望值。即将(1.10)和(1.16)式改写成

$$\left.\begin{aligned} H(X) &= E[-\log p(x_i)] \quad (\text{离散时}) \\ H(X) &= E[-\log f(x)] \quad (\text{连续时}) \end{aligned}\right\} \quad (1.17)。$$

这种简明的写法不仅使离散与连续的熵的公式更统一一些，而且也使我们对熵加深了一层认识。

由于气温的概率分布密度本来就接近正态分布，按概率论中的中心极限定理，它的大样本平均值就更接近正态分布了。设某地某月气温的标准差 σ 为 2℃，则依正态分布其概率分布密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-a}{\sigma}\right)^2}$$

式中 a 为气温的数学期望值。e 为自然对数的底。

将上式代入(1.16)并积分得

$$H = \ln(\sqrt{2\pi e}\,\sigma)\,(\text{纳特}) \tag{1.18}$$

如用以 2 为底的对数计算则求得之熵即以比特为单位而有