第44卷 第2期　　气象学报　　Vol. 44, No. 2
1986年5月　　ACTA METEOROLOGICA SINICA　　May, 1986

相对分布函数和气象熵*

张学文

（新疆气象台）

提　　要

针对某变量在一个物质系统中的分布状况，本文提出了相对分布函数的概念。文中论证了它与概率密度分布函数的等价性、列举了一些相对分布函数实例。由于熵与概率分布的已知关系和后者与相对分布函数等价，作者给出了计算气象要素场的熵值的方法。文中认为开展熵气象学研究可以为气象学找出新的出路。

一、引　　言

熵是热力学中引入的概念。统计物理和信息论问世以后，对它的认识又大为深化。今天熵的概念也和信息概念一样在迅速地伸向各个学科。在这里表现了熵有着极强地横贯各个领域的能力。这和能量概念在各个领域的广泛应用是十分相似的。

如何把熵的概念准确、有效、巧妙地用到气象学中来是一件很重要的事。这个问题的恰当解决有可能为气象学找出新出路。

熵是测度给定系统的总体状态的一种物理量。我们可以把地球大气看成一个系统，而气象要素在空间上的分布也就是大气的总体状态。这里既然涉及了“系统”的“状态”，所以也必然存在着一个描述气象系统的状态的熵值。熵在这里就是测度气象要素在系统中分布状态的物理量。本文先引入一个与概率分布等价的相对分布函数的概念，进而从概率分布引入气象熵。

二、相对分布函数

早在上世纪分子还仅仅是科学家的假说之时，马克斯威尔即论证了气体分子的速率分布。他的结论可用下式表达

$$f(v)=4\pi\left(\frac{m}{2\pi kT}\right)^{3/2}e^{-\frac{mv^2}{2kT}}v^2 \tag{1}$$

这里的 $f(v)$ 就是分子运动速率在 $v\pm0.5$ 范围内的分子个数占分子总个数的比值。现在

* 本文于1985年1月31日收到。

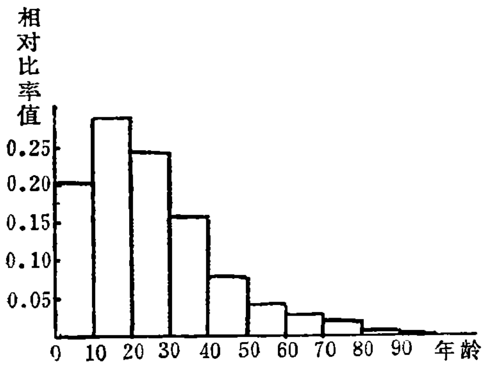

图 1　人口随年龄的相对分布

社会学家也可以根据人口调查计算出各种年龄的人占了人口总数多大的比重。图 1 就是一个人口在每十岁一个年龄组所占的人口总数的比重的一个示例图。

由(1)式可知，给了分子速度的范围，可以求出它在总体中占多大比重的值。给了年龄的大小，可以求出它在总体中占多大的比重值(图 1)。对一定系统而言，给了某变量(如速度、年龄)一个出现范围，就有一个它在总体中占的比重值。从这里我们可以说：比重值是变量值的函数。我们就把这个函数称为相对分布函数。

这里的变量当然也可以是气象上的诸如气温、气压、降水、风速……等等。这里的总体当然也可以是气象上关心的某一区城、北半球的地表面、某气压(或等高)面或者大气总体。气温在 10±0.5℃ 者占了地球表面积总体的百分之多少？风速为 5±0.5m/s 者占了大气总质量的百分比是多少?！降水为 5±0.5mm 者占了雨区面积的百分比是多少？……这都是相对分布问题。

只要有实际的地理分布，不难求出相对分布函数。例如有一张北半球气温分布图。我们可以用求积仪测出在各等温线之间包括了多少面积。不同气温占据了不同的面积百分比值。这就是气温的相对分布函数了。

一般而言，如变量出现于 x 到 $x+\Delta x$ 之间者占了地球面积的大小为 Δs。在 Δx 充分小时，可以认为 Δs 与 Δx 成正比。而这个比值在 x 为不同值时不一定相同。故一般认为这个比值是 x 的函数。我们把 Δx 充分小时上述比值除以地球总面积 S 称为相对分布函数并记为 $f(x)$。故有

$$ds=Sf(x)dx \tag{2}$$

或写成

$$\frac{ds}{dx}=Sf(x) \tag{3}$$

如令 x 代表气温，设 $x=10$ 时的 f 值为 0.01，那么依(3)式，可以看出

$$\frac{\Delta s}{S}=0.01\Delta x$$

在 $\Delta x=1$ 时 $\Delta s/S=0.01$。这表明气温在 10±0.5℃ 者占了地球总面积的百分之一。有了 $f(x)$ 我们可以求得 x 出现在任一区间者占了总系统的比重是多少。如 x 的最低值为 x_1，最大值为 x_2，那么欲求 x 出现于 $x_1\to x_2$ 者占了总体的比重值可依(3)式得：

$$\int_0^S ds=\int_{x_1}^{x_2} Sf(x)dx$$

或

$$\int_{x_1}^{x_2} f(x)dx=1$$

一般而言可有

$$\int_{-\infty}^{\infty} f(x)dx=1 \tag{4}$$

即相对分布函数对自变量从下界积分到上界，其值应当等于 1。

我们注意到相对分布函数的这种归一性质是与概率密度分布函数的归一性一致的。根据以上分析，我们可以把相对分布函数正式定义如下：在总体体积为 V 的系统中，某变量 x 在其中每一点有确定值。我们把变量值界于 x 到 $x+dx$ 之间所占据的体积值 dv 与 Vdx 的比值称为变量 x 在体积为 V 的系统中的相对分布函数 $f(x)$。即

$$\frac{1}{V}\cdot\frac{dv}{dx}=f(x) \tag{5}$$

这里的体积 V 泛指系统的总量。它可以是长度、面积、体积或质量。在研究气体分子运动时，它可以是分子总个数。在研究人口时，它是人口总数。而 x 可以是标量也可以是矢量。

相对分布函数表示了系统中的物质有多大的比例是其变量 x 值恰好在 $x\pm0.5$ 之间。或者说相对分布函数表示了系统中的物质处于某单位(1)变量范围内的相对丰度。在已知系统中物质在每一点的 x 值的情况下完全可以确知相对分布函数。已知每一点的气温可以求出气温分布函数。已知每人的年令可以求出年龄的相对分布函数。已知雨区内每一点的雨量可以求出降水的相对分布函数。而马克斯威尔从理论上推出了气体分子速率的相对分布函数。

三、相对分布函数与概率密度分布函数的等价性

设有三个系统：1. 一升 15°C 的氧气；2. 1985 年 1 月 1 日的中国人口；3. 1985 年 1 月 1 日零时的地球表面大气。这三个系统必然存在三个相对分布函数。它们分别表示，不同速度的分子占了总体多大的比例；不同年龄的人占了总体多大的比例；不同气温的空气占了总体多大的比例。现在对系统作一次抽样调查。对气体调查分子速度、对人调查年令、对空气调查气温。如抽样是随机的，那么分子速度、人的年龄、气温会在不同抽样中有不同值。我们显然应当称这些变量为随机变量。在统计学中，它们都有一个出现不同值的概率问题。这里的变量，速度、年龄、气温都是连续的随机变量。准确地说，抽得不同变量值都有一个确定的概率密度分布函数。所以这三个系统也就分别存在着三个概率密度分布函数。

我们现在要用一个假想实验来说明相对分布函数和概率密度分布函数实际上是同一个函数。在古典概率定义中，事件 A 在 n 次实验中如出现了 m 次，那么 n 充分大时的 m 与 n 的比值 (m/n) 即为 A 事件出现的概率。如人口总数为 10^9。经人口普查得知年龄在 30 岁者有 0.025×10^9 个人。即占人口总数的 2.5% 或说相对分布函数在年龄为 30 时的值为 0.025。现在进行一个假想实验，这实验是把人口普查重覆进行 k 次。每次普查每人被调查一遍。k 次普查每个人都被调查了 k 次。现在问在重覆 k 次调查中，年龄为 30 岁的出现了多少人次？显然，由于 30 岁的人占了总人数的 2.5%，所以以总人数(10^9)乘以

年龄为 30 时的相对函数值(2.5%)再乘以 k 即为 30 岁的被调查的人次数。如果问在总人口中仅作一次抽样(调查一个人)被调查者恰为 30 岁的概率是多少？这依概率定义可用下式求得

$$\text{年龄为 30 岁的概率} = m/n$$

这里的 n 是实验总次数，m 是年龄为 30 的出现次数。而 n 是充分大的。利用前述假想实验，这时的 n 即为 k 乘以人口总数。而 m 就是前述的总人口数乘相对函数值再乘以 k 即

$$\text{年龄为 30 岁的概率} = \frac{m}{n} = \frac{(\text{人口总数}) \times k \times \text{相对函数值}}{(\text{人口总数}) \times k}$$

在 k 充分大时 n 和 m 也充分大，所以上式符合概率定义。把分子分母中都有的 k 和人口总数相消后可得

$$\text{年龄为 30 岁的概率} = \text{相对分布函数值}$$

这个等式显然在年龄为任何值时都是成立的。所以我们得出：概率值与相对分布函数值在自变量为任何值时都是相等的。严格地讲，所谓年龄为 30 岁者实际上是指诞生恰好 30±0.5 年者。即我们研究的是自变量(年龄)增加单位值(1 岁)时概率的增加量。所以前述等式实应理解为在自变量为任何值时，概率密度分布函数的值恰好与同一自变量值情况下的相对分布函数值相等。

既然概率密度分布函数与相对分布函数用的是同一的自变量而在自变量为任何值，这两个函数的对应的函数值也相等。所以说相对分布函数与概率密度分布函数尽管直观物理含义不同，但实际上是同一个函数。这两个函数不仅都有前述的归一性，而且因次是一样。它们无论用解析式表达，编一个函数表或者用图形表示都是完全相同的。

如果一升氧气中分子运动速度为 400±0.5 m/s 者占了 0.1%，那么任取一个分子时，抽中的分子的运动速度为 400±0.5 m/s 者的概率也是 0.001(0.1%)。如果 30 岁的人占了 2.5%，那么任调查一个人，其年岁为 30 岁的概率也是 0.025。如果大气中有百分之一的空气的温度在 10±0.5℃的范围内，那么随机地对大气采样时，采得的空气的温度为 10±0.5℃者的概率也是 0.01。

在这里我们把变量在系统中的分布(有时是地理分布)的比例值与从系统中任抽一个样本时样本为某个值的概率值等同了起来。在这里我们看到相对分布函数与概率密度分布函数实际上是等价的。

这两个函数的等价性为我们把概率理论和熵与信息理论用于对一个系统的研究提供了具体途径。

四、气象上的相对分布函数

气象学中对于气温的分布、气压的分布、风的分布和降水的分布的种种地理分布的研究由来已久，这不是什么新课题。但是从这些分布所对应的相对分布函数是什么形态？这是一个新问题。有多少空气在向东移动？有多少空气的温度在 10±0.5℃之间？500 hPa 上位势高度是那些值所包围的面积最大？一个气旋造成了大片降水，有多少面积上的降

水比平均雨量多 8 倍？……凡此都可以在知道了对应的相对分布函数以后很容易得知。波尔兹曼研究了气体分子的能量的分布律对理解一系列气体性质作出了重要贡献。我们如果研究一下大气能量的相对分布函数是否也会有新的大气规律被揭露？

笔者在"暴雨的时面深的理论关系"一文中[1]研究了不同雨深在雨区面积上的相对分布函数。从熵自动达到最大的原理推导出雨深 x 在雨区面积上的相对分布函数为

$$f(x)=\frac{1}{\bar{x}}\exp(-x/\bar{x}) \tag{6}$$

式中的 $\bar{x}$ 为雨区的平均雨深。文中还揭示降水强度在总降水历时上的相对分布函数也有相同的形状。

在"风的分布律"一文中[2]，作者从北半球风场资料综合出一个风速绝对值(c)的相对分布函数的经验方程为

$$f(c)=\frac{4c}{\bar{c}^2}\exp(-2c/\bar{c}) \tag{7}$$

式中 $\bar{c}$ 为半球平均风速。

郭爱卿同志研究月平均北半球气压场的相对分布函数时[3]发现它遵守正态分布。戴新刚同志在研究对流层等压面位势高度在北半球的相对分布函数时发现它呈双峰分布。

五、大 气 熵

在信息论中给出一个连续的随机变量 x 的概率密度分布函数如为 $f(x)$，那么这个变量 x 的熵值 H 由下式计算

$$H=-\int_{-\infty}^{\infty}f(x)\ln f(x)dx \tag{8}$$

此时熵 H 的单位是纳特。在这里只要有了概率密度分布函数即可求得熵值。我们已经论证相对分布函数是与概率密度分布函数等价的。所以对相对分布函数依上式作积分也就求得了对应的熵值。如果研究的是气象要素的地理分布所决定的相对分布函数，显然由此算出来的熵值也就表示了气象要素场的熵值。换言之，我们找到了一个计算气象要素场的熵值的方法。一个随机变量的熵值表示了在一次抽样中变量出现什么结局的(取什么值)不确定性的大小。熵在这里度量了结局的混乱程度。一个气象要素分布决定了一个相对分布函数。由它进而算出来的熵则表示了该气象要素在分布上的不均一性。熵在这里也度量了分布的混乱程度。

在物理学和化学中，科学家利用一个系统的熵自动会达到极大值的原则推出了不少著名的方程式。我们可否在找到熵的表示方法后也借助熵原理推出某些气象规律来？熵原则在其他科学中起了巨大的作用，难道它在气象上几乎无所作为？

有了气象要素场的熵值计算办法以后，我们可以从每天的天气图上计算这个熵值

1) 1981 年刊于新疆气象杂志第 12 期。
2) 刊于 1983 年新疆气象杂志第 1 期。
3) 刊于 1984 年新疆气象杂志第 1 期。

(注意是一个分布场仅对应一个熵，而不是天气图上的每一点有一个熵值)。这是个新的统计特征量。天气形势每天都在变，这个熵也在变？戴新刚同志在研究中就发现甚至在寒潮过程中，位势场的北半球熵值就基本不变。难道这里存在着等熵过程(这里的等熵过程与大气热力学中的等熵在含义上是有原则不同的)？这无论从理论和实际气象预告上看都是十分重要的问题。

降水的地理分布一直是气象上的难题。但在前面述及的工作中，实际上作者借助新含义下的熵达到极大的原理(即热力学第二定律)反而较方便地求出了它的相对分布函数。这应当看成在气象上利用熵原理找出规律的例证。

应当指出不同的要素分布对应着不同的熵值，而且同一要素不同含义下的分布也有不同熵值。在文献[1]中作者从气象预告和气候分析角度研究过不少气象熵。本文研究的气象熵是专门与相对分布函数对应意义下的一种气象熵。它与文献[1]中论及的是不相同的。本文论及的熵可能便于联系统计力学理论在气象上的应用，而文献[1]中论及的熵则便于用于天气预告实用业务。

我们觉得相对分布函数这一概念的引入、它与概率密度分布有等价性的论证和恰当地计算气象要素场的熵等等会为概率理论、信息和熵的原理引入气象理论领域带来好处。我们相信深入研究熵在气象上的应用可以建立起熵气象学。这可能为气象学打开新的出路。

参 考 文 献

[1] 张学文，气象预告问题的信息分析，科学出版社.28—56.1981。

RELATIVE DISTRIBUTION FUNCTION AND METEOROLOGICAL ENTROPY

Zhang Xuewen

(*Xinjiang Meteorological Observatory*)

Abstract

A new concept, the relative distribution function, is brought up in this paper for the analysis of the distribution of certain variable in a material system. It is demonstrated that the function is equal to the probability density function and several examples of the relative distribution function are given. Because of the known relation between entropy and the probability density function, and the equality between the latter function and the relative distribution function, a method is given to calculate the entropy value of the meteorological element fields.