# 非平衡热力学进展

## ——曾丹苓论文选集

重庆大学热力学及节能技术研究所 选编

科学出版社
www.sciencep.com

# 非平衡热力学进展

## ——曾丹苓论文选集

重庆大学热力学及节能技术研究所　选编

科　学　出　版　社

北　京

## 内 容 简 介

本书选编了曾丹苓教授近 30 年发表在国际国内重要期刊上的具有代表性的学术论文 40 篇。内容主要涉及非平衡热力学理论及其在相关学科的应用，包括热力学基础、两相流动解的几何拓扑分析、两相流动声速及激波、相变及稳定性理论、微重力下的热毛细对流、热现象的分子动力学模拟等。

本书可供从事工程热物理及其相关领域的科研人员、高校教师以及大专院校的研究生及本科生阅读和参考。

**图书在版编目(CIP)数据**

非平衡热力学进展：曾丹苓论文选集/重庆大学热力学及节能技术研究所选编. —北京：科学出版社，2009

ISBN 978-7-03-024383-6

Ⅰ.非… Ⅱ.重… Ⅲ.非平衡(热力学)-文集 Ⅳ.O414.14-53

中国版本图书馆 CIP 数据核字 (2009) 第 053866 号

责任编辑：王志欣 闫井夫 范 勃／责任校对：包志虹
责任印制：赵 博／封面设计：王 浩

科学出版社 出版
北京东黄城根北街 16 号
邮政编码：100717
http://www.sciencep.com

中国科学院印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2009 年 4 月第 一 版 开本：787×1092 1/16
2009 年 4 月第一次印刷 印张：18 3/4
印数：1—1 200 字数：428 000

**定价：80.00 元**

(如有印装质量问题，我社负责调换〈科印〉)

## 曾丹苓教授简介

曾丹苓教授 1935 年 1 月 13 日生于四川省自贡市，1953 年从自贡市第二中学考入重庆大学动力系热能动力装置专业，1957 年大学毕业后留校任教。1962 年晋升讲师职称，1981~1984 年曾在美国布朗大学作访问学者，1983 年晋升副教授，1986 年晋升教授、博士生导师，享受国务院专家津贴，2005 年退休。2008 年 4 月 18 日因病去世。

# 序　她留下的，我们珍视的

作序非易事，字字须斟酌。此番应邀为我已故的老朋友、我国工程热力学和非平衡热力学学科的带头人、在国内外享有盛名的热力学专家和教育家——曾丹苓教授的论文选集作序，下笔之际我更是凝思迟疑，不知如何遣词，方能真切达意。

一九三五年的寒冷冬日，她在盐都自贡出生；七十三年后的明媚春天，她在学习、工作和生活了五十五年的山城重庆，走完了她为师育人、治学报国的不凡人生旅程，留下了如此丰硕的教学科研成果，也留下了朋友、同事和学生们的不尽追念。

曾丹苓教授先后担任中国工程热物理学会理事、工程热力学专业委员会副主任，《工程热物理学报》杂志编委，教育部工程热物理专业指导委员会委员，第五、第六届国家自然科学基金学科评审组成员，重庆市人民政府参事。曾获得全国教育系统“巾帼建功”标兵、四川省“巾帼英雄”及重庆市“三八红旗手”等荣誉称号。

春风化雨，孜孜不倦，呕心沥血，甘之如饴，是曾丹苓教授作为教师留给学生们的印象。她数十年如一日在三尺讲台潜心耕耘，培养出数以千百计的学生，包括五十多名博士和硕士生。她主编的《工程热力学》在全国首届优秀教材评选中被评为国家级优秀教材，她主持的教改项目曾获省级优秀教学成果一等奖。

严谨勤奋，才思敏捷，广思厚积，独辟蹊径，是曾丹苓教授作为科研工作者留给我的印象。她的研究成果涉及非平衡热力学理论、气液两相流动、相转变及界面现象、微重力下热毛细对流及微尺度下热现象的分子动力学模拟等诸多方面。她先后主持八项国家级研究课题、三项部省级课题及多项工程课题，获国家教委科技进步二、三等奖和中国高校科学技术二等奖等多项，在国内外发表论文二百余篇。她提出了建立“工程非平衡热力学”学科的构想，著有学术专著《工程非平衡热动力学》。

据我所知，工科的科研人员，很少得到诺贝尔奖。而搞非平衡热动力学的普里高津就是得奖者之一，说明此学科的创新发展前景很大。在我国从事此学科的学者很少，我曾希望，曾丹苓教授能发展此方向，为祖国争光。可惜，天不如人愿，曾教授过早仙逝。

而今，她为之操劳一生的学生们恭恭敬敬地梳理她的教学科研成果，齐心协力选编了40篇论文结集出版，以表缅怀，字里行间应可见她毕生治学育人的苦辣酸甜。若一生为师的曾丹苓教授九泉有知，当感欣慰！

这本选集，浓缩并集中反映了曾丹苓教授五十多年来对我国教学科研工作尤其是非平衡热力学方面的突出贡献，我相信这是对她最好的纪念，也是留给我们的宝贵财富。

学高行洁，桃李满园，这是她留下的，我们珍视的。

蔡睿贤

中国工程热物理学会前理事长　院士

# 目　录

## CONTENTS

# 热力学与信息论

曾丹苓

## 一、引言

狭义来说，信息论可认为是近代数学中概率论和数理统计的一个分支。在生产及人们生活中脱离不了信息的传递，将有关信息及其加工、存储、传输等基本问题以普遍化的形式提出，并寻求普遍化的解答，这就是信息论探讨的范围。近几十年信息论得到了极为迅速的发展，目前已形成一门系统的科学。它的高度概括性使它所涉及的范围远远超出了通信这类学科的范畴，而发展到许多其它的科学领域，甚至进入像生物学、生理学、心理学、社会学这样的学科领域内。同时，信息论也可用来作为热力学理论发展的基础。本文着重研讨信息论与热力学理论的相互联系，文中除系统综述了有关文献中的若干论述外，并导出了某些关系式，同时对之进行了分析。

## 二、信息论中的某些基本概念

### Ⅰ．不确定性的度量，熵

任何随机事件的基本性质是：对它们的出现与否没有完全确定的把握，在进行与这些事件相关的各种实验时，某事件的出现具有一定的不确定性。在不同的情况下，这种不确定性的程度是完全不同的。在实践上，重要的是能从数值上估计各种实验的不确定性的程度，并使它们有相互进行比较的可能。

对于具有 $K$ 个等概结局的实验而言，每个这种实验的不确定性程度是由数 $K$ 确定的。当 $K=1$ 时，实验的结局是一个必然事件，根本不是随机的。当 $K$ 较大时，则实验结果是很难予料的。可见，实验结果的不确定性程度的数值特征应依赖于 $K$，也即是说是 $K$ 的函数；当 $K=1$ 时此函数应为 0，而当 $K$ 增大时，此函数也应增大。

为了确定此函数 $f(K)$，让我们来讨论由两个独立事件 A、B 组合的复合事件，假定实验 A 具有 $k$ 个等概结局，而实验 B 具有 $l$ 个等概结局，则实验 A、B 同时进行的复合实验 AB 的不确定性程度显然应分别大于实验 A 或 B，而应该等于 A 及 B 这两个实验的不确定性程度之和。又因复合实验 AB 具有 $kl$ 个等概结局，因此我们所引出的函数 $f(K)$应满足如下条件：

$$f(kl) = f(k) + f(l)$$

这就使我们想到可以用 $\lg K$(或者 $\log_2 K$，$\ln K$，这是无关紧要的)作为有 $K$ 个等概结局的实验的不确定性的度量。即

$$f(K) = \lg K$$

此函数显然满足我们最先对函数 $f(K)$所提出的基本要求：当 $K=1$ 时，$f(K)=0$，而当 $K$ 增大时，$f(K)$增大。

---

原载于：重庆大学学报第 2 期, pp. 21-39(1979 年)。

根据我们的条件，实验总的不确定性等于 $\lg K$，则对于有 $K$ 个等概结局的实验而言，每个结局的不确定性就等于 $\frac{1}{K}\lg K=-\frac{1}{K}\lg\frac{1}{K}$。

再设想有 $N$ 个系统，其中有 $N_1$，$N_2$，…，$N_K$ 个系统分别是不可分辨的，则

$$N=N_1+N_2+\cdots+N_K$$

因此系统呈现各种排列的总数为

$$\frac{N!}{\prod_{i=1}^{K}N_i!}$$

系统对应于上述排列的不确定性程度可写作

$$\lg\frac{N!}{\prod_{i=1}^{K}N_i!}$$

将斯特林近似公式 $\ln N!=N\ln N-N$ 写作 $\lg N!=N\lg N-aN$(其中 $a$ 为换底时引入的常数)则上式可展开为：

$$\begin{aligned}\lg\frac{N!}{\prod_{i=1}^{K}N_i!}&=\lg N!-\sum_{i=1}^{K}\lg N_i!=-\left(\sum_{i=1}^{K}\lg N_i!-\lg N!\right)\\&=-\left(\sum_{i=1}^{K}\lg N_i-a\sum_{i=1}^{K}N_i-N\lg N+aN\right)\\&=-\left(\sum_{i=1}^{K}N_i\lg N_i-N\lg N\right)=-\left(\sum_{i=1}^{K}N_i\lg N_i-\lg N\sum_{i=1}^{K}N_i\right)\\&=-\sum_{i=1}^{K}N_i\lg\frac{N_i}{N}\end{aligned}$$

这样，包含在每一个系统中的不确定性程度为

$$\frac{1}{N}\lg\frac{N!}{\prod_{i=1}^{K}N_i!}=-\sum_{i=1}^{K}\frac{N_i}{N}\lg\frac{N_i}{N}\tag{1}$$

式中，$\frac{N_i}{N}$ 为呈现第 $i$ 种配置时的概率

$$\frac{N_i}{N}=P_i$$

故(1)式亦可等于

$$-\sum_i P_i\lg P_i$$

因此，一般说来，如果实验的各结局不是等概的，则对于具有 $K$ 个可能结局，$A_1$，$A_2$，…，$A_K$ 的实验 $\alpha$ 而言，如果各结局的概率分布为

实验结局　　$A_1$，　$A_2$，…，$A_K$

相应的概率，$P(A_1)$，$P(A_2)$，…，$P(A_K)$

则实验 $a$ 的不确定性可用以下数值度量

$$-\sum_i P_i \lg P_i = -P(A_1)\lg P(A_1) - P(A_2)\lg P(A_2) - \cdots - P(A_K)\lg P(A_K) \tag{2}$$

这个数我们称之为实验$\alpha$的熵，并用 $H(\alpha)$来表示。这种度量任意具有 $K$ 个可能结局的实验的不确定度的方法是由先农(Shannon)提出的，实际上它是取各结局的不确定性的平均值来作为实验 $a$ 总的不确定性的度量，即分别以等于 $P(A_1)$，$P(A_2)$，…，$P(A_K)$的概率，取值$-\lg P(A_1)$，$-\lg P(A_2)$，…，$-\lg P(A_K)$的随机变量的平均值来作为度量的。

熵 $H(a)$具有下列性质：

(1) $H \geqslant 0$；

(2) $H$ 对 $P_i$ 连续；

(3) 当所有 $P_i = \dfrac{1}{K}$ 时，$H$ 是 $K$ 的非递减函数；

(4) 当选择可以划分为两步时，则熵 $H$ 为两步熵的加权和。

$-P\lg P$—$P$ 图形如图 1 所示。由图可见，当 $P\to 0$ 时，乘积$-P\lg P$ 无限减小而趋于 0，即

$$\lim_{p\to 0}(-P\lg P) = 0$$

这就意味着，若 $a$ 实验中，结局 $A_i$ 的概率 $P(A_i)$为 0，则其所对应的项$-P(A_i)\lg(A_i)$为零，而可在熵的表达式中去掉。反之，当 $P\to 1$ 时 $\lg P$ 趋于零，若结局 $A_i$ 为必然事件，$P(A_i)=1$ 时，$\lg P(A_i)=0$，则$-P(A_i)\lg P(A_i)=0$。因此，对于整个实验 $a$ 而言，实验的熵 $H(a)$只有当概率 $P(A_1)$，$P(A_2)$，…，$P(A_K)$中之任何一个等于 1，而其余为 0 时，方等于零。显然，这是与 $H(a)$作为不确定性度量的意义完全符合的。因为在这种情况，实验根本不存在任何不确定性。

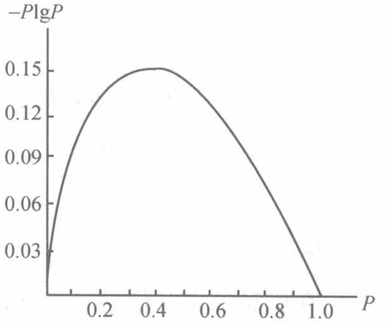

图 1

同时，我们还可进一步推论出当 $P_1=P_2=\cdots=P_K=\dfrac{1}{K}$ 时，熵 $H$ 达到最大值，等于 $\lg K$。为说明此问题，令

$$P_K = 1 - \sum_{i=1}^{K-1} P_i$$

则

$$H = -\sum_{i=1}^{K-1} P_i \lg P_i - P_K \lg P_K$$
$$= -\sum_{i=1}^{K-1} P_i \lg P_i - \left(1 - \sum_{i=1}^{K-1} P_i\right)\lg\left(1 - \sum_{i=1}^{K-1} P_i\right)$$

为求 $H$ 的极值，将上式对前 $K-1$ 个 $P_i$ 分别求导，并令各偏导数等于 0，即得

$$-\lg P_i - 1 + \lg\left[1 - \sum_{i=1}^{K-1} P_i\right] + 1 = 0 \quad (i = 1，2，\cdots，K-1)$$

由此得到

$$P_i = 1 - \sum_{i=1}^{K-1} P_i = P_K$$

即当 $H$ 取得极值时任意一个 $P_i$ 均等于 $P_K$，即

$$P_1 = P_2 = \cdots = P_K$$

又

$$P_1 + P_2 + \cdots + P_K = 1$$

故得

$$P_i = \frac{1}{K}$$

又，$H$ 对任意 $P_i$ 的二阶导数为

$$-\frac{1}{P_i} - \frac{1}{1 - \sum_{i=1}^{K-1} P_i} < 0$$

故当 $P_i = \dfrac{1}{K}$ 时，熵 $H$ 达到极大值。这即是说，在某实验 $\alpha$ 中，当各种结局的概率彼此相等时，实验是最不确定的，要判断实验的结果将最为困难，与此实验相应的熵具有最大值。

## Ⅱ. 信息的度量

在信息论中，信息量是这样来度量的：如果实验 $\alpha$ 与实验 $\beta$ 不是相互独立的，实现实验 $\alpha$ 将减小实验 $\beta$ 的不确定性程度，则在实现实验 $\alpha$ 后对实验 $\beta$ 的不确定性程度的减小量，即称为包含在实验 $\alpha$ 中关于实验 $\beta$ 的信息量，或简单地说：称为包含在 $\alpha$ 中的关于 $\beta$ 的信息。

因此

$$I(\alpha，\beta) = H(\beta) - H(\beta \mid \alpha) \tag{3}$$

这里，$H(\beta)$为实验 $\beta$ 原有的熵(无条件熵)，$H(\beta \mid \alpha)$为在实现实验 $\alpha$ 的条件下实验 $\beta$ 的条件熵，二者的差值代表通过实验 $\alpha$ 引起实验 $\beta$ 的熵的变化量，$I(\alpha，\beta)$即是包含在实验 $\alpha$ 中关于实验 $\beta$ 的信息。上式给我们提供了度量信息数值的可能性。

1. 条件熵 $H(\beta \mid \alpha)$的特性

① $H(\beta \mid \alpha) \geqslant 0$；

② 易于证明

$$\begin{aligned} H(\alpha\beta) &= H(\alpha) + \{P(A_1)H(\beta \mid A_1) + P(A_2)H(\beta \mid A_2) + \cdots \\ &\quad + P(A_K)H(\beta \mid A_K)\} \\ &= H(\alpha) + H(\beta \mid \alpha) \end{aligned} \tag{4}$$

可见，$H(\beta \mid \alpha)$实际上是实现实验 $\alpha$ 的条件下实验 $\beta$ 的平均条件熵。根据(4)式，若在实验 $\alpha$ 中所有概率 $P(A_1)$，$P(A_2)\cdots$，$P(A_K)$均不为零，即实验 $\alpha$ 有 $K$ 个结局，则当且仅当 $H(\beta \mid$

$A_1)=H(\beta\mid A_2)=\cdots=H(\beta\mid A_K)=0$ 时，即在实验 $\alpha$ 的任何结局上，实验 $\beta$ 的结局都完全成为肯定时，等式 $H(\beta\mid\alpha)=0$ 方成立。这时，由(4)式得到；

$$H(\alpha\beta)=H(\alpha)$$

③ 若 $\alpha$，$\beta$ 两个实验是独立的，则

$$H(\beta\mid A_1)=H(\beta\mid A_2)=\cdots=H(\beta\mid A_K)=H(\beta)$$

故得

$$H(\beta\mid\alpha)=H(\beta)$$

而由(4)式有

$$H(\alpha\beta)=H(\alpha)+H(\beta)$$

④ 在任何情况下， 条件熵 $H(\beta\mid\alpha)$总是在零和实验 $\beta$ 的无条件熵 $H(\beta)$之间变化即

$$0\leqslant H(\beta\mid\alpha)\leqslant H(\beta)$$

而实验 $\beta$ 的结局完全由实验 $\alpha$ 决定$\{H(\beta\mid\alpha)=0\}$，或实验 $\alpha$ 与实验 $\beta$ 无关$\{H(\beta\mid\alpha)=H(\beta)\}$，只不过是两种极端的情况而已。

2. $I(\alpha,\ \beta)$的特性

① $I(\alpha,\ \beta)\geqslant 0$。

② 根据定义，包含在实验 $\alpha$ 中关于实验 $\beta$ 的信息即是与实验 $\alpha$ 的单个结局 $A_i$ 有关的随机变量 $H(\beta)-H(\beta\mid A_i)$的平均值，也即是包含在 $\alpha$ 中关于 $\beta$ 的平均信息。

③ $0\leqslant I(\alpha,\ \beta)\leqslant H(\beta)$。

当实验 $\alpha$ 与 $\beta$ 彼此互相独立时，$H(\beta\mid\alpha)=H(\beta)$，则

$$I(\alpha,\ \beta)=0$$

而只有当 $H(\beta\mid\alpha)=0$，即当实验 $\alpha$ 唯一地确定于 $\beta$ 的结局时，方有

$$I(\alpha,\ \beta)=H(\beta)$$

一般地说，$I(\alpha,\ \beta)$是大于 0 且小于 $H(\beta)$的。

④ 包含在实验 $\alpha$ 中关于实验 $\beta$ 的信息恒等于包含在 $\beta$ 中关于 $\alpha$ 的信息，这点是易于证明的，根据(1)式

$$H(\alpha\beta)=H(\alpha)+H(\beta\mid\alpha)$$

而

$$H(\alpha\beta)=H(\beta\alpha)$$

故得

$$H(\alpha)+H(\beta\mid\alpha)=H(\beta)+H(\alpha\mid\beta)$$

所以

$$I(\alpha,\ \beta)=H(\beta)-H(\beta\mid\alpha)=H(\alpha)-H(\alpha\mid\beta)=I(\beta,\ \alpha) \tag{5}$$

这样，包含在实验 $\alpha$ 中关于实验 $\beta$ 的信息 $I(\alpha,\beta)$可称为实验 $\alpha$ 与 $\beta$ 的相互信息。联合(4)、(5)式可写出

$$I(\alpha,\ \beta)=H(\alpha)+H(\beta)-H(\alpha\beta) \tag{6}$$

这个公式在许多情况下是极为有用的。

⑤ 实验 $\alpha$ 的熵 $H(\alpha)$可定义为包含在实验 $\alpha$ 本身中关于它自己的信息(实际上，实验

$\alpha$ 完成后，其结局将完全确定，因而 $H(\alpha \mid \alpha)=0$)

⑥ 公式 $I(\alpha, \beta)=I(\beta, \alpha)$可写作

$$I(\alpha, \beta)=H(\alpha)-H(\alpha \mid \beta)$$

由此可见，包含在实验 $\alpha$ 中关于实验 $\beta$ 的信息 $I(\alpha, \beta)$不大于实验 $\alpha$ 的熵 $H(\alpha)$，也即是说不大于包含在 $\alpha$ 中关于自己本身的信息。因此熵 $H(\alpha)$也可定义为实验 $\alpha$ 可能包含的最大信息(关于 $\alpha$ 的完全信息)。实验 $\alpha$ 的熵 $H(\alpha)$等于实现这个实验之后我们所得到的信息。也即包含在实验 $\alpha$ 的一个结局中的平均信息量。

## 三、一些重要的关系式

为了将信息论与热力学紧密地、系统地联系起来，我们来对由大量粒子构成的系统导出以下重要的关系式。

为了讨论一般的情况，我们假定所讨论的系统存在以下已知的约束条件

$$\sum_i P_i = 1 \tag{7}$$

$$\sum_i P_i g_r(x_i) = \langle g_r \rangle \quad r = 1, 2, 3, \cdots \tag{8}$$

这里，$P_i$ 为系统各个别可能状态的概率，$g_r(x_i)$及$\langle g_r \rangle$分别为所给出的各种约束的可能取值及这些可能值的数学期望(或某种平均值)。

下面我们来讨论熵达最大值(亦即包含最大平均信息量的一种特殊情况)时的概率分布。我们知道在一个信号中包含的平均信息为：(在以下热力学的熵的表示式中，我们加入了常数 $K$ 并取自然对数，且按热力学的习惯，用符号 $S$ 表示熵。

$$S = -K\sum_i P_i \ln P_i \tag{9}$$

为了求在满足约束条件(7)，(8)下 $S$ 的极大值，我们采用格拉朗日乘子法，用乘子 $-(\lambda_0-1)$乘(7)式(这里比乘 $\lambda_0$ 要方便些)，得到

$$-(\lambda_0 - 1)\left(\sum_i P_i - 1\right) = 0$$

再用 $-\lambda_r$ 乘(8)式，并对 $r$ 求和，得到

$$-\sum_r \lambda_r \left[\sum_i P_i g_r(x_i) - \langle g_r \rangle\right] = 0$$

又由于在(9)公式中 $K$ 为某常数，故求 $S$ 的极值相当于求

$$-\sum_i P_i \ln P_i$$

的极值。将以上三个函数相加， 即可作出以下拉格朗日函数

$$-\sum_i P_i \ln P_i - (\lambda_0 - 1)\left(\sum_i P_i - 1\right) - \sum_r \lambda_r \left[\sum_i P_i g_r(x_i) - \langle g_r \rangle\right] \tag{10}$$

在产生极值时此辅助函数对各变量的偏导数均应为 0，将辅助函数对各 $P_i$ 取偏导数并令之等于 0，得到

$$-(\ln P_i+1)-(\lambda_0-1)-\sum_r \lambda_r g_r(x_i)=0$$

$$-\ln P_i-\lambda_0-\sum_r \lambda_r g_r(x_i)=0$$

则

$$P_i=e^{-\lambda_0-\sum_r \lambda_r g_r(x_i)}=e^{-\lambda_0-\lambda_1 g_1(x_i)-\lambda_2 g_2(x_i)-\cdots} \tag{11}$$

这即是具有从 0 阶到 r 阶，即 r+1 个拉格朗日乘子的分布函数。

现在来确定各阶拉格朗日乘子，将(11)式代入(7)式得到

$$\sum_i P_i=\sum_i e^{-\lambda_0-\sum_r \lambda_r g_r(x_i)}=1$$

$$e^{-\lambda_0}\cdot\sum_i e^{-\sum_r \lambda g_r(x_i)}=1$$

$$e^{-\lambda_0}=\sum_i e^{-\sum_r \lambda_r g_r(x_i)} \tag{12}$$

所以

$$\lambda_0=\ln\sum_i e^{-\sum_r \lambda_r g_r(x_i)} \tag{13}$$

为了得到其它拉格朗日乘子的方程，取偏导数$\dfrac{\partial\lambda_0}{\partial\lambda_r}$，则

$$\frac{\partial\lambda_0}{\partial\lambda_r}=\frac{\partial\ln\sum_i e^{-\lambda_r g_r(x_i)}}{\partial\lambda_r}=-\frac{\sum_i e^{-\sum_i \lambda_r g_r(x_i)} g_r(x_i)}{\sum_i e^{-\sum_r \lambda_r g_r(x_i)}}$$

又利用(12)式，上式可改写为

$$\frac{\partial\lambda_0}{\partial\lambda_r}=-e^{-\lambda_0}\sum_i e^{-\sum_r \lambda_r g_r(x_i)} g_r(x_i)$$

将 $e^{-\lambda_0}$ 放入求和符号内，得到

$$\frac{\partial\lambda_0}{\partial\lambda_r}=-\sum_i P_i g_r(x_i) \tag{14}$$

上式即 $g_r(x_i)$之数学期望$\langle g_r\rangle$，即

$$\langle g_r\rangle=-\frac{\partial\lambda_0}{\partial\lambda_r} \tag{15}$$

由(13)式可知，零阶拉格朗日乘子 $\lambda_0$ 是其它拉格朗日乘子的函数。显然，根据(15)式，一旦找出了 $\lambda_0$ 的函数，则可通过简单微分而得到各期望值$\langle g_1\rangle$、$\langle g_2\rangle$、…。

在给定拉格朗日乘子 $\lambda_0$ 后，不仅函数 $g_r(x_i)$的期望值可以算得，其方差也可由此确定。

利用(14)式， 求 $\dfrac{\partial^2\lambda_0}{\partial\lambda_r^2}$ 得

$$\frac{\partial^2\lambda_0}{\partial\lambda_r^2}=-e^{-\lambda_0}\sum_i g_r(x_i)e^{-\sum_r\lambda_r g_r(x_i)}[-g_r(x_i)]$$

$$-\sum_i e^{-\sum_r\lambda_r g_r(x_i)}g_r(x_i)e^{-\lambda^0}\left(-\frac{\partial\lambda_0}{\partial\lambda_r}\right)$$

$$=\sum_i e^{-\lambda_0-\sum_r\lambda_r g_r(x_i)}[g_r(x_i)]^2-\left(-\frac{\partial\lambda_0}{\partial\lambda_r}\right)\sum_i e^{-\lambda_0-\sum_r\lambda_r g_r(x_i)}g_r(x_i)$$

$$=\sum_i P_i[g_r(x_i)]^2-<g_r(x_i)>^2=<[g_r(x_i)]^2>-<(g_r x_i)>^2$$

$$=\sigma^2[g_r(x_i)]$$

故得方差为

$$\sigma^2[g_r(x_i)]=\frac{\partial^2\lambda_0}{\partial\lambda_r^2} \tag{16}$$

将由(11)式得到的 $P_i$ 值代入(9)式得到不定度为：

$$S=-K\sum_i P_i\ln P_i$$

$$=-K\sum_i P_i\left[-\lambda_o-\sum_r\lambda_r g_r(x_i)\right] \tag{17}$$

$$=K\lambda_0\sum_i P_i+K\sum_i P_i\sum_r\lambda_r g_r(x_i)$$

利用(8)式，(17)式得

$$S=K\left[\lambda_0+\lambda_1<g_1>+\lambda_2<g_2>+\cdots\right] \tag{18}$$

$$=K(\lambda_0+\sum_r\lambda_r<g_r>) \tag{19}$$

这即是系统所包含的平均信息量的最大值。

## 四、信息论对各热力学定律的说明

### I. 温度及第零定律

下面我们来讨论拉格朗日乘子的意义，从而导出温度及第零定律。

考虑有 $N$ 个系统，令 $N_i$ 代表具有能量为 $\varepsilon_i$ 的系统数，而 $N$ 个系统具有的总能量为 $NE$，则可写出

$$\sum_l N_l=N$$

$$\sum_l N_l\varepsilon_l=NE$$

用 $N$ 除上式，得

$$\sum_l P_l=1 \tag{20}$$

$$\sum_l P_l \varepsilon_l = <E> \tag{21}$$

为求 $S=-K\sum_l P_l \ln P_l$ 的极大值，引入拉格朗日乘子 $(\psi-1)$ 及 $\beta$ ，而写出

$$\mathrm{d}S=-K\sum_l(\ln P_l+1)\mathrm{d}P_l$$

$$(\psi-1)\sum_l \mathrm{d}P_l=0$$

$$\beta\sum_l \varepsilon_l \mathrm{d}P_l=0$$

令 $\mathrm{d}S=0$ ，去掉 $K$，而将以上三式相加，得到

$$\sum_l(\ln P_l+1+\psi-1+\beta\varepsilon_l)\mathrm{d}P_l=0$$

$$P_l=\mathrm{e}^{-\psi-\beta\varepsilon_l} \tag{22}$$

利用前面已导出的关系式，可得到

$$\psi=\ln\sum_l \mathrm{e}^{-\beta\varepsilon_l} \tag{23}$$

$$S=K[\psi+\beta<E>] \tag{24}$$

$$<E>=-\frac{\partial\psi}{\partial\beta} \tag{25}$$

$$\frac{\partial<E>}{\partial\beta}=-\frac{\partial^2\psi}{\partial\beta^2}=-\sigma^2(E) \tag{26}$$

由以上的关系可知，如果系统的 $\varepsilon_l$ 为已知，则能量的期望值 $<E>$ 仅随特性 $\beta$ 变化，且(26)式可见，$<E>$ 为 $\beta$ 的单调递减函数 $\left(\text{因}\,\sigma^2(E)\,\text{恒为正值}\,\frac{\partial<E>}{\partial\beta}<0\right)$。因此，当外部坐标一定，因而 $\varepsilon_l$ 一定时，$\beta$ 可视为能量期望值的指标。即可视为反映系统中能量的概率分布的一个数学特征。这即是拉格朗日乘子 $\beta$ 的统计特征。

现考虑两个系统 $A$ 和 $B$，最初，其能量的期望值分别为 $<E_l>$ 和 $<E_R>$。对以上两系统根据前面导出的关系可分别绘出图 2 中的图线(a)、(b)。由图可见，$<E_A>$ 及 $<E_B>$ 分别随 $\beta_A$、$\beta_B$ 单调递减。今若使 $A$、$B$ 系统相互接触，彼此发生能量交换，则达平衡时复合系统 $C(AB)$ 的不确定性增加。对系统 $C$ 同样可作出图 2(c)所示图线。这时，$A$、$B$ 系统应是同一特性来描述的。而在图 2(c)上仅有一个 $\beta$ 可与 $<E_C>$ 相对应。这样，如果系统 $A$，$B$ 在相互作用前具有不同的 $\beta$ 值，则在相互作用过程中将引起一个系统 $\beta$ 的下降，而另一个系统 $\beta$ 上升，直至二者等于某共同的 $\beta$（即 $\beta_C$）值时为止。同时，由图可见，能

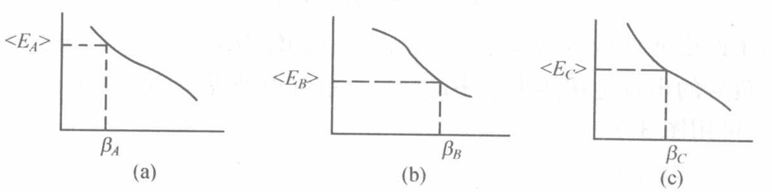


图 2

量将从具有较低 $\beta$ 值的系统流向具有较高 $\beta$ 值的系统。

因此，对于由大量粒子组成的系统而言，若处于平衡时，必有某一特征 $\beta$ 存在。当两个具有不同特征 $\beta$ 的系统相互接触时，则相互作用将使它们达到某共同的 $\beta$ 值时为止，我们称此种平衡为“热平衡”。显然，若 $A$ 与 $B$ 处于热平衡，且 $A$ 与 $C$ 也处于热平衡，则 $B$ 和 $C$ 必然也处于热平衡，这即是第零定律所提示的事实。

对于理想气体而言，不难证明

$$\beta=\frac{1}{KT} \tag{27}$$

所以我们说 $\beta$ 对于某给定系统而言，是说明其内部能量的概率分布的一个数字特征，我们把它的某单值函数 $T=\dfrac{1}{\beta K}$ 叫做温度。所以，从统计观点来看，温度是用来确定给系统概率分布特性的一个统计参数。

## II. 热力学第一定律

前面我们分析了拉格朗日乘子 $\beta$ 的意义，我们说它是能量概率分布的某种数学特征，我们称它作温度的某种度量。下面我们来讨论另外一个拉格朗日乘子 $\psi$ 的涵义，并由此导出热力学第一定律的统计解释。

1. 交换能量的两种方式——热与功

在经典热力学中常把热和功作为两个基本的概念，然而，在信息论中，我们把热和功视为导出量。

前已得到

$$\langle E\rangle=\sum_l P_l\varepsilon_l \tag{28}$$

将上式微分得

$$\mathrm{d}\langle E\rangle=\sum_l \varepsilon_l \mathrm{d}P_l+\sum_l P_l \mathrm{d}\varepsilon_l \tag{29}$$

采用以下符号

$$\mathrm{d}Q_r=\sum_l \varepsilon_l \mathrm{d}P_l \tag{30}$$

$$\mathrm{d}W_r=-\sum_l P_l \mathrm{d}\varepsilon_l \tag{31}$$

则得

$$\mathrm{d}\langle E\rangle=\mathrm{d}Q_r-\mathrm{d}W_r \tag{32}$$

可见，$\mathrm{d}Q_r$ 为不改变 $\varepsilon_l$ 时改变<E>的一种方式，而 $\mathrm{d}W_r$ 为不改变 $P_l$ 而改变<E>的另一种方式。(31)式中负号的出现是按热力学中的规定，系统由外界得到功时取负号。两种能量交换方式的区别可用图 3 表示。

这里，图 3 表示 $\varepsilon_l$ 随 $l$ 的变化，$\mathrm{d}W_r$ 型的变化意味着图 3(a)中各图线的变化；而图 3(b)表示 $P_l$ 随 $l$ 的变化。$\mathrm{d}Q_r$ 型的变化意味着图 3(b)中图线的变化。图 3(c)则表示 $P_l\varepsilon_l$ 随 $l$ 的