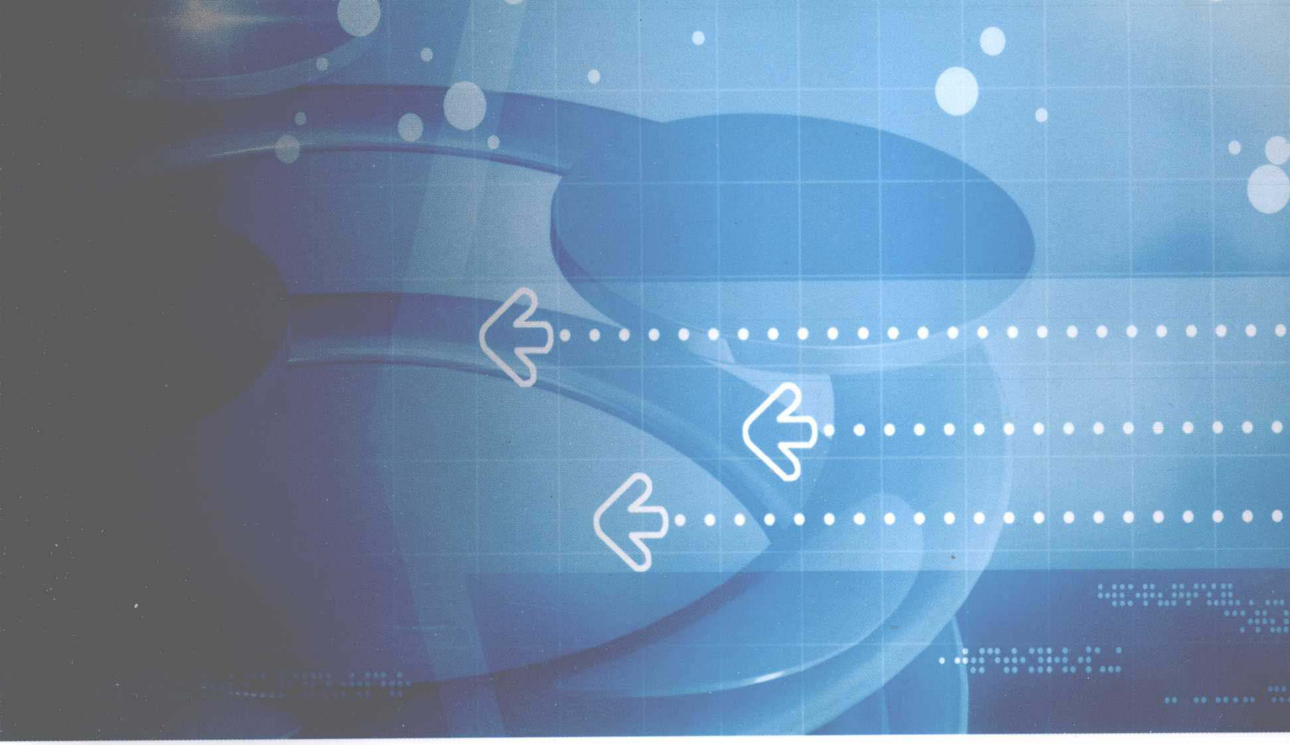

复杂系统方法学与中医证候建模

西广成 著

科学出版社
www.sciencep.com

复杂系统方法学
与中医证候建模

西广成　著

科学出版社
北　京

内 容 简 介

在有关中医药学的973计划科研项目中，对中医证候及其与疾病、方剂的相关性开展研究，获得一些阶段性成果。本书是这些成果的总结，主要内容包括：①证候的发生机制、存在性及主要特点；②证候、疾病、方剂复杂系统广义特征指标；③疾病映射模型及其算法；④中医广义症状（四诊信息、西医病理信息、各类理化指标等）与证候之间的相关性；⑤中医证候和方剂之间的相关性；⑥症状对证候的贡献度；⑦中医大规模流行病学数据的无监督聚类方法；⑧中医辨证论治的智能系统模型等。

本书的特点是理论与实践相统一，具有科学性和原创性。本书可供从事复杂系统、生物信息学、脑模型和中医药的科研与教学工作人员参考；还可作为复杂系统建模与分析领域高年级本科生及研究生的参考书。

图书在版编目(CIP)数据

复杂系统方法学与中医证候建模/西广成著. —北京：科学出版社，2010

ISBN 978-7-03-028738-0

Ⅰ.①复… Ⅱ.①西… Ⅲ.①辨证论治-系统建模 Ⅳ.R241-39

中国版本图书馆CIP数据核字（2010）第163405号

责任编辑：张海娜/责任校对：赵桂芬

责任印制：赵 博/封面设计：耕者设计工作室

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

双青印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2010年8月第 一 版　开本：B5（720×1000）

2010年8月第一次印刷　印张：16 1/2　彩插：1

印数：1—2 000　字数：315 000

定价：58.00元

前　言

2004年9月份以来，我和我的学生有幸参加973项目在中医药领域中“证候规范及其与疾病、方剂相关的基础研究”（2003CB517100）和“络病学说与针灸理论的基础研究”（2005CB523300）的科研工作，任务分别是基于抽象神经自动机理论和复杂系统建模中的熵方法，研究中医证候及其与疾病、方剂的相关性和络病与血管病变相关性的复杂系统建模分析。在这几年时间里，研究工作取得了一些成果，对这些成果进行总结并向学术界介绍和展示，向那些在学术研究道路上关心我、鼓励我但至今未曾谋面的学术前辈进行汇报（相关信件请见附录）；另外，我已退休，行将淡出学界“江湖”，向祖国和人民汇报我的部分科研成果，这一直是我真诚和强烈的愿望，也是我写本书的主要意图。

本书共包括11章。其中第0章是分析信息论的经典初步介绍，此章的目的有两个：一是对理解本书后面的内容有帮助，为初学者看本领域国内外水平提供一种视角；二是为那些立志在数学信息论特别是在将信息论应用于复杂系统研究领域从事科研工作并渴望取得高水平研究成果的本科生及研究生准备一些导言性的内容。第1章和第2章是自20世纪80年代初至2003年本书作者以唯一作者发表的部分研究成果。主要内容是复杂系统研究中的熵理论与方法，以及抽象神经自动机理论和智能系统理论，这些内容在目前及今后相当长的时期内都是复杂系统建模与分析研究领域的前沿热点，一旦取得实质性突破，有可能对该领域乃至整个科学领域形成重大影响。这两章的内容可简单概括为如下三点：①成果“复杂系统分划（聚集）的熵方法及其应用”于1993年7月份评审（中国科学院和科技部两级评审），于1995年选为国家级优秀科技成果。该成果是一项有创建性的复杂系统研究中的新理论、新方法，对复杂系统研究很有价值，为开展社会—生态—经济复杂系统区域划分的理论与实践的研究，以及脑神经系统、复杂Markov系统的研究提供了新的有效途径。在科学研究领域，分划（聚集）的思想具有一般的意义，而熵分划抓到了问题的本质，这对于深刻理解熵理论的人是不难理解的。②基于借助条件概率及其正则性描述的随机场理论、神经网络理论、神经生理学和神经心理学理论提出并初步建立的抽象神经自动机理论，是关于人脑模型思维的理论，抽象神经自动机是无穷维随机神经网络，是认识和思维的机器。③通过相对熵最小化理论与神经网络理论相结合，给出智能系统的一般理论框架，这一理论的精华是给出并证明命题：在智能系统中，熵与智能共增减。这一命题不仅直接导出了两种训练神经网络的学习算法（Boltzmann机学习

算法和确定性 Boltzmann 机学习算法)，而且具有更深刻更巨大的价值：它可能彻底抹去热力学第二定律在人们心灵上投下的阴影，满怀信心地面对生活。

全书其余章节内容是上述理论成果在中医证候及其与疾病、方剂相关性建模中的应用、发展和完善，主要工作是由博士生赵亚丽、陈建新、孙占全、陈静等完成的。本书大部分论述的素材来自我们共同发表的论文和他们的学位论文。他们把用复杂系统科学语言和中医理论相结合的学术论文在国际重要学术杂志和重要学术会议上发表，引起了人们的强烈兴趣，为世界人民特别是学术界理解中医提供一种新的科学视觉。第 3 章介绍中医复杂系统中广义症状和证候之间的相关性研究；第 4 章介绍中医复杂系统研究中的熵分划方法研究；第 5 章介绍一种基于熵的复杂系统综合评价指标的建立方法；第 6 章介绍特征提取方法及其在中医中的应用；第 7 章介绍中医流行病学数据的无监督分类方法研究；第 8 章介绍中医辨证的神经计算模型；第 9 章介绍熵理论在疾病诊断标准研究中的应用；第 10 章介绍熵在多时点数据中的演化及其应用。

向那些在复杂系统研究工作中曾经对我寄予希望并给予热情鼓励但至今未曾谋面的学术界前辈表示衷心感谢!

感谢戴汝为院士，在我多年的科研工作中，他对同事充满关爱的支持和鼓励让我感到温暖并在重要时刻增强了克服困难的信心和力量，他是我与中医“相识”的第一位介绍人。感谢王永炎院士和黄璐琦博士，他们让我与中医“结缘”，在将自己的学科知识与中医的结合中获益匪浅。感谢 973 项目“证候规范及其与疾病、方剂相关的基础研究”和“络病学说与针灸理论的基础研究”项目组的全体老师和同仁，特别是这两个项目的首席科学家王庆国教授和吴以岭院士，本书中关于中医证候建模的一些成果直接得益于他们在项目的实施过程中及时提出问题并在问题讨论中的不吝赐教，以及在资料数据等方面的大力支持。本书所涉及的科研工作得到国家自然科学基金和 973 计划的资金资助，在此表示由衷的感谢。我的学生赵亚丽、陈建新、孙占全和陈静等在复杂系统建模与分析和中医理论相结合的研究中，做出了出色的成果，感谢他们为本书的完成付出的辛勤劳动。另外，孙占全、陈建新、陈静和郑荣尧等在为书稿资料准备及录入方面做了大量的工作，对他们深表谢意。

目　录

第 0 章　熵一般化描述中的某些基本概念

在初步了解关于离散随机变量、连续随机变量以及随机过程的熵概念之后，可以从更一般的意义上对熵进行更深刻的描述，本章的简单介绍对理解本书后面的内容，能提供些许启发，为初学者看本领域国内外水平提供一种视觉，希望引导善于思索的初学者进入更高的理论层次。

0.1　熵发展史概述

“熵”一词是由法国物理学家 Clausius 于 19 世纪中叶（1854～1864）作为热力学概念而引入的，随后产生了热力学第二定律的熵表达，即孤立系统的熵之非减原理。在基于概率论的原子论热力学——统计力学中，Boltzmann 等对熵（记为 S）作出了微观的解释。在全同的气体粒子的组成的系统中，有著名的公式 $S=k\ln\Omega$，其中 k 为 Boltzmann 常数，Ω 为与某一宏观状态对应的微观状态数，或者是系统的热力学几率。H 定理将 S 表示为 $S=-k\int f\ln f$，其中 f 是气体粒子相位空间上的概率分布密度，这个公式表示系统的非平衡态熵原来就是分布密度的对数对该密度本身的积分取负值[1]。

将系统相位空间分划为有限个单元，然后应用概率矢量（元素为 p_i，系统的宏观态）表示在每个分划单元中粒子的比例。这时，Boltzmann 熵是关于系统状态的不确定性程度的度量，它可以被与系统宏观状态对应的系统状态（微观状态）数的对数来度量。对于由 n 个粒子组成的系统，这个数等于 $\left(\prod_i (np_i)!\right)\Big/n!$。Stirling 公式给出熵的连续统极限为

$$\lim_{n\to\infty}\frac{\left(\prod_i (np_i)!\right)\Big/n!}{n\cdot n!}=\sum_i p_i\ln p_i$$

系统的最可几状态可通过最大化 $\sum_i p_i\ln p_i$ 来实现。

Shannon 考虑了一个有限的离散符号集合，其第 i 个符号被赋予概率 p_i，寻找一个函数以测量从这些符号中选择一个符号的不确定性，忽略一个常数该函数被确定为 $H=\sum_i p_i\ln p_i$，它是关于 p_i 的唯一的连续函数，当 $(p_1,\cdots,p_n)=\left(\frac{1}{n},\cdots,\frac{1}{n}\right)$时，随 n 增大而单调增大，它是 $\ln p_i$ 的加权平均，对于联合独立分

布具有可加性，如果选择分为相继的两个步骤，那么原先的 H 将等于各个 H 值的加权和。因此，某概率空间一个有限或可列可测分划 ξ 的 Shannon 熵是

$$H(\xi):=H_{\mu}(\xi):=-\sum_{C\in\xi}\mu(C)\ln\mu(C)\geqslant 0$$

式中，$0\ln 0=0$，对于可列的 ξ，这个熵可能有限也可能无穷大。

Shannon 的原创性的工作成为数学信息论的基础。这一理论与平稳随机过程理论的关系对于概率论专家是很明显的。1956 年，Khinchin 以平稳随机过程的熵为中心，对信息论给出了极为漂亮的严格的处理。他充分认识到，具有有限状态空间的平稳随机过程与具有有限分划 ξ 的保测变换之间的等价性，从而发展了熵的基本理论。对保测变换 T 定义联合分划

$$\xi_{-n}^{T}:=\bigvee_{i=1}^{n}T^{1-i}(\xi)$$

式中，$\xi\vee\eta:=\{C\cap D|C\in\xi,D\in\eta\}$。称

$$h(T,\xi):=h_{\mu}(T,\xi):=\lim_{n\to\infty}\frac{H(\xi_{-n}^{T})}{n}$$

为变换 T 关于分划 ξ 的度量熵。容易看出，上述极限是存在的。借助于条件熵

$$H(\xi|\eta):=-\sum_{D\in\eta}\mu(D)\sum_{C\in\xi}\mu(C|D)\ln\mu(C|D)$$

式中，$\mu(C|D):=\mu(C\cap D)/\mu(D)$，有限状态随机过程（保测变换关于给定的有限分划）的熵可等价地定义为给定过去完备知识，一步所得的平均信息量，即

$$h(T,\xi)=H(\xi|\xi_{-\infty}^{T})$$

这时，熵提供了描述随机过程复杂性一个参数，该随机过程具有有限多个状态。给定可测空间（X,μ）的一个可测变换 T，通过固定一个 X 的有限分划 ξ，将其元素视为具有 Card（ξ）个符号的一个字母表中的字母，并且借助于这个分划将变换编码，人们就可制造这样的随机过程。该随机过程的熵通常表示为 $h(T,\xi)$。同样的变换可用许多不同的分划编码，相应随机过程的熵可能不同。Kolmogorov 认识到这可以用于定义一个描述保测变换固有复杂性的量，即一个不变量。如果属于 $\bigvee_{i=-n}^{n}T^{-i}(\xi)$ 的分划集合在具有 Rokhlin 度量

$$d_{\mathrm{R}}(\xi,\eta):=H(\xi|\eta)+H(\eta|\xi)$$

的有限熵分划的集合中是稠密的，那么具有有限熵的一个可测分划 ξ 是一个保测变换 T 的生成元。Kolmogorov 指出，一个保测变换 T 的所有生成元有相同的熵，并将 T 的熵定义为这个公共的值（如果 T 有一个生成元），否则，T 的熵是 ∞。Sinai 建立了一种自然的方法，使得这一概念有了新的发展并将保测变换的熵（或度量熵）定义为 $\operatorname{Sup}_{\xi}\{h(T,\xi):H(\xi)<\infty\}$，即 $h(T):=h_{\mu}(T):=\operatorname{Sup}_{\xi}\{h(T,\xi):H(\xi)<\infty\}$，人们常称此熵为 K-S 熵。

熵诞生于热力学系统，“降生”伊始，将热力学第二定律定量表示出来，使

其蕴涵更丰富深刻，意义更伟大深远；熵进入统计力学系统，建立了宏观与微观之间的联系；熵进入通信系统，使得信息定量表述成为可能；熵进入动力系统理论，成为现代动力系统理论中的关键概念，创造了灿烂辉煌的业绩。

在许多令人感兴趣的情况下，人们获得了如下的公式：

$$h_{\text{K-S}} = \sum_{\lambda_i > 0} \lambda_i$$

式中，λ_i 是系统的 Lyapunov 指数。$h_{\text{K-S}} > 0$，系统混沌。当然，有些情况下，混沌拯救了热力学第二定律，从而拯救了熵。

通常所说的信息论主要是指 Shannon 概率信息论，其代表函数是概率熵。但现实世界中，信息并非必然有概率性质。在某些非概率情况下，存在所谓的“非概率信息”。而 Shannon 熵无法度量这种非概率信息。于是，人们又提出了无概率信息、定性信息、偶发信息、模糊信息等一系列信息熵的概念和公式。

天体物理学家 Eddington 指出：“如果让我们把下列概念分成两类：距离、质量、电场、熵、美、旋律，我想，把熵放在美和旋律里，是有最强的理由的。熵只有当部分联合在一起时才能被发现，正如美和旋律只有当部分联合在一起时才能被看见和听到”。在经典系统和量子（quantum）系统中，熵无时无处不在。熵已经渗透到生物学、化学、经济学、哲学、社会学、认识论等众多学科。特别是，熵经 Kolmogorov 引入到遍历论，使得遍历论在自身的发展中树立新的里程碑。自此，熵不仅成功解决了一系列测度空间与保测变换的同构等价问题，而且熵的概念对系统的复杂性、混沌等一系列不同领域中的重要问题都是一个很好的刻划工具。

熵经 Shannon 之手与信息紧紧相连，为熵从热力学进入信息科学、生命科学等当代前沿科学铺平道路，并产生了生物信息学、神经信息学等新学科。熵在数据分析、通信理论、图像处理、模式识别、生命科学等方面获得广泛的应用并取得显著成果。

0.2　熵的定义及其特性[2,3]

假设 μ，ν 是可测空间（Ω,F）上的有限测度，对任意 $A_i \in F$，$A_j \in F$，$i \neq j$，$A_i \cap A_j = \varnothing$，$i,j=1,2,\cdots,n$，$\sum_{i=1}^{n} A_i = \Omega, A = \{A_i\}$ 是（Ω,F）的有限可测分划，或称为 Ω 的有限分划。对于 Ω 的有限分划 A，令

$$H_A(\mu,\nu) = \sum_{i=1}^{n} \mu(A_i) \ln \frac{\mu(A_i)}{\nu(A_i)} \tag{0.1}$$

定义 0.1　称对于可测空间（Ω,F）上的一切有限可测分划的 $H_A(\mu,\nu)$ 的上确界

$$H(\mu,\nu) = \sup_{A}\{H_A(\mu,\nu); A: \Omega \text{ 的有限分划}\} \tag{0.2}$$

为关于测度 ν 的测度 μ 的熵（也称为测度 μ 与 ν 之间的相对熵（relative entropy），或测度 μ 与 ν 的信息分歧（I-divergence），或测度 μ 与 ν 之间的 K-L 距离（Kullback-Leibler distance））。

定理 0.1

$$\begin{cases} H(\mu,\nu) \geqslant 0, \mu(\Omega) = \nu(\Omega) < \infty \\ H(\mu,\nu) = 0 \Leftrightarrow \mu = \nu \end{cases} \tag{0.3}$$

证明 使用不等式 $\ln x \leqslant x - 1$。对 Ω 的任意有限分划 $A = \{A_i\}$，有

$$\begin{aligned} H_A(\mu,\nu) &= -\sum_i \mu(A_i) \ln \frac{\nu(A_i)}{\mu(A_i)} \geqslant \sum_i \mu(A_i)\left[\frac{\nu(A_i)}{\mu(A_i)} - 1\right] \\ &= \sum_i \mu(A_i) - \sum_i \nu(A_i) = \mu(\Omega) - \nu(\Omega) = 0 \end{aligned} \tag{0.4}$$

由 $H(\mu,\nu)$ 的定义，$H(\mu,\nu) \geqslant 0$；若 $H(\mu,\nu) = 0$，则对 Ω 的一切有限分划 $A = \{A_i\}$ 都有 $H_A(\mu,\nu) = 0$，从式（0.4）可知，对一切 i，$\mu(A_i) = \nu(A_i)$，即 $\mu = \nu$。相反，$\mu = \nu$，则 $H(\mu,\nu) = 0$，证毕。

在 Ω 的两个有限分划 $A = \{A_i, i = 1,2,\cdots,m\}, B = \{B_j, j = 1,2,\cdots,n\}$ 上，对任意 $B_j \in B$，存在 $A_i \in A$，使 $A_i \supset B_j$，这时称 B 是 A 的加细。经适当替换和排列，假定 $B = \{A_{ij}, i = 1,2,\cdots,m, j = 1,2,\cdots,k_i\}$，则有 $\sum_{j=1}^{k_i} A_{ij} = A_i$。

引理 0.1 假设分划 $B = \{A_{ij}, i = 1,2,\cdots,m, j = 1,2,\cdots,k_i\}$ 是 Ω 的有限分划 $A = \{A_i, i = 1,2,\cdots,m\}$ 的加细，则

$$H_A(\mu,\nu) \leqslant H_B(\mu,\nu) \tag{0.5}$$

证明 令 $d_i = \dfrac{\mu(A_i)}{\nu(A_i)}, \alpha_{ij} = \dfrac{\mu(A_{ij})}{\nu(A_{ij})}, i = 1,2,\cdots,m, j = 1,2,\cdots,k_i$，由 $\sum_{i,j} \alpha_i \nu(A_{ij}) = \sum_{i,j} \alpha_{ij} \nu(A_{ij}) = \mu(\Omega) < \infty$ 以及对任意概率分布 $(q_1, q_2, \cdots, q_m)$：

$$H(q_1, q_2, \cdots, q_m) \leqslant -\sum_{i=1}^{m} p_i \ln q_i \tag{0.6}$$

有

$$\begin{aligned} &\sum_{i,j} \alpha_{ij} \nu(A_{ij}) \ln(\alpha_i \nu(A_{ij})) \\ \leqslant &\sum_{i,j} \alpha_{ij} \nu(A_{ij}) \ln(\alpha_{ij} \nu(A_{ij})) = \sum_{i,j} \mu(A_{ij}) \ln \mu(A_{ij}) \end{aligned}$$

然而

$$\begin{aligned} &\sum_{i,j} \alpha_{ij} \nu(A_{ij}) \ln(\alpha_i \nu(A_{ij})) \\ = &\sum_i \mu(A_i) \ln \frac{\mu(A_i)}{\nu(A_i)} + \sum_{i,j} \mu(A_{ij}) \ln \nu(A_{ij}) \end{aligned}$$

从以上两式得式（0.5），证毕。

利用引理 0.1，可得 $H(\mu,\nu)$ 的积分表示，这件事情与测度的绝对连续性有关。对于 (Ω,F) 上的两个测度 μ,ν，若 $\nu(A)=0$ 可得 $\mu(A)=0$，则称 μ 关于 ν 绝对连续，记为 $\mu\ll\nu$。若 ν 有限且 $\mu\ll\nu$，则存在可测函数 $\varphi(x)$，使得

$$\mu(A)=\int_A\varphi(x)\mathrm{d}\nu(x),\quad A\in F$$

$\varphi(x)$ 称为 Radon-Nikodym 导函数，表示为 $\varphi(x)=\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)$。有时也写成 $\mathrm{d}\mu(x)=\varphi(x)\mathrm{d}\nu(x)$。

定理 0.2　假定 μ,ν 是 (Ω,F) 上的有限测度：

(1) μ 关于 ν 不绝对连续，则

$$H(\mu,\nu)=\infty \tag{0.7}$$

(2) $\mu\ll\nu$，令 $\varphi(x)=\frac{\mathrm{d}\mu}{\mathrm{d}\nu}(x)$，则

$$H(\mu,\nu)=\int_\Omega\ln\varphi(x)\mathrm{d}\mu(x) \tag{0.8}$$

证明

(1) 由假定，存在 $A\in F$，使 $\nu(A)=0$ 且 $\mu(A)>0$。考虑分划 $\xi=\{A,A^c\}$，有

$$\begin{aligned}H(\mu,\nu)&\geqslant H_\xi(\mu,\nu)\\&=\mu(A)\ln\frac{\mu(A)}{\nu(A)}+\mu(A^c)\ln\frac{\mu(A^c)}{\nu(A^c)}=\infty\end{aligned}$$

式（0.7）成立。

(2) $A=\{A_i,i=1,2,\cdots,n\}$ 是任意有限分划，定义可测函数

$$\psi(x)=\frac{\mu(A_i)}{\nu(A_i)},\quad x\in A_i,i=1,2,\cdots,n$$

则 $\int_\Omega\varphi(x)\mathrm{d}\nu(x)=\int_\Omega\psi(x)\mathrm{d}\nu(x)=\mu(\Omega)$，因此有

$$\int_\Omega\varphi(x)\ln\varphi(x)\mathrm{d}\nu(x)\leqslant\int_\Omega\varphi(x)\ln\psi(x)\mathrm{d}\nu(x)=\int_\Omega\ln\varphi(x)\mathrm{d}\mu(x)$$

然而，根据 $\psi(x)$ 的定义，有

$$\int_\Omega\varphi(x)\ln\psi(x)\mathrm{d}\nu(x)=\int_\Omega\ln\psi(x)\mathrm{d}(x)=H_A(\mu,\nu)$$

因此，$H_A(\mu,\nu)\leqslant\int_\Omega\ln\varphi(x)\mathrm{d}\mu(x)$。因为 A 是任意分划，所以不等式

$$H(\mu,\nu)\leqslant\int_\Omega\ln\varphi(x)\mathrm{d}\mu(x) \tag{0.9}$$

成立。

下面给出式（0.9）的逆向不等式。令 $C_m=\{x\in\Omega;|\ln\phi(x)|>m\},m\to\infty\Rightarrow\mu(C_m)\to 0$。因此，对任意正数 ε，存在 M，若 $m\geqslant M$，有

$$\mu(C_m)\ln\frac{\mu(C_m)}{\nu(C_m)}\geqslant\mu(C_m)\ln\frac{\mu(C_m)}{\nu(\Omega)}\geqslant-\varepsilon \tag{0.10}$$

固定 $m(\geqslant M)$，给出有限分划 $A=\{A_i,i=0,\pm1,\pm2,\cdots,\pm n\}$

$$A_i=\left\{x\in\Omega;\frac{(i-1)m}{n}<\ln\varphi(x)\leqslant\frac{im}{n}\right\},\quad -n+2\leqslant i\leqslant n$$

$$A_{-n+1}=\left\{x\in\Omega;-m\leqslant\ln\varphi(x)\leqslant\frac{(-n+1)m}{n}\right\},\quad A_{-n}=C_m$$

当 $x\in A_i,-n<i\leqslant n$，容易看出

$$\ln\varphi(x)\leqslant\frac{im}{n}\leqslant\ln\frac{\mu(A_i)}{\nu(A_i)}+\frac{m}{n} \tag{0.11}$$

从而，由式（0.10）和式（0.11）可得

$$\int_{C_m}\ln\varphi(x)\mathrm{d}\mu(x)\leqslant\sum_{i=-n+1}^{n}\mu(A_i)\ln\frac{\mu(A_i)}{\nu(A_i)}+\frac{m}{n}\mu(\Omega)$$

$$\leqslant H_A(\mu,\nu)+\varepsilon+\frac{m}{n}\mu(\Omega)\leqslant H(\mu,\nu)+\varepsilon+\frac{m}{n}\mu(\Omega)$$

先令 $n\to\infty$，后令 $\varepsilon\to 0$，得到

$$\int_{\Omega}\ln\varphi(x)\mathrm{d}\mu(x)\leqslant H(\mu,\nu)$$

证毕。

通过 $H(\mu,\nu)$ 的积分表示，使得在无限测度的情况下也能定义 $H(\mu,\nu)$ 。

定义 0.2　假定 μ,ν 是 (Ω,F) 上的测度，ν 有限。若 μ 关于 ν 不绝对连续，μ 关于 ν 的熵 $H(\mu,\nu)$ 是无穷大；若 μ 关于 ν 绝对连续，则 $H(\mu,\nu)$ 用式（0.8）定义。

下面的例子是令人感兴趣的。

例 0.1　假定 μ,ν 是 $\mathbf{R}^d$ 上的测度，且关于 Lebesgue 测度 dx 绝对连续，它们的 Radon-Nikodym 导函数分别为 $p(x),q(x)$。若 $\mu\ll\nu$，则 $\frac{\mathrm{d}\mu(x)}{\mathrm{d}\nu(x)}=\frac{p(x)}{q(x)}$，由定义 0.2，有

$$H(\mu,\nu)=\int_{\mathbf{R}^d}\ln\frac{\mathrm{d}\mu(x)}{\mathrm{d}\nu(x)}\mathrm{d}\mu(x)=\int_{\mathbf{R}^d}p(x)\ln\frac{p(x)}{q(x)}\mathrm{d}x$$

特别地，若 μ 是概率测度，作为标准化的测度 ν_0 选择 Lebesque 测度（这时 $q(x)\equiv 1$），则有

$$H(\mu,\nu)=\int_{\mathbf{R}^d}p(x)\ln p(x)\mathrm{d}x=-h(p)$$

可见，连续随机变量的熵是将 Lebesque 测度标准化的熵取负号。

一般的提法是，设 X，Y 是概率空间 (Ω,F,P) 上两个绝对连续的随机变量，密度函数分别是 $p(x),q(x)$，则 X 关于 Y 的熵

$$H(X,Y)=\int_{\mathbf{R}^d}[p(x)\ln p(x)-p(x)\ln q(x)]\mathrm{d}x=\int_{\mathbf{R}^d}p(x)\ln\frac{p(x)}{q(x)}\mathrm{d}x \tag{0.12}$$

在双方都有限情况下，显然有

$$H(\mu,\nu)=H(X,Y) \tag{0.13}$$

定义 0.2 的熵也可由随机变量定义。为此，先引入一些记号。

设 $X(\omega),Y(\omega)$ 是定义在概率空间 (Ω,F,P) 上的非负随机变量，使得

$$\mu(A)=\int_A X(\omega)\mathrm{d}p(\omega),\quad \nu(A)=\int_A Y(\omega)\mathrm{d}p(\omega),\quad A\in F \tag{0.14}$$

是概率测度。由 Radon-Nikodym 微分定理：

$$\mathrm{d}\mu=X\mathrm{d}p,\quad \mathrm{d}\nu=Y\mathrm{d}p \tag{0.15}$$

$X(\omega),Y(\omega)$ 可能被认为是广义概率密度。假定 μ 关于 ν 绝对连续，即 $\mu\ll\nu$。

有了以上这些记号，定义 0.2 的熵用随机变量定义为

$$\widetilde{H}(X,Y)=\int X(\omega)\ln\frac{X(\omega)}{Y(\omega)}\mathrm{d}p(\omega) \tag{0.16}$$

可以证明

$$\widetilde{H}(X,Y)=H(\mu,\nu) \tag{0.17}$$

定理 0.3　假定概率测度满足

$$\lim_{n\to\infty}u_n=\mu,\quad \lim_{n\to\infty}\nu_n=\nu$$

则有

$$\liminf_{n\to\infty}H(\mu_n,\nu_n)\geqslant H(\mu,\nu) \tag{0.18}$$

即函数 $H(\mu,\nu)$ 是下半连续的。

证明　从函数 $x\ln x(x\geqslant 0)$ 的凸性，对于 $\nu(A_i)\neq 0$ 的任意 $A_i\in F$ 和可测函数 $f\geqslant 0$，有

$$\left(\frac{1}{\nu(A_i)}\int_{A_i}f\mathrm{d}\nu\right)\ln\left(\frac{1}{\nu(A_i)}\int_{A_i}f\mathrm{d}\nu\right)\leqslant\frac{1}{\nu(A_i)}\int_{A_i}f\ln f\mathrm{d}\nu$$

令 $f=\dfrac{\mathrm{d}\nu}{\mathrm{d}\mu}$，则

$$\nu(A_i)\ln\frac{\mu(A_i)}{\nu(A_i)}\leqslant\int_{A_i}\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\ln\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\mathrm{d}\nu$$

进而有

$$\sum_i\nu(A_i)\ln\frac{\mu(A_i)}{\nu(A_i)}\leqslant\int\left(\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\ln\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\right)\mathrm{d}\nu=\int X(\omega)\ln\frac{X(\omega)}{Y(\omega)}\mathrm{d}p(\omega) \tag{0.19}$$