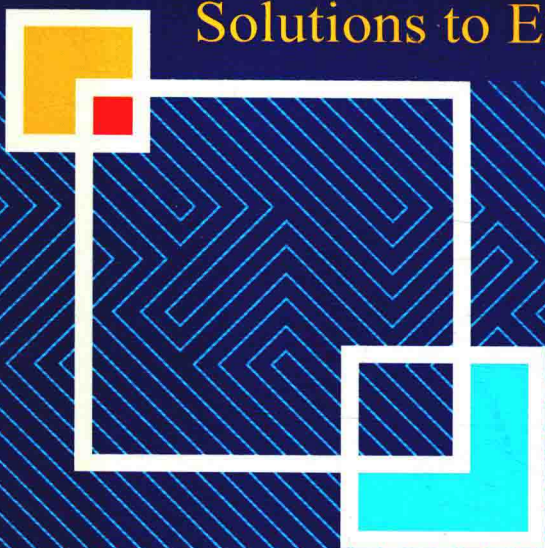


普通高等学校电类规划教材

信息论基础 习题解答 (第2版)

许文俊 田宝玉 杨洁 贺志强 编著

Elements of Information Theory:
Solutions to Exercises (2nd Edition)



 中国工信出版集团

 人民邮电出版社
POSTS & TELECOM PRESS


名师名校

普通高等学校电类规划教材

信息论基础 习题解答 (第2版)

许文俊 田宝玉 杨洁 贺志强 编著

Elements of Information Theory:
Solutions to Exercises (2nd Edition)

人民邮电出版社

北京



名师名校

图书在版编目 (CIP) 数据

信息论基础习题解答 / 许文俊等编著. — 2版. —
北京: 人民邮电出版社, 2017.8
普通高等学校电类规划教材
ISBN 978-7-115-46523-8

I. ①信… II. ①许… III. ①信息论—高等学校—题
解 IV. ①G201-44

中国版本图书馆CIP数据核字(2017)第180359号

内 容 提 要

《信息论基础习题解答(第2版)》是本科生教材《信息论基础(第2版)》配套的教学辅助资料,主要目的是为学生提供更多的信息论基本问题解题示范,开阔学生的解题思路,提高学生解决与信息论有关的基础性或综合性问题的能力,进一步提高信息论课程的教学质量。本书很多习题来自传统或经典的国内外教科书,同时还包含相当数量的通过一线教师多年的教学、科研实践提炼并得到验证的典型习题。

本书与主教材结构相同,也对应为11章。内容包括“知识要点”“例题精解”“习题解答”“补充题解”(部分章节)4个部分。另外,有些章节安排了编写计算机程序解答的实验型习题,以加强信息论课程的实践性。通过类似的训练,学生可提高利用计算机解决用解析方法难以解决或需要大量计算的问题的能力。

本书可作为电子信息类专业的本科生与研究生的学习参考用书,也可作为信息论教师的教学参考资料,还适合于信息技术领域需要自学信息论的读者。

◆ 编 著 许文俊 田宝玉 杨 洁 贺志强

责任编辑 李 召

责任印制 陈 犇

◆ 人民邮电出版社出版发行 北京市丰台区成寿寺路11号

邮编 100164 电子邮件 315@ptpress.com.cn

网址 <http://www.ptpress.com.cn>

大厂聚鑫印刷有限责任公司印刷

◆ 开本: 787×1092 1/16

印张: 18.25

2017年8月第2版

字数: 448千字

2017年8月河北第1次印刷

定价: 49.80元

读者服务热线: (010)81055256 印装质量热线: (010)81055316

反盗版热线: (010)81055315

广告经营许可证: 京东工商广登字 20170147号

21世纪被称为信息化社会,信息已经融入人类生活、生产的方方面面,我国诸多国家重大发展战略,如“互联网+”“大众创业、万众创新”“中国制造2025”“一带一路”等,均与信息领域息息相关。为了创新与普及信息理论与信息技术的应用,社会需要大力培养具备信息处理全局观和扎实理论基础的信息技术高级人才,能够在信息处理工程实践中自觉运用甚至思考创新信息理论知识,“知其然”且“知其所以然”,从而以系统观、全局观最优化信息系统设计。因此,需要一批协助学生掌握、理解、自觉运用信息理论到信息处理工程实际中的教材及专著,助推我国信息技术从业人员整体理论水平及工程能力的提升。

由香农(C. E. Shannon)创立的信息论基础是信息处理不同环节,如信息度量、压缩、传输、存储等的理论基石,因此它是通信与电子信息类专业的重要基础课,也是绝大多数电子信息类专业高年级本科生和研究生的必修课程。信息论基础课程以信息熵、平均互信息为基本概念,以香农三个编码定理——无失真信源编码定理、有噪信道编码定理和限失真信源编码定理为核心内容,研究信息系统中信息的度量、信源的压缩以及信息通过信道有效和可靠传输等理论问题。

为使学生更好地理解基本原理、掌握基本方法、提高分析能力,我们经过多年的教学和科研实践的积累,编写了大学工科本科生教材《信息论基础(第2版)》(普通高等教育“十一五”国家级规划教材,田宝玉、杨洁、贺志强、许文俊编著,人民邮电出版社,2016年1月出版)。国内外信息论课程教学实践证明:一本优秀的教科书,必须配备精心挑选的习题,而这些习题都应该配有详细的解答,这对学生深入理解基本概念,掌握和运用基本的解题方法,提高解决问题的能力非常必要,对从事本课程教学的教师也有重要的参考价值。

《信息论基础习题解答(第2版)》(以下简称《题解》)就是为上述本科生教材《信息论基础(第2版)》配套的辅助教学资料,主要目的是为学生提供更多的信息论基本问题解题示范,开阔学生的解题思路,提高学生解决与信息论有关的基础性或综合性问题的能力,进一步提高信息论课程的教学质量。

当前,市场上已有一些信息论习题解答的图书问世,但随着信息理论教学与研究的广泛开展,还需要更多具有自身特点的习题解答图书出现,相互借鉴,取长补短,以使初学者能够了解多种风格的问题描述、分析方式及解题技巧,有利于学生更全面地掌握基本解题方法,扩大眼界,开拓思路。

本书是在参考了大量的国内外优秀教科书和专著的基础上编写而成的,很多习题来自传

统或经典的国内外教科书,同时本书还包含较多通过我们多年的教学实践提炼并得到验证的典型习题。

本书与主教材结构相同,也对应为11章。内容包括下述四部分。第一部分是“知识要点”,扼要总结了本章的基本概念、重要的定理与技术方法。第二部分是“例题精解”,对本章所涉及的重要或典型问题进行细致地分析与求解。为使学生灵活运用所学的知识 and 充分拓展学生的思维,对同一个问题尽可能提供多种不同的解决方案。第三部分是“习题解答”,与教材每章的习题严格对应,解题步骤与第二部分相比略有简化,但确保学生容易理解、易于掌握。第四部分是“补充题解”,为了扩展学生对课本知识掌握的广度与深度,部分章节有针对性地增加了一些习题及解答。为保证较大的信息量,本书不包含原教材中的例题(虽然也很重要)。本书在某些章节还安排了编写计算机程序解题的实验型习题,以加强信息论课程的实践性。读者通过类似的训练,可提高利用计算机解决用解析方法难以解决或需要大量计算的问题的能力。

在第1版的基础上,本书第2章到第10章都新增了习题及对应解答,重点增加了经典信源编码、信道编码方法相关的习题及解答,以及率失真函数、多用户信息论相关的习题及解答。前者更加契合当前通信与电子信息类、电子信息类本科专业信息论基础与编码同一课程授课的趋势,后者更能满足相应专业的研究生深化信息论基础学习的需求。

为使读者不陷入枯燥和抽象理论问题的迷宫,真切体会信息论作为基础理论指导实际问题的巨大潜力,本书还尽量选择贴近实际而且有趣的问题,如天气分析预测与信息度量的问题、最佳决策与信息编码的问题、鉴别假币与信息可加性的问题、现代蜂窝通信与信息传输容量的问题等,以便提高学生学习本课程的兴趣。实践证明:上述内容对不是专门从事信息论研究的人员也有很好的参考价值。

本书由许文俊、田宝玉、杨洁、贺志强共同编写。

在本书写作过程中,作者得到了北京邮电大学信息与通信工程学院的支持与资助,得到了北京邮电大学信息论课程组广大教师以及新科研楼619室学生的支持,特别是黄春雷博士以及唐霁霁、韩晓、段皓、王译铎、刘宇琦5位硕士的仔细校对,在此对帮助过我们的单位、老师、学生等表示最诚挚的感谢。编者在本书的编写过程中参考了大量的国内外著作,在此向有关同行表示深切谢意。

编者

2017年7月于北京邮电大学

目 录

第 1 章 绪论	1	5.1.2 定长码	93
第 2 章 离散信息的度量	4	5.1.3 变长码	94
2.1 知识要点	4	5.1.4 最优编码	95
2.1.1 自信息和互信息	4	5.1.5 几种实用的信源编码方法	96
2.1.2 信息熵	6	5.2 例题精解	97
2.1.3 平均互信息	9	5.3 习题解答	103
2.2 例题精解	11	5.4 补充题解	119
2.3 习题解答	15	第 6 章 离散信道及其容量	124
第 3 章 离散信源	34	6.1 知识要点	124
3.1 知识要点	34	6.1.1 基本概念	124
3.1.1 离散信源的分类与数学模型	34	6.1.2 单符号离散信道及其容量	125
3.1.2 离散无记忆信源的扩展	35	6.1.3 级联信道及其容量	126
3.1.3 离散平稳信源的熵	35	6.1.4 多维矢量信道及其容量	127
3.1.4 有限状态马尔可夫链	36	6.1.5 信道容量的迭代算法	129
3.1.5 马尔可夫信源	37	6.1.6 有约束信道的容量	129
3.1.6 信源的相关性与剩余度	38	6.2 例题精解	130
3.2 例题精解	38	6.3 习题解答	136
3.3 习题解答	42	第 7 章 有噪信道编码	155
第 4 章 连续信息与连续信源	64	7.1 知识要点	155
4.1 知识要点	64	7.1.1 基本概念	155
4.1.1 连续随机变量的熵	64	7.1.2 最佳判决与译码准则	156
4.1.2 离散时间高斯随机变量的熵	65	7.1.3 信道编码与最佳译码	157
4.1.3 连续最大熵定理	66	7.1.4 费诺不等式	157
4.1.4 连续随机变量之间的平均互 信息	67	7.1.5 有噪信道编码定理	158
4.1.5 离散集与连续随机变量之间的 互信息	67	7.1.6 纠错编码技术简介	159
4.1.6 几种重要的连续信源	68	7.2 例题精解	160
4.2 例题精解	68	7.3 习题解答	166
4.3 习题解答	74	7.4 补充题解	181
4.4 补充题解	89	第 8 章 波形信道	182
第 5 章 无失真信源编码	92	8.1 知识要点	182
5.1 知识要点	92	8.1.1 离散时间连续信道	182
5.1.1 概述	92	8.1.2 加性噪声信道与容量	183
		8.1.3 AWGN 信道的容量	184
		8.1.4 有色高斯噪声信道	185

8.1.5 数字调制系统的信道容量	186	10.1 知识要点	235
8.2 例题精解	186	10.1.1 基本概念	235
8.3 习题解答	190	10.1.2 多址接入信道	236
第9章 信息率失真函数	204	10.1.3 广播信道	237
9.1 知识要点	204	10.1.4 中继信道	237
9.1.1 基本概念	204	10.1.5 分布信源编码	237
9.1.2 信息率失真函数定义	205	10.2 习题解答	238
9.1.3 限失真信源编码定理	205	第11章 信息理论方法及其应用	262
9.1.4 离散 $R(D)$ 函数的性质与 计算	205	11.1 知识要点	262
9.1.5 连续 $R(D)$ 函数的性质与 计算	206	11.1.1 离散信源熵的估计	262
9.1.6 高斯信源的 $R(D)$ 函数	207	11.1.2 最大熵原理	262
9.1.7 一般连续信源的 $R(D)$ 函数	207	11.1.3 最小交叉熵原理	264
9.1.8 有损数据压缩技术简介	207	11.1.4 信息理论方法的应用	265
9.2 例题精解	207	11.2 例题精解	265
9.3 习题解答	213	11.3 习题解答	267
9.4 补充题解	232	期末考试模拟试题	277
第10章 网络信息论初步	235	期末考试模拟试题标准答案	279
		参考文献	285

1. 信息的基本概念

(1) 组成客观世界的三大基本要素：物质、能量和信息。

(2) 通信的基本问题：在一点精确地或近似地恢复另一点所选择的消息。

(3) 通信系统的3项主要性能指标：传输的有效性、传输的可靠性、传输的安全性。

(4) 3项性能指标所对应的3项基本技术：数据压缩、数据纠错和数据加密。

(5) 信息有许多与物质、能量相同的特征（信息可以产生、消失、携带、处理和量度），信息也有与物质、能量不同的特征（信息可以共享，可以无限制地复制等）。

(6) 一种比较普遍的对信息的描述：信息是认识主体（人、生物、机器）所感受的和所表达的事物运动的状态和运动状态变化的方式。

(7) 信息的3个基本层次。

- 语法 (Syntactic) 信息：反映事物运动状态及其变化方式的外在形式。
- 语义 (Semantic) 信息：反映事物运动状态及其变化方式的内在含义。
- 语用 (Pragmatic) 信息：反映事物运动状态及其变化方式的效用价值。

语义信息以语法信息为基础，语用信息以语义信息和语法信息为基础。三者之间，语法信息是最简单、最基本的层次，语用信息则是最复杂、最实用的层次。

(8) 通信信息的3个层次：信号、消息与信息。其中信号为最低的层次，信息是最高的层次。消息是信息的携带者，信息包含于消息之中；信号是消息的载体，消息是信号的具体内容。

2. 通信系统模型

(1) 模型的主要组成部分：信源、编码器、信道（包括噪声）、译码器、信宿。

● 信源：信源就是信息的来源；按符号的取值分为离散信源和连续信源；按输出符号之间的依赖关系分类，分为无记忆信源和有记忆信源；信源研究的核心问题是信源消息所包含信息的量度。

● 编码器：编码器的功能是将消息变成适合于信道传输的信号，包括：信源编码器、信道编码器、调制器。信源编码器的功能是将信源消息变成符号，目的是提高传输有效性，也就是压缩每个信源符号传输所需代码（通常为二进制代码）的数目（对二进制代码称比

特数)。信道编码器给信源编码符号增加冗余符号,目的是提高传输可靠性。调制器功能是将编码器的输出符号变成适合信道传输的信号,目的是提高传输效率(使远距离传输成为可能)。

- 信道:信道是信号从编码器传输到译码器的中间媒介,分为无噪声信道和有噪声信道。噪声主要有两种:加性噪声和乘性噪声。这里主要研究加性噪声。其中最普遍的是理想加性高斯白噪声(AWGN)信道。信道还可分为离散信道、离散时间连续信道和波形信道(或模拟信道)。其中,离散信道和离散时间连续信道输入与输出都是符号序列,只不过符号取值不同,前者取离散值,而后者取连续值。而波形信道的输入与输出均为时间的连续波形。信道有无记忆信道和有记忆信道的区分,离散信道和离散时间连续信道可以是无记忆的,也可以是有记忆的。而波形信道通常是有记忆的。

- 译码器:译码器实现与编码器相反的功能,即从信号中恢复消息。译码器包括解调器、信道译码器、信源译码器。解调器功能是将信道输出信号恢复成符号。信道译码器的功能是去掉解调器输出符号中的冗余符号,并对非冗余符号进行检错或纠错。信源译码器的功能是将信道译码器输出符号变成消息。

- 信宿:信宿的功能是接收信息,包括人或设备。

(2) 通信系统性能指标的评价。

- 有效性用频谱复用程度(模拟系统)或频谱利用率(数字系统)来衡量。提高有效性的措施是,采用性能好的信源编码以压缩码率,采用频谱利用率高的调制减小传输带宽。

- 可靠性用输出信噪比(模拟系统)和传输错误率(数字系统)来衡量。对于数字系统,提高可靠性的措施是,采用高性能的信道编码以降低错误率。

- 安全性用信息加密强度来衡量。提高安全性的措施是,采用强度高的密码与信息隐藏或伪装技术。

3. 信息论研究的内容

信息论研究的基本问题是信源、信道及编码问题,核心是3个编码定理。

(1) 关于信源信息的度量。信息熵作为信源所含信息的量度,是信息论中最重要的概念。

(2) 关于无失真信源编码。无失真信源编码定理(香农第一定理)是信源无损压缩的理论基础,其内容是:如果信源编码码率(编码后传送信源符号所需比特数)不小于信源的熵,就存在无失真编码;反之,不存在无失真编码。可以简述为

$$R \geq H \Leftrightarrow \text{存在无失真信源编码}$$

其中, R 为信源编码码率, H 为信源的熵。

(3) 关于信道容量与信息的可靠传输。有噪信道编码定理(香农第二定理)是信道编码的理论基础,其内容是:如果信息传输速率小于信道容量,则总可找到一种编码方式使得当编码序列足够长时传输差错任意小;反之不存在使传输差错任意小的编码。可以简述为

$$R \leq C \Leftrightarrow \text{存在译码差错任意小的信道编码}$$

其中, R 为信息传输速率,也称信道编码码率, C 为信道容量。

(4) 信息率失真理论。限失真信源编码定理(香农第三定理)是信源有损压缩的理论基础,其内容是:对任何失真测度 $D \geq 0$,只要码字足够长,总可找到一种编码,使得当信源编码的码率 $\geq R(D)$ 时,码的平均失真 $\leq D$;反之,如果信源编码的码率 $< R(D)$,就不存在

使平均失真 $\leq D$ 的编码。可以简述为

$$R \geq R(D) \Leftrightarrow \text{存在平均失真} \leq D \text{ 信源编码}$$

其中, R 为信源编码码率, $R(D)$ 称为信息率失真函数, 是满足失真准则 (平均失真 $\leq D$) 下, 每信源符号所需最小编码比特数。

(5) 信息论的特点。

- 以概率论、随机过程为基本研究工具。
- 研究通信系统的整个过程, 而不是单个环节, 并以编码器、译码器为重点。
- 关心最优系统的性能和怎样达到这个性能, 并不具体设计系统。
- 研究语法信息中的概率信息, 要求信源为随机过程。

2.1 知识要点

符号规定：大写字母表示集合，小写字母表示集合中的事件。

2.1.1 自信息和互信息

1. 自信息

事件集合 X 中的事件 $x = a_i$ 的自信息定义为

$$I_X(a_i) = -\log P_X(a_i) \quad (2.1)$$

简记为

$$I(x) = -\log p(x) \quad (2.2)$$

注意：

(1) $a_i \in A$ ，且 $\sum_{i=1}^n P_X(a_i) = 1$ ， $0 \leq P_X(a_i) \leq 1$ ，简记为 $0 \leq p(x) \leq 1$ 。

(2) 要求自信息 I 为非负值，所以对数的底必须大于 1。单位换算如下：

$$1 \text{ Nat} = \log_e e \text{ Nat} = \log_2 e \text{ bit} = 1.443 \text{ bit}$$

$$1 \text{ Dit} = \log_{10} 10 \text{ Dit} = \log_2 10 \text{ bit} = 1 / \log_{10} 2 \text{ bit} = 3.32 \text{ bit}$$

(3) 自信息为随机变量，且 $I(x)$ 是 $p(x)$ 的单调递减函数。

(4) 自信息含义包含如下两个方面：

① 表示事件发生前，事件发生的不确定性；

② 表示事件发生后，事件所包含的信息量，是提供给信宿的信息量，也是解除这种不确定性所需要的信息量。

2. 联合自信息

联合事件集合 XY 中的事件 $x = a_i$ ， $y = b_j$ 的联合自信息定义为

$$I_{XY}(a_i, b_j) = -\log P_{XY}(a_i, b_j) \quad (2.3)$$

简记为

$$I(xy) = -\log p(xy) \quad (2.4)$$

其中, $p(xy)$ 要满足非负和归一化条件。实际上, 如果把联合事件 xy 看成一个单一事件, 那么联合自信息的含义与自信息的含义相同。

3. 条件自信息

事件 $x = a_i$ 在事件 $y = b_j$ 给定条件下的条件自信息定义为

$$I_{x|Y}(a_i | b_j) = -\log P_{x|Y}(a_i | b_j) \quad (2.5)$$

简记为

$$I(x|y) = -\log p(x|y) \quad (2.6)$$

其中, 条件概率 $p(x|y)$ 也要满足非负和归一化条件。

4. 自信息、条件自信息和联合自信息之间的关系

$$I(xy) = I(x) + I(y|x) = I(y) + I(x|y)$$

5. 互信息

设联合集 XY , 离散随机事件 $x = a_i$ 和 $y = b_j$ 之间的互信息 ($x \in X, y \in Y$) 定义为

$$I_{x;y}(a_i; b_j) = \log \frac{P_{XY}(a_i | b_j)}{P_X(a_i)} \quad (2.7)$$

简记为

$$I(x; y) = \log \frac{p(x|y)}{p(x)} \quad (2.8)$$

通过计算可得

$$I(x; y) = I(x) - I(x|y) \quad (2.9)$$

注意:

(1) 互信息反映了两个不同的随机变量集合中两个随机事件之间的统计关联程度; $I(x; y)$ 表示当 y 发生后 x 不确定性的变化, 表示由 y 得到的关于 x 的信息量;

(2) 互信息具有对称性;

(3) 互信息可正可负;

(4) 互信息是随机变量;

(5) 互信息的单位与自信息单位相同。

6. 条件互信息

设联合集 XYZ , 在给定 $z \in Z$ 条件下, $x(\in X)$ 与 $y(\in Y)$ 之间的条件互信息定义为

$$I(x; y|z) = \log \frac{p(x|yz)}{p(x|z)} \quad (2.10)$$

除具有条件外, 条件互信息的含义与互信息的含义与性质都相同。

2.1.2 信息熵

1. 信息熵

- 离散随机变量集合 X 的熵定义为自信息的平均值, 记为 $H(X)$

$$H(X) = E_{p(x)} [I(x)] = -\sum_x p(x) \log p(x) \quad (2.11)$$

简记为

$$H(X) = H(p_1, p_2, \dots, p_n) \quad (2.12)$$

注意:

- (1) $I(x)$ 为事件 x 的自信息, $E_{p(x)}$ 表示对随机变量用 $p(x)$ 取平均运算; 熵的单位为比特 (奈特) / 信源符号;

- (2) $\sum_{i=1}^n p_i = 1$, $0 \leq p_i \leq 1$, 特别是当 $n=2$ 时, $H(X) = H(p, 1-p) = H(p)$ 。

- 信息熵是从平均意义上表征信源总体特性的一个量, 其含义体现在如下几方面:

- (1) 在信源输出前, 表示信源的平均不确定性;
- (2) 在信源输出后, 表示每个信源符号所提供的平均信息量;
- (3) 表示信源随机性大小, $H(X)$ 大的信源, 随机性大;
- (4) 当信源输出后, 不确定性就解除, 熵可看成解除信源不确定性平均所需的信息量。

2. 联合熵

联合集 XY 上, 联合自信息 $I(xy)$ 的平均值称为联合熵, 表示为

$$H(XY) = E_{p(xy)} [I(xy)] = -\sum_x \sum_y p(xy) \log p(xy) \quad (2.13)$$

3. 条件熵

联合集 XY 上, 条件自信息 $I(y|x)$ 的平均值定义为条件熵, 表示为

$$\begin{aligned} H(Y|X) &= E_{p(xy)} [I(y|x)] \\ &= -\sum_x \sum_y p(xy) \log p(y|x) \\ &= \sum_x p(x) [-\sum_y p(y|x) \log p(y|x)] \\ &= \sum_x p(x) H(Y|x) \end{aligned} \quad (2.14)$$

其中, $H(Y|x) = -\sum_y p(y|x) \log p(y|x)$ 为在 x 取某一特定值时 Y 的熵。

4. 相对熵

- 散度的定义

若 P 和 Q 为定义在同一概率空间的两个概率测度, 定义 P 相对于 Q 的散度为

$$D(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \quad (2.15)$$

散度又称作相对熵、鉴别信息、方向散度、交叉熵、Kullback-Leibler 距离等。

- 散度的性质

如果在同一个有限字母表的概率空间上给定的两个概率测度为 $P(x)$ 和 $Q(x)$, 那么

$$D(P\|Q) \geq 0 \quad (2.16)$$

当且仅当对所有 x , $P(x) = Q(x)$ 时, 等式成立。

5. 各类熵之间的关系

- 熵与相对熵的关系

$$E_{p(x)} \log[1/Q(x)] \geq H(X)$$

同一概率空间的两随机变量集合, 如果一种分布的自信息用另一种分布做平均, 其值不小于另一种分布的熵。

- 条件熵与信息熵的关系

$$H(Y|X) \leq H(Y) \quad (2.17)$$

仅当 X 与 Y 相互独立时, 等式成立。这就是熵的不增原理: 在信息处理过程中, 条件越多, 熵越小。

6. 凸函数

- 上凸 (cap) 函数

多元函数 $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ 称为定义域上的上凸 (cap) 函数, 对于 $\alpha (0 \leq \alpha \leq 1)$ 及任意两矢量 $\mathbf{x}_1, \mathbf{x}_2$, 有

$$f[\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \geq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \quad (2.18)$$

成立。当且仅当 $\mathbf{x}_1 = \mathbf{x}_2$, 或 $\alpha = 0$ 或 1 时等式成立, 则称严格上凸函数。

- 下凸 (cup) 函数

多元函数 $f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$ 称为定义域上的下凸 (cup) 函数, 对于 $\alpha (0 \leq \alpha \leq 1)$ 及任意两矢量 $\mathbf{x}_1, \mathbf{x}_2$, 有

$$f[\alpha \mathbf{x}_1 + (1-\alpha)\mathbf{x}_2] \leq \alpha f(\mathbf{x}_1) + (1-\alpha)f(\mathbf{x}_2) \quad (2.19)$$

成立。当且仅当 $\mathbf{x}_1 = \mathbf{x}_2$, 或 $\alpha = 0$ 或 1 时等式成立, 则称严格下凸函数。

- Jenson 不等式

若 $f(\mathbf{x})$ 是定义在区间上的实值连续严格上凸函数, 则对于任意一组 $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_q$ 和任意一组 $\lambda_1, \lambda_2, \dots, \lambda_q$, $\sum \lambda_k = 1$, 有

$$f\left[\sum_{k=1}^q \lambda_k x_k\right] \geq \sum_{k=1}^q \lambda_k f(x_k) \quad (2.20)$$

当且仅当 $x_1 = x_2 = \dots = x_q$ 或 $\lambda_k = 1 (1 \leq k \leq q)$ 且 $\lambda_j = 0 (j \neq k)$ 时, 等式成立。特别是

$$E[\log x] \leq \log[E(x)] \quad (2.21)$$

- 重要不等式

$$1 - \frac{1}{x} \leq \ln x \leq x - 1 \quad (2.22)$$

仅当 $x = 1$ 时等号成立。

7. 熵的基本性质

- 对称性

概率矢量 $\mathbf{p} = (p_1, p_2, \dots, p_n)$ 中, 各分量的次序任意改变, 熵不变, 即

$$H(p_1, p_2, \dots, p_n) = H(p_2, p_3, \dots, p_n, p_1) = \dots = H(p_n, p_1, \dots, p_{n-1})$$

该性质说明熵仅与信源的总体特性有关, 而与随机变量的取值无关。

- 非负性

$$H(\mathbf{p}) = H(p_1, p_2, \dots, p_n) \geq 0 \quad (2.23)$$

仅当对某个 $p_i = 1$, 等式成立。

注意: 非负性仅对离散信源的熵有效, 而连续信源的熵不满足非负性。

- 确定性

$$H(1, 0) = H(1, 0, 0) = \dots = H(1, 0, \dots, 0) = 0$$

即当随机变量集合中任一事件概率为1时, 熵就为0。

- 扩展性

$$\lim_{\varepsilon \rightarrow 0} H_{n+1}(p_1, p_2, \dots, p_n - \varepsilon, \varepsilon) = H_n(p_1, p_2, \dots, p_n) \quad (2.24)$$

该式的含义就是, 小概率事件对熵的影响很小, 可以忽略。

- 可加性

设两个随机变量集合 X 、 Y 与它们的联合集 XY 的熵分别为 $H(X)$ 、 $H(Y)$ 、 $H(XY)$, 则

$$H(XY) = H(X) + H(Y|X) \quad (2.25)$$

熵的可加性可以推广到多随机变量集合的情况。设 n 维随机变量集 $X_1 X_2 \dots X_n$, 则有

$$\begin{aligned} H(X_1 X_2 \dots X_n) &= H(X_1) + H(X_2 | X_1) + \dots + H(X_n | X_1 \dots X_{n-1}) \\ &= \sum_{i=1}^n H(X_i | X_1 \dots X_{i-1}) \leq \sum_{i=1}^n H(X_i) \end{aligned} \quad (2.26)$$

上式称作熵的链原则。熵的可加性可以这样来理解:

- (1) 复合事件集合的不确定性为组成该复合事件的各简单事件集合的不确定性之和;
- (2) 对事件输出直接测量所得信息量等于分成若干步测量所得信息量的和;
- (3) 事件集合的平均不确定性可以分步解除, 各步解除不确定性的和等于信息熵。

- 极值性

对于有限离散随机变量集合，当集合中的事件等概率发生时，熵达到最大值。

注意：极值性不适用于无限离散信源。

- 上凸性

$H(\mathbf{p}) = H(p_1, p_2, \dots, p_n)$ 是 (p_1, p_2, \dots, p_n) 的严格上凸函数。

- 一一对应变换下的不变性

离散随机变量的变换包含两种含义，一是符号集中符号到符号的映射，二是符号序列到序列的变换。

(1) 离散随机变量（或矢量）经符号映射后的熵不大于原来的熵，仅当一一对应映射时熵不变。

(2) 离散随机序列经一一对应变换后，序列的熵不变，但单符号熵可能改变。

8. 熵函数的唯一性

如果要求熵函数满足以下条件：

- ① 是概率的连续函数；
- ② 当各事件等概率时是 n （信源符号数）的增函数；
- ③ 可加性。

那么，熵函数的表示是唯一的。

9. 有根概率树与熵的计算

在一棵有根概率树中，叶的平均深度等于除叶之外所有节点（包括根）概率的和。也就是说，如果有根概率树所有内部节点的概率分别为 q_1, q_2, \dots, q_M ，那么叶节点平均深度为

$$\bar{l} = \sum_{j=1}^M q_j \quad (2.27)$$

离散随机变量的熵等于所对应的有根概率树上所有内部节点（包括根节点，不包括叶）的分支熵用该节点概率加权的和，即

$$H(X) = \sum_i q(u_i) H(u_i) \quad (2.28)$$

其中， $q(u_i)$ 为节点 u_i 的概率， $H(u_i)$ 为节点 u_i 的分支熵。

2.1.3 平均互信息

1. 离散随机变量与事件之间的互信息

- 定义

$$I(x; Y) = \sum_y p(y|x) \log \frac{p(y|x)}{p(y)} \quad (2.29)$$

它表示由事件 x 提供的关于集合 Y 的信息量（注意：用条件概率平均）。

- 性质

$$I(x; Y) \geq 0 \quad (2.30)$$

仅当 x 与所有 y 独立时，等式成立。

2. 平均互信息

集合 X 、 Y 之间的平均互信息定义为

$$\begin{aligned} I(X;Y) &= \sum_x p(x) I(x;Y) = \sum_{x,y} p(x)p(y|x) \log \frac{p(y|x)}{\sum_x p(x)p(y|x)} \\ &= \sum_{i,j} p_i p_{ij} \log \frac{p_{ij}}{\sum_i p_i p_{ij}} = E_{p(xy)} [I(x;y)] \end{aligned} \quad (2.31)$$

平均互信息是从整体上表示一个随机变量 Y 所给出的关于另一个随机变量 X 的信息量。

3. 平均互信息与熵的关系

$$I(X;Y) = H(X) - H(X|Y) \quad (2.32)$$

$$I(X;Y) = H(Y) - H(Y|X) \quad (2.33)$$

$$I(X;Y) = H(X) + H(Y) - H(XY) \quad (2.34)$$

4. 平均互信息的性质

- 非负性

$$I(X;Y) \geq 0 \quad (2.35)$$

仅当 X 、 Y 独立时，等式成立。

- 对称性

$$I(X;Y) = I(Y;X) \quad (2.36)$$

- 凸函数性

(1) 为概率分布 $p(x)$ 的上凸函数。

(2) 为条件概率 $p(y|x)$ 的下凸函数。

- 极值性

$$I(X;Y) \leq H(X) \quad (2.37)$$

$$I(X;Y) \leq H(Y) \quad (2.38)$$

5. 平均条件互信息

• 设联合集 XYZ ，在 Z 条件下， X 与 Y 之间的平均互信息定义为条件互信息 $I(x;y|z)$ 的平均值，即

$$I(X;Y|Z) = E_{p(xyz)} [I(x;y|z)] = \sum_{x,y,z} p(xyz) \log \frac{p(x|yz)}{p(x|z)} \quad (2.39)$$

通过计算可得

$$\begin{aligned} I(X;YZ) &= I(X;Z|Y) + I(X;Y) \\ &= I(X;Y|Z) + I(X;Z) \end{aligned} \quad (2.40)$$