21世纪高等院校信息与通信工程规划教材

21st Century University Planned Textbooks of Information and Communication Engineering

信息论基础习题解答

田宝玉 杨洁 贺志强 许文俊 王晓湘 编著

Elements of Information Theory: Solutions to Exercises

名师名校

人民邮电出版社
POSTS & TELECOM PRESS

21世纪高等院校信息与通信工程规划教材

21st Century University Planned Textbooks of Information and Communication Engineering

信息论基础习题解答

田宝玉 杨洁 贺志强 许文俊 王晓湘 编著

Elements of Information Theory: Solutions to Exercises

名师名校

人民邮电出版社

北京

图书在版编目（CIP）数据

信息论基础习题解答 / 田宝玉等编著. -- 北京 :
人民邮电出版社, 2010.10
21世纪高等院校信息与通信工程规划教材
ISBN 978-7-115-22455-2

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

中国版本图书馆CIP数据核字(2010)第066571号

内 容 提 要

《信息论基础习题解答》是本科生教材《信息论基础》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理论教学的质量。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实践提炼并得到验证的典型题。

本书与主教材结构相同，也对应包含 12 章。内容包括“知识要点”、“例题精解”、“习题解答”、“补充题解”4 个部分。另外，有些章节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。通过类似的训练可提高读者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，特别适合于自学信息论的读者阅读，也可作为信息论教师的教学参考资料。

21 世纪高等院校信息与通信工程规划教材

信息论基础习题解答

◆ 编　著　田宝玉　杨　洁　贺志强　许文俊　王晓湘
责任编辑　贾　楠
◆ 人民邮电出版社出版发行　北京市崇文区夕照寺街 14 号
邮编　100061　电子函件　315@ptpress.com.cn
网址　http://www.ptpress.com.cn
中国铁道出版社印刷厂印刷
◆ 开本：787×1092　1/16
印张：15.5　2010 年 10 月第 1 版
字数：381 千字　2010 年 10 月北京第 1 次印刷

ISBN 978-7-115-22455-2

定价：29.00 元

读者服务热线：(010)67170985　印装质量热线：(010)67129223
反盗版热线：(010)67171154

前　言

信息论基础即香农（C. E. Shannon）信息论，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和研究生的必修课程。信息论基础课程以信息熵为基本概念，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，研究通信系统中信息的度量、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》（普通高等教育“十一五”国家级规划教材，田宝玉等编著，人民邮电出版社，2008 年 8 月出版）。国内外信息论课程教学实践证明，一本好的教科书，必须配备精心挑选的习题，而这些习题都应该有相应的解答。这对于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是很必要的；对从事本课程教学的教师也有重要的参考作用。

《信息论基础习题解答》（以下简称《题解》）就是为上述本科生教材《信息论基础》配套的辅助教学资料，主要目的就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理论教学的质量。

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，还需要更多有自身特点的习题解答的图书出现，以使初学者能够了解多种风格的问题描述与分析方式，相互借鉴，取长补短，有利于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写成的，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题。

本书与主教材结构相同，也对应包含 12 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较细致的分析与求解。为灵活运用所学的知识，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”，与教材每章的习题相对应，解题步骤与第二部分相比略有简化。第四部分是“补充题解”。为保证本书具有较大的信息量，本书不包含原教材中的例题（虽然也很重要）。

本书在有些章节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。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。

为使读者不陷入枯燥和抽象理论问题的迷宫，本书还尽量选择贴近实际而且有趣的问题，例如，天气与信息的问题、决策与编码的问题、鉴别假币与信息的问题等，以便提高学生学习本课程的兴趣。实践证明，这些内容对于不是专门从事信息论研究的人员也有一定的参考价值。

本书由田宝玉主编，杨洁、贺志强、许文俊、王晓湘参与编写。

本书在写作过程中，作者得到北京邮电大学信息与通信工程学院的支持与资助，得到北邮信息与通信工程学院领导与信息理论与技术教研中心广大教师的支持，特别是林雪红老师还仔细校对了本书的第 3 章和第 4 章，在此一并表示感谢。编者在本书的编写过程中参考了大量的国内外著作，在此向有关作者表示深切谢意。

因编者水平所限，错误与疏漏之处在所难免，敬请广大读者批评指正。

编　者

2009 年 12 月于北京邮电大学

目　录

第 1 章 绪论

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。主要有两种噪声：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

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

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

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

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$D \geqslant 0$，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$\geqslant R(D)$时，码的平均失真$\leqslant D$；反之，如果信源编码的码率$< R(D)$，就不存在使平均失真$\leqslant D$的编码。可以简述为

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

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$\leqslant D$）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。
- 关心的是最优系统的性能和怎样达到这个性能，并不具体设计系统。
- 研究的是语法信息中的概率信息，要求信源为随机过程。

第 2 章 离散信息的度量

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

$$I_X(a_i)=-\log P_X(a_i) \tag{2.1}$$

简记为

$$I(x)=-\log p(x) \tag{2.2}$$

注意：

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

（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

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

简记为

$$I(xy) = -\log p(xy) \tag{2.4}$$

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

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

简记为

$$I(x|y) = -\log p(x|y) \tag{2.6}$$

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

$$I_{X;Y}(a_i; b_j) = \log \frac{P_{X|Y}(a_i | b_j)}{P_X(a_i)} \tag{2.7}$$

简记为

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

通过计算可得

$$I(x;y) = I(x) - I(x|y) \tag{2.9}$$

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

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

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. 信息熵的定义与计算

- 离散随机变量集合 X 的熵定义为自信息的平均值，记为 $H(X)$

$$H(X)=\underset{p(x)}{E}[I(x)]=-\sum_{x}p(x)\log p(x) \tag{2.11}$$

简记为

$$H(X)=H(p_1,p_2,\cdots,p_n) \tag{2.12}$$

注意：

（1）$I(x)$ 为事件 x 的自信息，$\underset{p(x)}{E}$ 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）$\sum_{i=1}^{n}p_i=1$，$0\leqslant p_i\leqslant 1$，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. 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

$$\begin{aligned}H(Y|X)&=\underset{p(xy)}{E}[I(y|x)]\\&=-\sum_{x}\sum_{y}p(xy)\log p(y|x)\\&=\sum_{x}p(x)[-\sum_{y}p(y|x)\log p(y|x)]\\&=\sum_{x}p(x)H(Y|x)\end{aligned} \tag{2.13}$$

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。

3. 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

$$H(XY)=\underset{p(xy)}{E}[I(xy)]=-\sum_{x}\sum_{y}p(xy)\log p(xy) \tag{2.14}$$

4．凸函数

● 上凸（cap）函数

多元函数 $f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$ 称为定义域上的**上凸（cap）函数**，若对于 $\alpha(0\leqslant\alpha\leqslant1)$ 及任意两矢量 $\boldsymbol{x}_1,\boldsymbol{x}_2$，有

$$f[\alpha\boldsymbol{x}_1+(1-\alpha)\boldsymbol{x}_2]\geqslant\alpha f(\boldsymbol{x}_1)+(1-\alpha)f(\boldsymbol{x}_2) \tag{2.15}$$

成立。当且仅当 $\boldsymbol{x}_1=\boldsymbol{x}_2$、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 $f(\boldsymbol{x})=f(x_1,x_2,\cdots,x_n)$ 称为定义域上的**下凸（cup）函数**，若对于 $\alpha(0\leqslant\alpha\leqslant1)$ 及任意两矢量 $\boldsymbol{x}_1$，$\boldsymbol{x}_2$，有

$$f[\alpha\boldsymbol{x}_1+(1-\alpha)\boldsymbol{x}_2]\leqslant\alpha f(\boldsymbol{x}_1)+(1-\alpha)f(\boldsymbol{x}_2) \tag{2.16}$$

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

● Jenson 不等式

若 $f(\boldsymbol{x})$ 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 $\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_q$ 和任意一组 $\lambda_1,\lambda_2,\cdots,\lambda_q$，$\sum\lambda_k=1$，有

$$f[\sum_{k=1}^{q}\lambda_k\boldsymbol{x}_k]\geqslant\sum_{k=1}^{q}\lambda_k f(\boldsymbol{x}_k) \tag{2.17}$$

当且仅当 $\boldsymbol{x}_1=\boldsymbol{x}_2=\cdots=\boldsymbol{x}_q$ 或 $\lambda_k=1(1\leqslant k\leqslant q)$ 且 $\lambda_j=0(j\neq k)$ 时，等式成立。特别是

$$E[\log x]\leqslant\log[E(x)] \tag{2.18}$$

● 重要不等式

$$1-\frac{1}{x}\leqslant\ln x\leqslant x-1 \tag{2.19}$$

仅当 $x=1$ 时等号成立。

5．信息散度

● 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

$$D(P//Q)=\sum_{x}P(x)\log\frac{P(x)}{Q(x)} \tag{2.20}$$

散度又称做相对熵、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

$$D(P//Q)\geqslant0 \tag{2.21}$$

当且仅当对所有 x，$P(x)=Q(x)$ 时，等式成立。

6．熵的基本性质

- 对称性

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

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

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

$$H(\boldsymbol{p})=H(p_1,p_2,\cdots,p_n)\geqslant 0 \tag{2.22}$$

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。

- 扩展性

$$\lim_{\varepsilon\to 0} H_{n+1}(p_1,p_2,\cdots,p_n-\varepsilon,\varepsilon)=H_n(p_1,p_2,\cdots,p_n) \tag{2.23}$$

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

$$H(XY)=H(X)+H(Y\mid X) \tag{2.24}$$

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 $X_1X_2\cdots X_n$，则有

$$H(X_1X_2\cdots X_n)=H(X_1)+H(X_2\mid X_1)+\cdots+H(X_n\mid X_1\cdots X_{n-1}) \tag{2.25}$$

上式称做熵的链原则。熵的可加性可以这样来理解：复合事件集合的不确定性为各个分事件集合的不确定性之和。

- 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。

- 确定性

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

即当随机变量集合中任一事件概率为 1 时，熵就为 0。

- 上凸性

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

7．各类熵之间的关系

- 条件熵与信息熵的关系

$$H(Y\mid X)\leqslant H(Y) \tag{2.26}$$

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。

- 联合熵与信息熵的关系

$$H\left(X_1X_2\cdots X_n\right)\leqslant\sum_{i=1}^{n}H(X_i) \tag{2.27}$$

仅当各 X_i 相互独立时，等式成立。

2.1.3 平均互信息

1. 集合与事件之间的互信息

- 定义

$$I(x;Y)=\sum_{y}p(y\mid x)\log\frac{p(y\mid x)}{p(y)} \tag{2.28}$$

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

$$I\left(x;Y\right)\geqslant 0 \tag{2.29}$$

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

$$\begin{aligned}I(X;Y)&=\sum_{x}p(x)I(x;Y)=\sum_{x,y}p(x)p(y\mid x)\log\frac{p(y\mid x)}{\sum_{x}p(x)p(y\mid x)}\\&=\sum_{i,j}p_ip_{ij}\log\frac{p_{ij}}{\sum_{i}p_ip_{ij}}=\underset{p(xy)}{E}[I(x;y)]\end{aligned} \tag{2.30}$$

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

$$I(X;Y)=H(X)-H(X\mid Y) \tag{2.31}$$

$$I(X;Y)=H(Y)-H(Y\mid X) \tag{2.32}$$

$$I(X;Y)=H(X)+H(Y)-H(XY) \tag{2.33}$$

4. 平均互信息的性质

- 非负性

$$I(X;Y)\geqslant 0 \tag{2.34}$$

仅当 X ，Y 独立时，等式成立。

- 对称性

$$I(X;Y)=I(Y;X) \tag{2.35}$$

- 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

$$I(X;Y) \leqslant H(X) \tag{2.36}$$

$$I(X;Y) \leqslant H(Y) \tag{2.37}$$

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

$$I(X;Y|Z) = \underset{p(xyz)}{E}[I(x;y|z)] = \sum_{x,y,z} p(xyz)\log\frac{p(x|yz)}{p(x|z)} \tag{2.38}$$

通过计算可得

$$\begin{aligned} I(X;YZ) &= I(X;Z|Y) + I(X;Y) \\ &= I(X;Y|Z) + I(X;Z) \end{aligned} \tag{2.39}$$

- 平均条件互信息是非负的，即

$$I(X;Y|Z) \geqslant 0 \tag{2.40}$$

仅当 $p(x|z) = p(x|yz)$ 时，等式成立。

- 平均互信息的链原则

$$I(X_1X_2\cdots X_n;Y) = \sum_{i=1}^{n} I(X_i;Y|X_1X_2\cdots X_{i-1}) \tag{2.41}$$

- 其他

$$I(X;YZ) \geqslant I(X;Z) \tag{2.42}$$

仅当 $p(x|z) = p(x|yz)$ 时，等式成立。

$$I(X;YZ) \geqslant I(X;Y) \tag{2.43}$$

仅当 $p(x|y) = p(x|yz)$ 时，等式成立。

2.2 例题精解

例 2.1 在某城市，下雨和晴天的时间各占一半，而天气预报无论在雨天还是在晴天都有 2/3 的准确率。甲先生每天上班这样处理带伞问题：如果预报有雨，他就带雨伞上班；如果预报无雨，他也有 1/3 的时间带伞上班。

（1）求事件“在雨天条件下甲先生未带伞”所含的信息量。

（2）求“甲先生带伞条件下没有下雨”的信息量。

（3）求天气预报所得到的关于天气情况的信息量。

（4）求通过观察甲先生是否带伞所得到的关于天气情况的信息量。

解