高等学校通信教材

gaodeng xuexiao tongxin jiaocai

◎ 吴伟陵 编著

XINXI CHULI YU BIANMA

# 信息处理与编码（修订本）

人民邮电出版社
POSTS & TELECOMMUNICATIONS PRESS

高等学校通信教材

# 信息处理与编码(修订本)

吴伟陵　编著

人民邮电出版社

**图书在版编目(CIP)数据**

信息处理与编码(修订本)/吴伟陵编著. --北京:人民邮电出版社,2003.7
ISBN 7-115-11184-7

I. 信... II. 吴... III. ①信息处理—信息技术②信息—编码 IV. G202

中国版本图书馆 CIP 数据核字(2003)第 018297 号

## 内 容 提 要

本书重点介绍信息处理的理论基础以及实现原理与方法。

全书分两篇共计 7 章。其中第一篇重点讨论信息处理的理论基础:信息论,内容分为 3 章:无失真信源与信息熵,限失真信源与信息率失真 $R(D)$ 函数,信道容量。第二篇主要讨论信息与通信系统中信息处理的主要原理、手段与方法,内容分为 4 章:信息与通信系统的优化,提高有效性的信源编码,提高安全性的密码,提高可靠性的信道编码。

本书在原版本的基础上增加了部分内容,并根据广大读者的要求,对书中的所有习题(除第 4 章)都提供了参考答案。全书概念清晰、文字流畅、深入浅出。

本书作为教育部理工类重点教材,供信息工程、通信工程及其相关专业使用,亦可作为信息、通信、电子等部门教学、科研以及工程技术人员参考用书。

高等学校通信教材

**信息处理与编码(修订本)**

◆ 编　　著　吴伟陵
责任编辑　潘春燕

◆ 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
读者热线　010-67129260
北京汉魂图文设计有限公司制作
北京隆昌伟业印刷有限公司印刷
新华书店总店北京发行所经销

◆ 开本:787×1092　1/16
印张:24.75
字数:626 千字　　2003 年 7 月第 2 版
印数:6 501-11 500 册　　2003 年 7 月北京第 4 次印刷

ISBN 7-115-11184-7/TP・3394

定价:33.00 元

# 前　言

本书是为高等学校理工类信息、通信工程专业编写的教育部重点教材。主要内容是介绍信息与通信中信息处理的理论与技术。

为了适应高等学校教学改革和教材改革的需求，作者总结了近20多年在北京邮电大学为研究生、本科生开设《信息论基础》、《信息处理与编码》等多门课程的教学实践，认为有必要将信息工程专业中的主要专业基础课与相关的专业课，以信息与通信系统的优化为主线，力求简明、扼要、适用，编写一本以介绍信息工程中的基本理论、基本技术、基本方法为主的教材，以达到加强基础、拓宽专业、压缩专业课程的目的。

编写本书是一种尝试，它将两类不同要求的基础与专业、理论与技术以系统优化为主线有机地结合起来，在编写上有一定难度。不同的要求必然体现在前后两篇教材取材上的差异，前者以基本理论为主，而后者则以基本原理、基本技术与基本方法为主。

信息处理是现代信息工程的核心技术。信息处理的主要目的是为了提高系统对某一方面的要求以及优化系统某一方面的性能指标。信息处理的主要手段是变换，即编、译码。比如为了提高系统的有效性，可以通过信源编码来实现；为了提高系统的安全性，可以通过密码来实现；为了提高系统的可靠性，可以通过信道编码来实现等。

本书引用信息论的基本观点，重点探讨了在信息与通信系统中的不同性能指标要求下的优化理论、优化措施、实现原理，并着重介绍一些先进的信息处理方法。本书第一篇，即前3章为理论基础，主要介绍信息的基本理论与基本概念：信息熵、互信息、信息率失真 $R(D)$ 函数，信道容量以及它们的性质与计算。这一部分也可以作为单独的一门专业基础课“信息论基础”开设。本书的第二篇，即后4章则侧重于应用，主要介绍信息与通信系统中的优化及其实现手段与方法。其中第4章着重从系统的总体上讨论系统的优化，它是本书承上启下的部分。第5章介绍提高系统有效性的信源编码，包括信源编码定理、无失真的统计匹配编码、解除相关性的预测编码与变换编码，以及实用性的传真游程编码、话音编码与图像编码等。第6章介绍提高系统安全性的密码，包括仙农密码定理与保密理论、经典保密、序列(流)加密、分组(块)加密，并且简介了双钥体制中的公开密钥与认证技术等。第7章介绍提高系统可靠性的信道编码，包括信道编码定理、分组码的基本理论与方法、卷积码的基本理论与方法，以及交织技术、级连码、信道编码的构造性能界限等。

本书为了适应本科与研究生不同层次以及不同专业及专业化的需求，将内容分为基本内容与提高部分两个层次。一些较深的提高部分的章节均加“*”号以示区别，以备选择。

本书作为教材，在每个章节后面还配有习题，并附习题解答。

本书在写作过程中，得到教研室同志大力支持和鼓励。周炯槃院士亲自审阅了原稿，并提出很多建设性意见；教研室主任田宝玉教授详细地校阅了原稿，也提出很多宝贵意见，并亲自遴选了各章配套习题与解答。研究生牛凯、黄贻青、高露、周胜、曹昊加等先后为本书整理录入、打印、排版、画图，作了大量的工作。本书在原版本基础上增加了部分内容，并对书中的所

有习题(除第 4 章)都提供了参考答案。答案部分由博士生吴湛击、鲁燕萍完成,在此对他们表示感谢。

鉴于本人水平有限,错误难免,望多批评指教。

吴伟陵　于北京邮电大学

2002 年 8 月

# 目　录

## 第一篇　信息论基础


**第 1 章　无失真信源与信息熵** …… 2

1.1　信源特性与分类 …… 2

1.1.1　信源的统计特性 …… 2

1.1.2　信源的描述与分类 …… 2

1.2　离散信源的信息熵 …… 7

1.2.1　信息熵和信息量的基本概念 …… 7

1.2.2　熵的数学性质 …… 9

＊1.2.3　熵的公理化结构 …… 12

1.3　离散序列信源的熵 …… 14

1.3.1　离散无记忆信源的序列熵 $H(\boldsymbol{U})$ 与消息熵 $H_L(U)$ …… 14

1.3.2　离散有记忆信源的序列熵 $H(\boldsymbol{U})$ 与消息熵 $H_L(U)$ …… 15

1.4　互信息 …… 18

1.4.1　单个消息的互信息 …… 18

＊1.4.2　消息序列的互信息 $I(\boldsymbol{U};\boldsymbol{V})$ …… 21

1.4.3　信息不增性原理 …… 23

1.5　冗余度 …… 24

1.6　连续信源的熵与互信息 …… 26

习题 …… 34

**第 2 章　限失真信源与信息率失真函数** …… 39

2.1　引言 …… 39

＊2.2　$R(D)$ 函数的性质 …… 42

2.3　离散信源 $R(D)$ 函数的计算 …… 44

2.3.1　等概率对称性失真信源 $R(D)$ 函数的计算 …… 44

＊2.3.2　一般情况下的参量表达式 …… 47

＊2.3.3　$R(D)$ 函数的迭代算法 …… 50

2.4　连续(模拟)信源的信息率失真函数 $R(D)$ …… 52

习题 …… 59

**第 3 章　信道与信道容量** …… 62

3.1　信道的分类与描述 …… 62

3.1.1　信道的分类 …… 62

3.1.2　信道描述 …… 63

3.2 无干扰离散信道 …… 64
3.3 离散单个消息(符号)信道及其容量 …… 68
*3.4 离散消息序列信道及其容量 …… 74
*3.4.1 无记忆离散消息序列信道 …… 74
*3.4.2 有记忆离散消息序列信道 …… 75
3.5 连续信道及其容量 …… 76
3.5.1 连续单个消息信道及其容量 …… 76
3.5.2 一般迭加性干扰的单消息连续信道 …… 77
3.5.3 限时限频限功率的白色高斯噪声信道 …… 79
*3.5.4 有公共约束的连续消息序列信道 …… 81
3.6 信道容量代价函数 $C(F)$ 及信道冗余度 …… 84
3.6.1 信道容量代价函数 $C(F)$ …… 84
3.6.2 信道冗余度 …… 84
3.7 多用户信道 …… 85
3.7.1 引言 …… 85
3.7.2 多址信道 …… 87
*3.7.3 广播信道 …… 90
*3.7.4 相关信源的多用户信道 …… 93
习题 …… 95

## 第二篇 信息处理的实现方法

**第4章 信息与通信系统的优化** …… 100
4.1 信息与通信系统的物理和数学模型 …… 100
4.2 信息与通信系统的单指标优化 …… 103
习题 …… 107
**第5章 信源编码** …… 109
5.1 无失真信源编码 …… 109
5.1.1 等长编码定理 …… 111
5.1.2 变长编码定理 …… 113
5.1.3 最佳变长编码—哈夫曼编码 …… 118
*5.1.4 算术编码 …… 122
*5.2 限失真信源编码定理 …… 126
*5.3 矢量量化编码 …… 130
*5.3.1 最佳标量量化编码 …… 130
*5.3.2 矢量量化编码 …… 131
5.4 预测编码 …… 134
5.4.1 预测编码的基本原理 …… 134
5.4.2 预测编码的基本类型 …… 136
5.5 变换编码 …… 140

5.5.1　正交变换的基本数学知识 …… 141
5.5.2　几种主要变换编码 …… 141
*5.5.3　小波变换编码 …… 150
5.6　传真编码 …… 151
5.6.1　文件传真的基本特性 …… 152
*5.6.2　三、四类传真机的实用化压缩编码 …… 155
5.7　语音压缩编码 …… 161
5.7.1　波形编码 ADPCM 基本原理 …… 163
5.7.2　参量编码的线性预测编码器 LPC …… 164
5.7.3　混合编码的各类方法 …… 165
*5.7.4　低延迟码激励线性预测(LD-CELP)编码器 …… 167
*5.7.5　共轭结构——代数码激励线性预测编码器 …… 168
*5.7.6　第三代移动通信中的语音编码 …… 169
5.8　图像编码 …… 172
5.8.1　静止图像压缩编码及其技术标准 JPEG …… 173
5.8.2　面向通信的视频压缩编码及其技术标准 H.261 …… 175
5.8.3　活动图像压缩编码及其技术标准 MPEG …… 178
*5.8.4　第二代视频编码 …… 180
习题 …… 183
**第6章　密码** …… 186
6.1　密码学的基本概念 …… 186
6.2　保密学的理论基础 …… 190
6.3　序列(流)密码 …… 198
6.4　分组(块)密码 …… 207
*6.5　公开密钥密码 …… 218
*6.6　认证系统 …… 223
6.7　模拟消息加密体制 …… 230
*6.8　GSM 的鉴权与加密 …… 236
习题 …… 238
**第7章　信道编码** …… 242
7.1　信道编码的基本概念 …… 242
7.2　线性分组码 …… 247
7.3　循环码 …… 256
7.4　BCH 码 …… 263
7.5　卷积码 …… 269
7.5.1　卷积码编码 …… 270
7.5.2　卷积码的译码 …… 275
*7.5.3　卷积码的距离特性 …… 283
7.6　纠正突发错误码 …… 284

7.7 交织码 …… 287
7.8 级连码 …… 291
*7.9 信道编码的性能界限 …… 293
*7.9.1 信道编码定理 …… 294
*7.9.2 信道编码的构造性能界限 …… 298
*7.10 实际信道编码应用 …… 301
7.11 Turbo码 …… 306
7.12 高效率信道编码 TCM …… 310
习题 …… 317

## 附录 习题参考答案

参考文献 …… 386

# 第一篇　信息论基础

本篇主要讨论信息处理的理论基础，并侧重从“概念、描述、度量、分析与计算”5个方面予以讨论。主要内容包括：第1章无失真信源与信息熵，它将以单个离散消息信源为主体，讨论熵的基本性质与计算。第2章限失真信源及信息率失真函数，它也以单个离散消息信源为主体，讨论信息率失真函数的基本性质与计算。第3章信道与信道容量，它仍以输入单个离散消息为主体，讨论各类信道及其容量的计算。

当今科学技术的发展已使人类进入了一个新的时代，这个时代的主要特征之一是对信息的需求与利用，所以人们又称它为信息时代。

那么，什么是信息呢？确切地说信息至今无定义，但它是一个人人皆知的概念，这一概念具有两大特征：广泛性与抽象性。

所谓广泛性是指客观世界充满着信息，上至宇宙天体、下至地面矿藏无不含有客观的特征信息；人类生存离不开信息，人的五官不停地接收信息，人的神经系统在不断地传递信息，人的大脑则不停地处理、存储和利用信息。

所谓抽象性则体现在它是组成客观世界的三大支柱之一：物质、能量与信息。其中物质是最基本、最具体的，能量可以看成物质的运动形式，然而信息既不是物质也不是能量，但又离不开物质和能量，它是人类认识和改造世界的新层次。

下面，我们暂避开上述广义信息的概念，集中精力探讨通信中信息的概念。在通信中，信息表达有3个层次：信号、消息与信息。其中信号最具体，它是一个物理量，可测量、可显示、可描述，同时它又是载荷信息的实体，故称它为信息的物理表达层；消息或称为符号，它是为了从数学上进一步描述与表达不同形式的信号，将信号划分为离散（数字）消息与连续（模拟）消息两大类以便于采用随机变量、随机序列与随机过程进行分析，称它为数学表达层，它也是信息论中主要描述方式；3个层次中最抽象的是哲学表达层的信息。从哲学上看，3者关系是属于一种内涵与外延的关系，即信息是具体信号与消息的内涵，是信号载荷的内容，是消息描述的对象。反过来，信号则是信息在物理表达上的外延，消息则是信息在数学表达上的外延。同一信息，可以采用不同形式物理量（声音、图像、文字）来载荷，也可以采用不同的数学描述方式（数学或模拟），同样，同一类型信号或消息也可以代表不同内容的信息。

# 第 1 章　无失真信源与信息熵

## 1.1　信源特性与分类

### 1.1.1　信源的统计特性

首先讨论什么是信源？直观地说，信源是信息的来源。在实际通信中最常见的信源有话音、文字、图像、数据等。

在信息论中，确切地说信源是产生消息（符号）、消息序列和连续消息的来源。从数学上看，由于消息的不确定性，因此，信源是产生随机变量、随机序列和随机过程的源。

其次，讨论信源的特性。客观信源的基本特性是具有随机不确定性。这就是说在经典的仙农信息论中，仅讨论客观信源的概率统计特性以及在此基础上的客观概率信息。

### 1.1.2　信源的描述与分类

首先讨论离散单个消息（符号）信源。它是最简单的也是最基本的信源，是组成实际信源的最基本单元。根据单消息信源随机变量 $u$ 的取值范围和对应的概率 $p(u)$，共同组成一个二元有序对$[U,p(u)]$，用它来描述无失真单消息的离散信源，即

$$\begin{bmatrix} U \\ p(u) \end{bmatrix} = \begin{bmatrix} U=u_1, & U=u_2, & \cdots, & U=u_n \\ p_1, & p_2, & \cdots, & p_n \end{bmatrix} \tag{1-1-1}$$

例如：对二进制数字与数据信源 $U=\{0,1\}$

$$\begin{bmatrix} U \\ p \end{bmatrix} = \begin{bmatrix} 0,1 \\ p_0,p_1 \end{bmatrix} \xrightarrow{\text{当 } p_0=p_1=\frac{1}{2}} \begin{bmatrix} 0 & , & 1 \\ \frac{1}{2} & , & \frac{1}{2} \end{bmatrix}$$

显然，对于单个连续随机变量信源也可表示为：

$$\begin{bmatrix} U \\ p(u) \end{bmatrix} = \begin{bmatrix} (a,b) \\ p(u) \end{bmatrix}$$

其中 $u \in U=(-\infty,+\infty)$，$p(u)$ 为概率密度函数。

其次，讨论实际信源。实际信源不可能仅发送单个消息（符号），对离散信源而言，发送的是一组消息（符号）串，即一个随机序列；对连续信源而言则是一随机过程。前者如电报、数据以

及离散化后的话音与图像；后者如模拟话音、模拟图像等。

离散序列信源

$$\boldsymbol{U}=(U_1,U_2,\cdots,U_l,\cdots,U_L)\in U^L=\overbrace{U\times U\times\cdots\times U}^{L\text{个}} \tag{1-1-2}$$

其中 $l=1,2,\cdots,L$ 为离散消息序列的长度。而 $U_l$ 为第 $l$ 时刻的随机变量，且每个离散随机变量消息都有 $n$ 种可能取值。

这时，离散随机序列 $\boldsymbol{U}$ 的样值 $\boldsymbol{u}$ 可表示为：

$\boldsymbol{u}=(u_1,\cdots,u_l,\cdots,u_L)$

因而，$\boldsymbol{u}\in n^L=n\times n\times\cdots\times n$(共 $L$ 个)。即每个随机变量取值有 $n$ 种，那么 $L$ 个随机变量组成的随机序列，其样值共有 $n^L$ 种可能取值。其对应的概率为：

$$p(\boldsymbol{u})=p(u_1,u_2,\cdots,u_l,\cdots,u_L) \tag{1-1-3}$$

所以可以采用二元序对$[U^L,p(\boldsymbol{u})]$来描述，即

$$\begin{bmatrix}U^L\\p(\boldsymbol{u})\end{bmatrix}=\begin{bmatrix}\boldsymbol{U}=\boldsymbol{u}_1,\cdots, & \boldsymbol{U}=\boldsymbol{u}_l, & \cdots, & \boldsymbol{U}=\boldsymbol{u}_{n^L}\\ p(\boldsymbol{u}_1),\cdots, & p(\boldsymbol{u}_1), & \cdots, & p(\boldsymbol{u}_{n^L})\end{bmatrix} \tag{1-1-4}$$

例：以最简单的 3 位 PCM 信源为例，这时 $n=2,L=3$，即

$$\begin{bmatrix}U^3\\p(\boldsymbol{u})\end{bmatrix}=\begin{bmatrix}\boldsymbol{U}=000, & \boldsymbol{U}=001, & \cdots, & \boldsymbol{U}=111\\ p_0^3, & p_0^2\cdot p^1, & \cdots, & p_1^3\end{bmatrix}$$

$$\underline{\underline{\text{当 } p_0=p_1=\frac{1}{2}}}\ \text{时}\begin{bmatrix}000, & 001, & \cdots, & 111\\ \frac{1}{8}, & \frac{1}{8}, & \cdots, & \frac{1}{8},\end{bmatrix}$$

对于这类离散序列信源，还可以进一步划分为离散无记忆、离散有记忆信源。对于离散无记忆信源我们有：

$$\begin{aligned}p(\boldsymbol{u})&=p(u_1,\cdots,u_L)\\&=\prod_{l=1}^{L}p(u_l)\qquad\text{(当满足无记忆条件时)}\\&=p^L\qquad\text{(当进一步满足平稳性时)}\end{aligned} \tag{1-1-5}$$

一般地，称这类信源为独立同分布信源，PCM 信源属于此类。

对于离散有记忆信源，由于大部分实际离散信源属于此类，对它的描述要比无记忆信源困难得多，尤其当记忆长度 $L$ 足够大时。然而，很多实际信源是符合有限记忆模型的，这时分析问题可大大简化。对于有限记忆模型数学上常采用马氏链来描述，而马氏链中最简单的是一阶情况。若将离散随机序列消息看作是一阶马氏链，这时信源输出的消息序列中任一时刻的消息 $u_l$，它仅与其前面的一个消息 $u_{l-1}$ 有关，而与更前面的消息无直接关系。这样，消息序列之间的关系就像一环扣一环的链锁。

$$\begin{aligned}p(\boldsymbol{u})&=p(u_1)P(u_2\mid u_1)P(u_3\mid u_2u_1)\cdots P(u_L\mid(u_{L-1}\cdots u_1)\\&=p(u_1)P(u_2\mid u_1)P(u_3\mid\boldsymbol{u}_1^2)\cdots P(u_L\mid\boldsymbol{u}_1^{L-1})\\&=p(u_1)P(u_2\mid u_1)P(u_3\mid u_2)\cdots P(u_L\mid u_{L-1})\quad\text{(对于马氏链)}\\&=p(u_1)\prod_{l=1}^{L-1}p(u_{l+1}\mid u_l)\\&=p_1\cdot P_{ji}^{L-1}\qquad\text{(对于齐次马氏链)}\\&\hat{=}P_{ji}^{L}\qquad\text{(对于齐次遍历马氏链)}\end{aligned} \tag{1-1-6}$$

在实际信源中,数字图像信源往往采用上述一阶马氏链模型作为其一阶近似。

下面,再讨论连续信源。在实际问题中,连续的模拟信源往往可以采用两种方法进行分析。一类是将连续信源离散化为随机序列信源,再采用前面的随机序列信源进行分析;另一类则是直接分析连续模拟信源,但是由于数学上的困难,只能分析单个连续消息变量的信源。以下,我们首先分析第一类可离散化的连续信源,即分析什么样的信源可以进行离散化处理。

实际上,连续模拟信源只要满足一个非常宽松的条件即可,凡满足限时($T$)、限频($F$)的实际连续过程均可以离散化为随机序列。

类似于在信号分析中对周期性确知信号的正交展开,这里也可以对非确知的连续随机信号在满足上述限时($T$)、限频($F$)条件下展开成离散随机序列信号。

下面,我们给出3类最常用展开式:傅氏级数展开、取样函数展开及K-L展开。

首先,介绍对限时($T$)、限频($F$)过程的傅氏展开。对一随机过程$U(t,\omega)$,若满足:

$$\begin{cases} U(t,\omega) \overset{(a.e)}{=} 0 & \text{当} \mid t \mid > T \text{时} \\ U(t,\omega) \overset{(a.e)}{=} \overline{U}(t,\omega) & \text{当} \mid t \mid \leqslant T \text{时} \end{cases} \tag{1-1-7}$$

其中$\overline{U}(t,\omega)$为一周期性过程,(a.e)表示几乎处处收敛*。则我们可以类似于对周期性确知信号,在时域进行傅氏级数展开。

即当$\mid t \mid < T$时,我们有:

$$\begin{aligned} U(t,\omega) &\overset{(a.e)}{=} \overline{U}(t,\omega) \overset{(a.e)}{=} \sum_{n=-\infty}^{\infty} C_n(\omega) e^{j\frac{2n\pi}{T}t} \\ &\overset{\text{限频}^{**}}{=} \sum_{n=-N}^{N} C_n(\omega) e^{j\frac{2n\pi}{T}t} \end{aligned} \tag{1-1-8}$$

其中:

$$C_n(\omega) \overset{(a.e)}{=} \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} \overline{U}(t,\omega) e^{-j\frac{2n\pi}{T}t} dt \overset{(a.e)}{=} \frac{1}{T}\int_{-\frac{T}{2}}^{\frac{T}{2}} U(t,\omega) e^{-j\frac{2n\pi}{T}t} dt \tag{1-1-9}$$

由(1-1-8)式可见,这时随机过程$U(t,\omega)$可以展开为有限个$(2N+1)$随机变量$C_n(\omega)$序列和。

然后,再介绍限时($T$)、限频($F$)过程的取样函数展开。

若一频域随机过程$H(f,\omega)$满足下列条件:

$$\begin{cases} H(f,\omega) \overset{(a.e)}{=} 0 & \text{当} \mid f \mid > F \text{时} \\ H(f,\omega) \overset{(a.e)}{=} \overline{H}(f,\omega) & \text{当} \mid f \mid \leqslant F \text{时} \end{cases} \tag{1-1-10}$$

其中$\overline{H}(f,\omega)$为一频域周期性过程。

那么,我们可以类似于上面时域过程,在频域进行傅氏级数展开。

即当满足限频条件:$\mid f \mid \leqslant F$时,我们有

$$H(f,\omega) \overset{(a.e)}{=} \overline{H}(f,\omega) \overset{(a.e)}{=} \sum_{n=-\infty}^{\infty} g_n(\omega) e^{j\frac{2n\pi}{2F}f} \tag{1-1-11}$$

---

* 对随机变量、随机过程而言,等式只能在统计意义上相等,而(a.e)是almost every where的缩写,是一类最强的统计意义上的相等。

** 由测不准定理,严格限时(频)就不可能限频(时),这里,仅考虑工程上的近似的满足同时限时、限频。

其中：

$$g_n(\omega) \overset{(a.e)}{=} \frac{1}{2F}\int_{-F}^{F} \overline{H}(f,\omega)e^{-j\frac{2\pi n}{2F}f}df \overset{(a.e)}{=} \frac{1}{2F}\int_{-F}^{F} H(f,\omega)e^{-j\frac{2\pi n}{2F}f}df \qquad (1-1-12)$$

由于 $H(f,\omega)$ 与 $U(t,\omega)$ 为一对傅氏变换，即

$$U(t,\omega) \overset{(a.e)}{=} \int_{-F}^{F} H(f,\omega)e^{j2\pi ft}df \qquad (1-1-13)$$

现令 $t=-\frac{n}{2F}$ 则有：

$$\begin{aligned} g_n(\omega) &\overset{(a.e)}{=} \frac{1}{2F}\int_{-F}^{F} H(f,\omega)e^{j2\pi f(-\frac{n}{2F})}df \\ &\overset{(a.e)}{=} \frac{1}{2F}U(-\frac{n}{2F}) \end{aligned} \qquad (1-1-14)$$

所以，我们有：

$$\begin{aligned} U(t,\omega) &\overset{(a.e)}{=} \int_{-F}^{F} H(f,\omega)e^{j2\pi ft}df \overset{(a.e)}{=} \int_{-F}^{F}\left(\sum_{n=-\infty}^{\infty} g_n(\omega)e^{j\frac{2\pi n}{2F}f}\right)e^{j2\pi ft}df \\ &\overset{(a.e)}{=} \sum_{n=-\infty}^{\infty} g_n(\omega)\int_{-F}^{F} e^{j2\pi f(t+\frac{n}{2F})}df \\ &\overset{(a.e)}{=} \sum_{n=-\infty}^{\infty} \frac{1}{2F}U(-\frac{n}{2F})\frac{\sin 2\pi F(t+\frac{n}{2F})}{\pi(t+\frac{n}{2F})} \\ &\xlongequal{m=-n} \sum_{m=-\infty}^{\infty} \frac{1}{2F}U(\frac{m}{2F})\frac{\sin 2\pi F(t-\frac{m}{2F})}{\pi(t-\frac{m}{2F})} \\ &\xlongequal{\text{限时 T}} \sum_{m=-FT}^{FT} \frac{1}{2F}U(\frac{m}{2F})\frac{\sin 2\pi F(t-\frac{m}{2F})}{\pi(t-\frac{m}{2F})} \end{aligned} \qquad (1-1-15)$$

可见，当随机过程 $U(t,\omega)$ 满足限时($T$)、限频($F$) 条件时，它可以展开为一有限个$(2FT+1)$ 样值 $U(\frac{m}{2F})$ 序列来表示。

下面，我们介绍 3 类中的最后一类 K-L 展开。上述两类展开，虽然都可以将一个连续随机过程 $U(t,\omega)$ 在满足限时($T$)、限频($F$) 条件下展成一组有限的随机序列来表示。但是，在一般情况下，其展开的系数之间是统计关联的，即转化后的离散随机序列信源是有记忆的，这将给进一步分析带来一定的困难。对于傅氏展开，仅当 $U(t,\omega)$ 为广义(弱) 平稳时，而对于取样函数展开，仅当谱函数绝对值恒定时，即 $|H(f)|=C$，当 $|f|<F$ 时，即满足白噪声条件，展开后系数之间是统计无关的。能否在理论上寻找一类展开，使其展开后系数满足统计独立或线性无关?这类展开就是 K-L 展开。

设随机过程 $U(t,\omega)$，$t\in T$，若满足

$$\begin{cases} E[U(t,\omega)]=0 \\ R(t_1,t_2)=E[U(t_1,\omega)U(t_2,\omega)] \end{cases}$$

又设，在区间 $T=[a,b]$ 内有一组完备正交函数 $\varphi_i(t)$，$i=1,2,\ldots,n$，其中 $n$ 有限或可数，

即

$$\int_a^b \varphi_i(t)\varphi_j(t)\mathrm{d}t = \begin{cases} 1, & \text{当 } i=j \\ 0, & \text{当 } i\neq j \end{cases} \tag{1-1-16}$$

则：

$$\begin{cases} U(t,\omega) \overset{(a.e)}{=\!=} \sum_{i=1}^{\infty} a_i(\omega)\varphi_i(t) \\ a_i(\omega) \overset{(a.e)}{=\!=} \int_a^b U(t,\omega)\varphi_i(t)\mathrm{d}t \end{cases} \tag{1-1-17}$$

因为 $E[U(t,\omega)]=0$，所以 $E(a_i)=0$。而

$$R(t_1,t_2) = E[\sum_i a_i\varphi_i(t_1)\sum_i a_i\varphi_i(t_2)] \tag{1-1-18}$$

我们希望各 $a_i$ 之间线性无关，即

$$E(a_ia_j) = \begin{cases} \lambda_i=\sigma_i^2, & \text{当 } i=j \\ 0, & \text{当 } i\neq j \end{cases} \tag{1-1-19}$$

所以：

$$R(t_1,t_2) = \sum_i \lambda_i\varphi_i(t_1)\varphi_i(t_2) \tag{1-1-20}$$

两边同乘 $\varphi_j(t_2)$，并在$[a,b]$ 内对 $t_2$ 积分并由归一性得：

$$\begin{aligned} \int_a^b R(t_1,t_2)\varphi_j(t_2)\mathrm{d}t_2 &= \int_a^b \sum_i \lambda_i\varphi_i(t_1)\varphi_i(t_2)\varphi_j(t_2)\mathrm{d}t_2 \\ &\overset{\text{当 } i=j \text{ 时}}{=\!=\!=} \lambda_j\varphi_j(t_1)\int_a^b \varphi_j(t_2)\varphi_j(t_2)\mathrm{d}t_2 = \lambda_j\varphi_j(t_1) \end{aligned} \tag{1-1-21}$$

可见，所需正交函数系应满足下列积分方程：

$$\int_a^b R(t_1,t_2)\varphi(t_2)\mathrm{d}t_2 = \lambda\varphi(t_1) \tag{1-1-22}$$

所谓积分方程是指未知函数在积分号内的方程式。我们这里所讨论的是最常见的线性积分方程，即 $a(x)\Psi(x)+f(x)=\int_a^b k(x,\xi)\varphi(\xi)\mathrm{d}\xi$。对照上述积分方程，可得 $a(x)=\lambda$，$f(x)=0$ 而积分核 $k(x,\xi)=R(t_1,t_2)$，仅有 $\varphi(\xi)$ 未知。这类积分方程又称为齐次第二类线性积分方程，其核为对称型，它是比较容易求解的。但是它要求其特性值$\lambda_i$ 为某些离散值，而与之对应的正交函数是积分方程的特征函数 $\varphi_i(t)$。可见，当 $R(t_1,t_2)$ 已知时，可解上述积分方程求得特征值 $\lambda_i$ 与特征函数 $\varphi_i(t)$，然后可以将 $U(t,\omega)$ 展成：

$$U(t,\omega) \overset{(a.e)}{=\!=} \sum_{i=1}^{\infty} a_i(\omega)\varphi_i(t) \tag{1-1-23}$$

其展开所得的系数 $a_i(\omega)$ 是线性无关的随机变量。若 $U(t,\omega)$ 为一正态过程，则 $a_i(\omega)$ 是一组统计独立的随机变量。

可见，K-L 展开主要优点在于所展开的系数是线性无关的，对正态过程则是统计独立的。因此展开后可以作为无记忆信源来处理。此外它的收敛速度也是比较快的。但可惜由于目前尚未找到相应的快速算法。另外在概念上又不像傅氏级数和抽样函数展开那样直观，所以通常仅将它作为理论上最佳变换的一个参考标准，而实际上则很少使用。

## 1.2　离散信源的信息熵

### 1.2.1　信息熵和信息量的基本概念

上一节我们讨论了信源的统计特性，它可以用概率来描述。然而信源是信息的来源，那么信息与概率之间是什么关系呢？又怎样才能用概率来定量地描述信源的信息量呢？

本节，我们首先从直观概念出发推导出单个消息（符号）信源的信息测度：信息熵 $H(U)$，再讨论它的基本概念与基本性质，最后再用严格的公理化结构证明熵的惟一性。

信息的定量化是人们长期以来追求的目标，1928 年，信息论的先驱者之一哈特莱（Hartley）首先研究了具有 $N^m$ 个组合的单个消息信源。他对这类非概率（实际是等概率）信源进行了研究，并给出了最早的信息度量公式：

$$I = \log N^m = m\log N \tag{1-2-1}$$

实际上，这一信息度量公式给后来仙农（Shannon）建立概率信息度量以很大的启发。仙农保留了其对数度量的合理性，它直接反映信息具有可加性。并将特殊的等概率情况进一步推广至一般的不等概率情况的概率信源。下面，我们首先从直观概念推导出概率信源的信息度量。

通常，对于单个消息（符号）信源 $U$，发送某一个具体消息 $U = u_i, i = 1,2,\ldots,n$。这时，信源输出的信息 $I$ 应该是消息概率 $p_i$ 的递降函数，即

$$\begin{cases} p_i \downarrow, I(p_i) \uparrow, \text{且当 } p_i \to 0 \text{ 时}, I(p_i) \to \infty \\ p_i \uparrow, I(p_i) \downarrow, \text{且当 } p_i \to 1 \text{ 时}, I(p_i) \to 0 \end{cases} \tag{1-2-2}$$

这就是说，消息 $u_i$ 出现的概率 $p_i$ 越小，一旦它出现以后必然使人们感到很意外，给人的信息量 $I(p_i)$ 就越大。当几乎不可能的消息出现了，则会一鸣惊人，给人以巨大的信息量。反之，消息出现的概率越大，越是在人们预料之中，一旦出现，也不会给人们意外，故信息量 $I(p_i)$ 很小，而对于必然要出现的消息，则不具有任何信息量。

另外，从人们的直观概念上讲，由两个不同的消息（相互统计独立）所提供的信息，等于它们分别提供信息之和，即满足可加性。

由上述两方面，即信息对消息概率的递降性与可加性，可以推导出这类函数应是对数函数，即

$$I(p_i) = \log \frac{1}{p_j} = -\log p_i \tag{1-2-3}$$

通常，我们称 $I(p_i)$ 为单个信源消息的非平均自信息量。它给出某个具体消息 $U = u_i$ 时，信源的信息度量。

同理，可定义消息 $U = u_j$，对应概率为 $q_j$ 时，信源非平均自信息量为：

$$I(q_j) = \log \frac{1}{q_j} = -\log q_j \tag{1-2-3'}$$

另外，可以定义两个独立消息 $u_i$、$u_j$ 同时出现联合概率为 $r_{ij}$ 时的非平均自信息量 $I(r_{ij})$ 为：

$$I(r_{ij}) = -\log p_i q_j = -\log p_i - \log q_j = I(q_j) + I(p_i) \tag{1-2-4}$$

可见,对数函数可以满足可加性,而公式中负号则表示对概率的递减。

若两个消息 $u_i$、$u_j$ 出现不是独立的,而是有相互联系的,这时可引用条件概率来描述。若将在信源消息 $u_i$(或 $u_j$)已出现的条件下,消息 $u_j$(或 $u_i$)出现的条件概率表示为 $P_{ji}$(或 $Q_{ij}$),则其非平均自信息量 $I(P_{ji})$([或 $I(Q_{ij})$])为:

$$I(P_{ji}) = \log\frac{1}{P_{ji}} = -\log P_{ji}$$
$$I(Q_{ij}) = \log\frac{1}{Q_{ij}} = -\log Q_{ij} \qquad (1-2-5)$$

上面,我们从直观概念出发,引入了信源输出单个消息(符号)的非平均自信息量的表达式。实际上,对单个消息信源,由于消息取值有 $n$ 种可能性,即 $u_i$,其中 $i=1,2,\cdots,n$。那么,对信源而言,我们应考虑的不再是某个具体消息的非平均自信息量,而应是单个离散消息信源所有可能取值的有限个($i=1,2,\cdots,n$)消息的统计平均信息量。显然,它与信源概率分布有关。即对单个消息离散信源:

$$\begin{bmatrix} U \\ p(u) \end{bmatrix} = \begin{bmatrix} U=u_1, & \cdots U=u_i, & \cdots, & U=u_n \\ p_1, & \cdots p_i, & \cdots, & p_n \end{bmatrix}$$

信源输出的平均信息量定义为信息熵:

$$H(U) = H(p_1\cdots p_n) \triangleq E[-\log p_i] = -\sum_{i=1}^{n} p_i \log p_i$$

同理可定义:$H(V) = E[I(q_j)] = E[-\log q_j] = -\sum_{j=1}^{m} q_j \log q_j$

$$H(V|U) = E[I(P_{ji})] = E[-\log p_{ji}] = -\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij}\log Q_{ji}$$

$$H(U|V) = E[I(Q_{ij})] = E[-\log Q_{ij}] = -\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij}\log Q_{ij}$$

$$H(U,V) = E[I(r_{ij})] = E[-\log r_{ij}] = -\sum_{i}^{n}\sum_{j}^{m} r_{ij}\log r_{ij} \qquad (1-2-6)$$

其中"$E$"表示求概率统计平均,即求数学期望。

显然,信息熵 $H(U)$ 是单消息信源非平均自信息量 $I(p_i)$ 的概率统计平均值,是消息概率的泛函数。它是描述信源统计特性的一个重要物理量。该式首先是由仙农在 1948 年发表的信息论奠基性论文中给出的。后来,费恩斯坦(Feinstein)等人又进一步证明了当信息满足对概率递减性和可加性条件下,上述信息熵表达式是惟一的。

熵这个名词是仙农从物理学的统计热力学中借用过来的。不过在那里称为热熵,它是用来表达统计热力学中的分子状态混乱程度的一个物理量。在这里,仙农引用它来描述信源平均不确定性,含义是类似的。但在热力学中,已知任何孤立系统的演化,热熵只能增加而不能减少;而在通信中,信息熵只会减少而不会增加,所以也有人称信息熵为负热熵。

信息熵的单位与公式中的对数取底有关。在通信中最常用的是以 2 为底,这时单位为比特(bit);有时,在理论推导中常用以 e 为底较为方便,这时,单位为奈特(Nat);在工程上有时采用以 10 为底运算方便,这时,对应单位为笛特(Det)。它们之间可以引用对数换底公式进行互换,比如:

$$1\text{bit} = 0.693\text{Nat} = 0.301\text{Det} \qquad (1-2-7)$$