信息量与信道容量计算

实验讲义

郭万里 李想 编著

西安电子科技大学

通信与信息工程实验教学中心

2019 年

第一章　绪论

本章节首先从对“信息”发展的介绍开始，引出信息度量的重要性。其中，本文所指的信息度量是指信息量与信道容量的计算。接着说明信息量与信道容量计算对系统设计与系统性能分析的重大意义。

研究背景

人类自诞生的那一天起，就翱翔在信息的浩瀚海洋中。人类社会的存在和发展，每时每刻都与接收信息、传递信息、处理信息和利用信息息息相关。

自古以来，信息的表达、传送、存储和处理等问题都受到了人们的关注。原始人的“结绳记事”是表达、传送和信息存储的最原始方法。古代的“烽火”突破了早期信息传递不及时的缺点，实现了快速而远距离地传递信息。语言和文字的出现解决了信息的表达、传递和储存问题，而使得该问题发生重大变革的则是造纸术和印刷术的这两种伟大发明。在此以后，电报、电话和电视相继被发明，信息加工和传输再次发生了巨大变革，实现了信息的快速便捷和远距离传送。近百年来，随着人类生产力与科学技术的发展，人们对信息的处理、传输、存储、提取和利用的手段及能力不断突破新的更高更强水平。近代以来，随着电子计算机的迅速发展和广泛运用，特别是个人微型计算机的普及，人们在处理、存储、控制和管理信息方面的能力得到了迅速提升。20 世纪 50 年代以后，随着计算机技术、微电子技术、卫星通信和移动通信技术、航空航天技术等新技术的迅猛发展和广泛应用，特别是进入二十一世纪以后，以计算机为主体的互联网行业的迅速兴起和发展，它们相互结合、相互促进，汇成了一股宏大的时代潮流，使得人类社会大踏步跨上了新信息时代。

在当今的信息化社会中，无论何种场合何种时段，各种生产、科学研究和社会活动都涉及信息的交换和利用。要想促进科学技术和国民经济快速发展，就必须实现信息的迅速获取、正确处理以及充分利用。总之，信息的重要性日益增强。

对信息的度量和利用涉及信息论这一学科。当代美国贝尔实验室杰出的科学家香农是信息论的奠基人，他于 1948 年发表了著名的文章——《通信的数学理论》，并为信息论奠定了理论基础。香农信息的定义和度量为我们理解、计算信息量和信道容量提供了理论指导，使得我们对信息的理解更为深刻和科学。

1.1 信息量与信道容量计算的意义

随着通信行业的迅猛发展，通信系统的性能要求也水涨船高，具体体现为对业务要求多样化，对网络要求快速化，对信号要求稳定化。现有的通信系统在不断地完善自身的过程中同样也面临着新的挑战，如通信环境更加复杂、信息安全形势更加严峻以及传输信息规模更加庞大等，这就对通信系统的可靠性和有效性提出了更高的要求。而可靠性与有效性的提高一般是通过增加与减少冗余度来实现的，二者存在矛盾。为了调和高可靠性和高有效性的矛盾，最根本的做法是提高通信信道的性能与容量。针对信息量与信道容量的计算为通信系统的性能估计和系统设计提供了理论指导，而能否又快又准确地预估信息量与信道容量是关乎系统有效性与可靠性的核心。

通信系统的种类多种多样，信源的输入分布和信道的统计特性甚至可能随时改变，因此，对通信过程中信息量与信道容量的准确计算受到了广泛关注。1948年香农提出的信息度量公式，使得人们对通信理论的研究从定性阶段进入到定量阶段，并为信息论等信息学科的发展奠定了坚实的理论基础。由信道容量的定义——信道的最大信息传输率（信息传输率为信道中平均每个符号所能传送的信息量），可知对信道容量的计算可在信息量计算的基础上进行。信道容量还是当前5G 移动通信的重要研究内容之一。但当前针对信息量和信道容量计算的研究并没有系统而完整的归纳，对信息量的计算研究更是寥寥无几，大多只是进行定性说明，定量研究也只是在香农信息量定义的基础上进行；对信道容量的计算大多仅讨论特定的理论模型下的计算方法，类型比较单一。

另外，目前对信息量与信道容量的研究大多只停留在书面的计算阶段，几乎没有人思考过此二者在各种现实场景中的具体意义，缺乏现实的意义支撑。

综上所述，通信系统的可靠性和稳定性的要求日益提高，准确预估信息量和信道容量对系统设计和系统性能分析起着非常重要的作用。系统性地整理信源和信道在不同情况下的计算方法，对工程设计者和研究者可起到指导作用。考虑到当前研究现状的不足，将基于信息论的基础知识，利用 MATLAB 实现对信息量和信道容量的计算。

1.2 信息量与信道容量计算的研究现状

在国内外研究现状中，对信息量的计算，大多以香农的信息论为基础进行，即统计分析输入集的概率分布。文献[1]提出了第二种方法，该方法使用从输入集提取出的拓扑参数，力求提高对概念间语义相似性的估值，实验证明该方法能提供更为精确的相似性估值并能显著提高对信息量的计算性能。

对信道容量的计算，主要的研究思路有三种：其一是使用线性方程组法，其二是使用信道特性值分析法,其三是使用迭代法。其中迭代算法的应用范围较广。

第一种研究思路发展现状：

对于一般的离散信道，其输入概率分布往往难以得知，因此难以直接用定理计算其信道容量。一般离散信道的信道容量计算可以采用求解线性方程组的方法[2]。信道容量就是在固定信道的条件下，寻找所有可能的输入概率分布 P(x)下平均互信息的极大值，在这里也是最大值。对该极大值的求解大多是运用拉格朗日乘子法。该方法是求解信道容量的一般算法，已经发展相对成熟，人工手算便可完成，但需要进行反复检验运算。

第二种研究思路发展现状：

对几种无噪信道、对称信道和可逆矩阵信道可以利用信道矩阵来计算信道容量。文献[3]针对 MIMO 系统模型，使用信道特征值分析方法给出了 MIMO 信道容量公式，并分析了在信道状态信息(CSI)未知和 CSI 已知这两种情况下的信道容量，并对 MIMO 信道容量进行仿真，验证了天线数目和平均信噪比对信道容量的影响。文献[4]根据相关联矩阵和均匀角分布分析了相关信道的信道容量，并进一步得出了计算信道容量的一般公式，该公式表明相关性的增加会导致信噪比的降低，并且接收天线阵列的半径或扩散角是决定信道容量的最主要因素。

第三种研究思路发展现状：

由于用求解线性方程组法所求出的最佳输入分布不一定是完备集，而且在运用该方法过程中，要求解非齐次线性方程组是较为困难的。即使求解出结果，也无法保证输入符号的概率都为非负值。因此求解的过程中必须进行反复试算，这就使得运算量大大提高。为了解决如此巨大运算量的困难，运用计算机进行运算的迭代算法应运而生。信道容量的迭代算法于 1972 年由 S.Arimoto 和 R.E.Blahut 提出。它是一种原理简单但切实可行的数值算法，能在满足任意给定精度的条件

下，在有限的迭代次数内计算出任意离散无记忆信道的信道容量。文献[5]给出了离散信道的信道容量线性乘法迭代以及线性常系数迭代两种算法，这两种算法都比现有的指数迭代算法高效，在所有单步迭代算法中它们可以说是最好的算法。文献[6]提出了一种近似计算离散无记忆信道信道容量的迭代算法，该信道的输入分配可能存在附加约束，并且基于凸规划二元性导出了信道容量的确切上下限，相比于 Blahut-Arimoto 算法的复杂度 $O(MN^2)$，该算法的复杂度降低为 $O(MN)$，该算法的另一优点是它提供了后验误差，在经过一定次数的迭代后就能准确算出当前的近似误差，这对实际计算而言十分重要。

未来的发展趋势主要集中在以下几点：第一，对 MIMO 系统信道容量的研究将继续成为热点，这是由于 MIMO 能在不增加传输功率和信号带宽的情况下显著增加信道容量。虽然在该方面已经开展了不少研究工作，并且知道了天线数目及信噪比都会影响信道容量的大小，但在实际应用中仍有待改进；第二，信道容量的上界和下界将会持续被突破，在某些情况下信道容量将会趋于无限大。拓宽信道容量的上下界是提高信道容量的根本方法，研究影响上下界的因素对提高信道容量十分有帮助；第三，香农的信息理论仍然是计算信息量和信道容量的基础，但更多更有效的方法将会持续出现，计算机将更广泛地参与到计算任务中。

作为通信系统设计和性能度量的基础，信息量和信道容量的计算是一个复杂且极具挑战性的课题。为了系统归纳各种条件下的信息量和信道容量计算方法，仍需要继续进行大量的研究工作。

第二章　离散信源与信道

通信系统一般由信源、信道和信宿三个部分组成（如图 2.1）。信源就是信息的发端，信道是传递信号的媒介或通道，信宿是消息传送的收端。信源和信宿可处于不同地点和不同时刻。单符号离散信源是指具有一定概率分布的离散符号的集合，单符号离散信道是最简单的离散信道。

图 2.1 通信系统模型

本章首先讨论离散信源和离散信道的数学模型及其统计特性，接着对信源的信息度量及信道传输的平均互信息进行定量研究，在此基础上，研究信道容量的计算方法，最后利用 MATLAB 完成仿真计算。

2.1　单符号离散信源与信道的数学模型

信源是产生消息和消息序列的源头。信息不是消息自身，但它又包含在消息之中，消息是信息的携带者。从本质上来说，信道是传递消息的通道，它能实现信息的传输和存储，是通信系统的重要组成部分。一般而言，通信过程可以简单地看成信源发出适合于信道传输的携带有信息的消息，通过信道传输，在接收端被信宿接收并进行分析处理。

2.1.1　单符号离散信源的数学模型

离散信源是由含有不确定性的消息所组合而成的集合。若这样的离散集合是由有限或无限可列个取值离散的符号（如数字、字母、文字等）组成，则称这样的符号发出集为离散信源。其中，每个消息含有一个符号的单符号离散信源是最基本的离散信源。

在通信系统中，信源发出的消息在被信宿接收之前，都是含有不确定性的随机变量，这样的不确定性本质上取决于信源的概率空间，因此我们可以用一维的离散型随机变量 X 来描述离散信源的输出。在这里，要声明的是，本文的研究对

象不是信源的内部结构，因此不关注信源为何及如何产生各种不同的、可能的消息，而只关注信源可能输出的变量及其不确定性。离散信源发出的每一个离散符号都具有随机性，都含有一定量的信息，但这样的随机性不是毫无规律可言的，而是遵循一定的概率分布来发出每一个符号的。因此，可以用一个样本空间及其概率分布——概率空间来表示信源的数学模型。

若信源 X 可能发出 r 种不同的消息，每个消息都只含有一个符号，则可以用 a_1，a_2，$a_3\cdots$，a_r来表示信源所有可能输出，其中 r 是有限的整数，对应的先验概率分别是$\mathrm{p}(a_1)$，$\mathrm{p}(a_2)$，$\mathrm{p}(a_3)\cdots\mathrm{p}(a_r)$，那么离散信源的概率空间表示为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_r \\ P(a_1) & P(a_2) & \cdots P(a_r) \end{bmatrix} \text{（2-1）}$$

其中

$$0 \le \mathrm{p}(a_i) \le 1 \quad \text{（i = 1，2，…，r）（2-2）}$$

$$\sum_{i=1}^{r} p(a_i) = 1 \text{（2-3）}$$

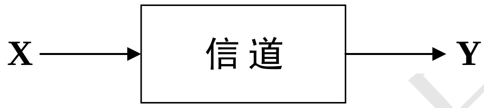
式（2-2）和（2-3）表示信源发出的可能的消息（符号）数 r 是有限的，并且每次的消息输出必定都是从这 r 个里选取一个。并且信源的概率分布是一个完备集。这样的单符号离散信源是最简单的离散信源。

信源与概率空间是一一对应的关系，不同信源所对应的概率空间不同。如果信源给定，其概率空间也就是确定的；反之，如果概率空间给定，则相应的信源也就确定。也就是说，离散信源的统计特性可以用概率空间来表征，因此概率空间也称为信源空间。要用信源空间表征信源的数学模型，必须满足信源符号集所对应的概率集先验可知或事先可被测定这样的前提。概率可测是香农信息论的基本前提条件。

若无论时间起点如何选取，信源发出的随机序列 X 的各维概率分布都不变，则称这样的信源为离散平稳信源[7]。如离散化平面灰度图像就是离散平稳信源。进一步，如果在不同时刻，离散平稳信源发出的符号X_i之间没有依赖，相互独立，则称为离散无记忆信源；相反地，若不同时刻信源发出的符号X_i之间相互关联，则这样的信源称为有记忆信源。表述有记忆信源比表述无记忆信源困难许多，实际的研究分析往往限制随机序列的记忆长度，这样的信源称为马尔可夫信源。无记忆信源的表述涉及复杂的条件概率，本文将不对其进行研究，而重点分析无记忆信源。

2.1.2 单符号离散信道的数学模型

在一般广义的通信系统中，信道是很重要的组成部分。信道的任务是对信源发出的消息进行传送和存储，消息在信道中是以信号的形式存在的。若信道的输入和输出随机序列的取值都是离散的，则称这样的信道为离散信道或数字信道。单符号离散信道是最简单的离散信道。图 2.2 是离散信道的数学模型。



$\mathbf{X}=(X_1,X_2,\cdots,X_i,\cdots,X_N)$ P(y|x) $\mathbf{Y}=(Y_1,Y_2,\cdots,Y_i,\cdots,Y_N)$

X:$[a_1,a_2,\cdots,a_r]$ Y:$[b_1,b_2,\cdots,b_s]$

$\sum_y P(y|x)=1$

图 2.2 离散信道的数学模型

信道中往往存在由噪声引起的随机干扰作用，信源 X 发出的信号经过信道传输后将会产生失真，使得信道的输出端出现输出符号集 Y 中的哪一个符号成为不确定的随机事件，所以信道的输入和输出符号之间的函数关系一般不是确定的，而只能用统计依赖关系来表征。因此，在确定信道特性时需要知道信道的输入信号、输出信号，以及此二者之间的统计依赖关系[11]。在此可以用条件概率 P(Y|X) 来表示输入与输出信号之间的统计依赖关系，条件概率体现了信道的传递特性，也称为信道的传递概率。由(r × s)个信道的传递概率和信道的输入符号集 X:$[a_1,a_2,\cdots,a_r]$、输出符号集 Y:$[b_1,b_2,\cdots,b_s]$构成的信道矩阵[P]，完整地描述了单符号离散信道的传递特性。因此，离散信道的数学模型除用图 2.2 表示外，还可以用信道的传递概率矩阵进行描述。

$$[\,P\,]=\begin{array}{c} \\ a_1 \\ a_2 \\ \vdots \\ a_r \end{array}\begin{array}{c} \begin{array}{cccc} b_1 & b_2 & \dots & b_s \end{array} \\ \begin{bmatrix} p(b_1|a_1) & p(b_2|a_1) & \cdots & p(b_s|a_1) \\ p(b_1|a_2) & p(b_2|a_2) & \cdots & p(b_s|a_2) \\ \vdots & \vdots & & \vdots \\ p(b_1|a_r) & p(b_2|a_r) & \cdots & p(b_s|a_r) \end{bmatrix} \end{array} \tag{2-4}$$

其中

$0\le \mathrm{p}\left(b_j|a_i\right)\le 1$ （i = 1，2， ⋯ ，r; j = 1，2， ⋯ ，s）（2-5）

$\sum_{j=1}^{s} p\left(b_j|a_i\right)=1\ (i=1，2，\cdots，r)$（2-6）

式（2-5）表示传递概率具有一般概率的性质，即信道矩阵[P]中的每一个元素都处于[0,1]区间；式（2-6）表示信道输入某信号$a_i(i=1,2,\cdots,r)$的前提下，即

使由于噪声干扰，信道输出的符号是不确定的，但一定是信道输出符号集 Y：$[b_1, b_2, \cdots, b_s]$中的某一种符号，而不可能是符号集 Y 之外的任何符号，即信道矩阵的每行元素之和均等于 1。

信道矩阵[P]所描述的信道传递特性，也可以用信道传递图形象直观地表示出来。最简单的二元对称信道（BSC）的信道传递图如图 2.3 所示。

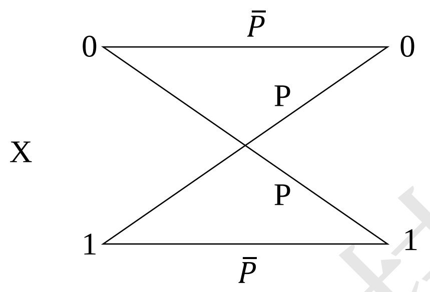


图 2.3　BSC 的信道传递图

根据信道传递概率 P(y|x)的不同，离散信道可做出以下三类划分：

- 无干扰信道（无噪信道）。无干扰信道中没有噪声或者噪声非常小可以忽略，输入信号 X 与输出信号 Y 之间的关系是确定的。条件概率满足
  $\mathrm{P(y|x)} = \begin{cases} 1 & y = f(x) \\ 0 & y \neq f(x) \end{cases}$（2-7）
- 有干扰无记忆信道。有干扰信道由于存在噪声，输入信号与输出信号之间没有确定的对应关系，因此信道的传递概率没有特殊性，只是一般的概率分布。无记忆信道是指信道在任意时刻的输出信号只统计依赖于对应时刻的输入信号，而与非对应时刻的输入信号及其它任何时刻的输出信号都没有关联。本章及下一章的研究关注的包含离散无记忆信道。
- 有干扰有记忆信道。有干扰有记忆信道在现实中更为常见。有记忆信道在某一时刻的输出信号不仅与对应时刻及过去时刻的输入信号有关，还与过去时刻的输出信号有关。对有记忆信道进行分析时，可以将信道近似为无记忆信道，或者采用马尔可夫链的联合条件概率的方法，这些方法都十分复杂，这里不对其进行研究。

另外，在此还要讨论一种特殊的单符号离散信道——串联信道。在实际通信系统中串联信道被广泛使用。例如，为了实现数据处理，常常需要对信道进行串接。微波中继接力通信就是一种串联信道，它是卫星到地面接收站的离散信道与

判决器处理系统的离散信道两种信道的串联。

设单符号离散信道 1 的输入变量为 X:{ $a_1, a_2, \cdots, a_r$ }，输出变量为 Y:{$b_1, b_2, \cdots, b_s$}，信道转移概率 P(y|x)=P($b_j|a_i$)；单符号离散信道 2 的输入变量为 Y，输出变量为 Z:{$z_1, z_2, \cdots, z_t$}，信道转移概率 P(z|xy)=P($c_k|a_i b_j$)，将这两个信道串接起来就成为串联信道。串联信道可以等价为一个总的离散信道，此信道的转移概率可以表示为$P(z|x) = \sum_Y P(y|x) \bullet P(z|xy)$　x∈X，y∈Y，z∈Z。如图 2.4 是串联信道的数学模型。

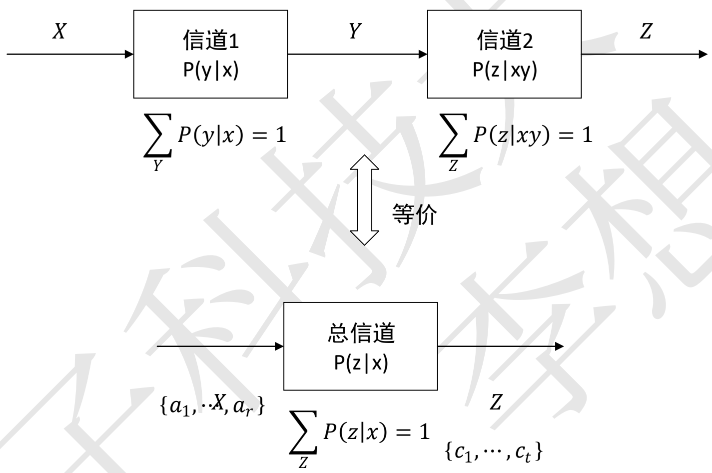


图 2.4 串联信道数学模型

2.2 单符号离散信源与信道的信息量

为了表征信息的多少，就必须对信息加以量化。目前信息尚且没有统一的度量方法，统一通用的方法还比较困难。信息论之父香农于 1948 年发表的著名论文《通信的数学理论》为信息论奠定了坚实的理论基础。他指出，任何信息都存在冗余，冗余大小与信息中每个符号出现的概率即不确定性有关，他把信息中排除了冗余后的平均信息量称为“信息熵”，并进一步研究了计算信息熵的数学方法。

不足的是，香农计算信息量的方法只适用于通信领域，在其它领域不能使用。

我国学者钟义信教授对信息做了如下三种分类。

- 语法信息。语法信息是信息的最抽象最基本的层次，它不涉及事件本身的含义和效用，而只关注事物运动时可能出现的所有状态及状态之间的关系。

香农对信息的定义正是属于这一类别，是从概率统计角度来对信息进行度量，因此属于基于概率性的语法信息，这种信息是能够用数学方法简单地度量出来的。语法信息能较好地解决通信系统中的信息传输问题。

- 语义信息：语义信息是事物运动状态和方式的具体意义，即信息的具体含义。
- 语用信息：语用信息是事物运动状态、方式及其含义对收信者的效用，它关注的是信息中存在的主观价值。语用信息所包含的信息量对于不同主体是不一样的，这已经超出了通信领域的范畴。

由于目前人们对信息本质的认识还不够完整统一，对信息的定义和测度不够深入，因此文中所讨论与计算的信息量主要是基于香农的信息理论。据此，信息量可以直观地定义为：收到某消息获得的信息量等于不确定性的减少量。本节将从信源的自信息、信息熵（平均自信息）以及信道的平均互信息三方面来展开论述。

2.2.1 单符号离散信源的自信息

在信息传输的一般情况下，信宿获得的信息量应等于通信前后不确定性的减少量。而事件发生的不确定性大小是与事件发生的概率高低有关的。事件发生的概率越小，收端猜测它是否发生的困难程度就越大，对信源存在的不确定性也就越大。那么，自信息量可以定义为事件 a_i 发生所包含的信息量，故自信息量是先验概率的函数，表示为：

$$\mathrm{I}(a_i) = \log\frac{1}{P(a_i)} \quad (2\text{-}8)$$

自信息$\mathrm{I}(a_i)$有两种含义：

- 在事件 a_i 发生之前，$\mathrm{I}(a_i)$表示事件 a_i 发生的不确定性；
- 在事件 a_i 发生之后，$\mathrm{I}(a_i)$表示事件 a_i 所提供的信息量。

自信息的单位与对数所选取的底有关，底数为 2 时单位为比特（bit）,底数为 e 时单位为奈特（nat），底数为 10 时单位为哈特（hart）。本文所采用的底数为 2，信息量的单位为比特，为了书写简洁，把底数“2”略去不写。需要注意的是，此处的“比特”是信息量的单位，与计算机学科中的“比特”没有关联。

2.2.2　单符号离散信源的信息熵

自信息是用于度量信源中某一事件的信息量的指标，但由于事件 a_i 是以 $P(a_i)$概率出现的随机事件，其信息量$\mathrm{I}(a_i)$也就是以$P(a_i)$为概率的随机变量，并且，信源发出的某一具体消息不能代表整个信源可能发出的所有消息，故自信息不能作为总体信源的信息度量指标。

为了表征整个信源，将自信息的数学期望定义为信源的平均自信息量是可行的，该平均值又称为信息熵。表示为：

$$\mathrm{H(X)} = \mathrm{E}\left[log\frac{1}{P(a_i)}\right] = -\sum_{i=1}^{r} P(a_i)logP(a_i)\text{（比特/符号）}\quad（2\text{-}9）$$

信源的信息熵是统计平均的总体信源的信息测度，是由 r 个符号 a_i (i=1，2，•••，r)组成的信源 X 的对应的 r 个随机的自信息量$\mathrm{I}(a_i)(\mathrm{i} = 1，2，\cdots，\mathrm{r})$在信源 X 的概率空间$\left[p(a_1)，p(a_2)，\cdots，p(a_r)\right]$中的统计平均值。对于特定的概率空间确定的信源，其信源熵是一个确定的量。

信息熵 H(X)有三种含义：

- 信源 X 发送符号之前，H(X)表示每一个符号存在的平均不确定性；
- 信源 X 发送符号之后，H(X)表示每一个符号所能提供的最大平均信息量；
- 收端能准确无误地收到信源 X 发送的每一个符号时，收端每收到一个符号所能获取的平均信息量。

信源熵有一些重要的基本性质：1）对称性；2）确定性；3）非负性；4）扩展性；5）可加性；6）强可加性；7）递增性；8）极值性。其中，极值性也称为最大离散熵定理，该定理表明，在含有 r 个信源符号的所有信源中，等概分布的信源的平均不确定性最大，即最大熵值为 logr。

2.2.3　单符号离散信道的平均互信息

信源的信息熵解决了定量估算信源每发一个符号提供的平均信息量，这是研究信源的核心问题。但对于通信系统来说，最根本的问题是信息传输问题，即如何定量估算信宿在收到经信道传输的消息后能从中能获取多少信息量，并根据信息量的大小来选择信道的问题。而信道所传输的信息量在数量上等于通信前后信宿对信源存在的不确定性的消除量，即输入随机变量 X 与输出随机变量 Y 之间的

互信息量。因此信息传输问题首先要解决的是 X 和 Y 两个随机变量之间的互信息的计算问题，最后再根据互信息量来选择最适合传输的信道。

当信道为无噪信道，此时信源与信宿之间不存在干扰，信宿能正确地接收到信源发出的消息，通信前对信源存在的所有不确定性将被消除，此时信源所发消息包含的全部信息量将被信宿完全接收。因此，无噪信道的信息传输首要问题可以看作是信源信息熵的计算问题。一般情况下，信道为有噪信道，噪声的随机干扰时时存在。消息通过有噪信道传输后往往发生了变形，即产生了错误和失真。信宿在收到加噪的失真的消息后，将无法彻底消除对信源存在的不确定性。根据定义，可以得出以下结论。

互信息量=通信前、后不确定性的减少量=先验不确定性－后验不确定性

$$= \log\frac{1}{\text{先验概率}} - log\frac{1}{\text{后验概率}} = log\frac{\text{后验概率}}{\text{先验概率}}$$

$$\mathrm{I}\left(a_i;b_j\right) = \mathrm{I}(a_i) - \mathrm{I}\left(a_i|b_j\right)$$

$$= \log\frac{1}{p(a_i)} - \log\frac{1}{p\left(a_i|b_j\right)} = \log\frac{p\left(a_i|b_j\right)}{p(a_i)} \quad (\mathrm{i} = 1,2,\cdots,\mathrm{s}) \text{（2-10）}$$

互信息能度量信源出现某个特定符号、信宿出现某个特定符号这一随机事件，它本身也是随机变量。对于 r 个输入符号、s 个输出符号的信道来说，存在 r×s 个互信息量$\mathrm{I}\left(a_i;b_j\right)(\mathrm{i} = 1,2,\cdots,\mathrm{s};\mathrm{j} = 1,2,\cdots,\mathrm{s})$，与自信息类似，这(r×s)个随机互信息量都不能从总体上度量信道传输的信息量。只有互信息量的统计平均值才能全面地度量信道传输信息量的大小，该值称为平均互信息量。

$$\begin{aligned}\mathrm{I}(\mathrm{X};\mathrm{Y}) &= \sum_{i=1}^{r}\sum_{j=1}^{s} p\left(a_i b_j\right)\mathrm{I}\left(a_i;b_j\right)\\ &= \mathrm{H}(\mathrm{X}) - \mathrm{H}(\mathrm{X}|\mathrm{Y})\\ &= \mathrm{H}(Y) - \mathrm{H}(\mathrm{Y}|\mathrm{X})\\ &= \mathrm{H}(\mathrm{X}) + \mathrm{H}(\mathrm{Y}) - \mathrm{H}(\mathrm{XY}) \text{（2-11）}\end{aligned}$$

平均互信息是从整体上衡量信道每传递一个信号所传输的平均信息量的，表征通信前后平均不确定性的消除，也表示输入与输出随机变量之间的统计约束程度。它是随机变量$\mathrm{I}\left(a_i;b_j\right)$在联合概率空间 P(XY)中的统计平均值，是一个确定量。

平均互信息有一些重要的基本性质：1）非负性；2）极值性；3）对称性；4）凸状性。其中，凸状性描述的是平均互信息 I(X;Y)是信源输入的概率分布 P(x)的

上凸函数，这意味着，当固定信道时，选择不同的信源（其概率分布不同）与信道相接，在信道的输出端收到每个符号后所获得的信息量是不同的。而且对于固定信道，总存在着一种信源，使得输出端获得的平均互信息量最大。

特别地，对于串联信道，其平均互信息的关系式 I(XY;Z)≥I(Y;Z)和 I(XY;Z)≥I(X;Z)正是信号处理定理，表明信号通过串联信道传输后只会丢失更多的信息，最多保持原来获得的信息。信息不增性原理表明，在信息传输系统中，一旦在某一环节丢失了一些信息，则在后面的环节中都无法恢复已经丢失的信息，也就是说，对于一系列的串接信道，H(X)≥I(X;Y)≥I(X;Z)≥I(X;W)≥…。

2.3 单符号离散信道的信道容量

研究信道的目的是要讨论信息传输问题，关注的是信道的信息传输率 R，即信道中每个消息符号所能传送的平均信息量。平均互信息 I(X;Y)即信道传输率可以表示如下

$$\mathrm{R} = \mathrm{I(X;Y)} = \mathrm{H(X)} - \mathrm{H(X|Y)} \quad (比特/符号) \text{（2-12）}$$

由上一节对平均互信息的描述可知，对每一个固定的信道，总可以找到一种信源(某种概率分布 P(x))，使得信道传输每个符号平均所能获得的信息量最大。换句话说，每个固定的信道都对应一个最大的信息传输率，即信道容量 C。

$$\mathrm{C} = \max_{P(x)}\{I(X;Y)\} \quad (比特/符号) \text{（2-13）}$$

当平均互信息量等于信道容量时，相应的输入概率分布 P(x)称为最佳输入分布。需要注意的是，信道容量是信道传递概率的函数，与信源输入分布没有关联，只是在信源为匹配信源时平均互信息量才能达到信道容量。因此，信道容量可以作为信道的特征参量之一。

2.4 MATLAB 实现

MATLAB 是美国 Math Works 公司出品的商业数学软件，是当今最优秀的数学、工程仿真工具之一。MATLAB 语言通俗易懂，编写简洁高效，广泛应用与通信、信号处理及科学计算等领域。本节将使用 MATLAB 完成单符号离散信源与信道的信息量与信道容量的仿真计算，并对计算结果进行简要分析。

2.4.1　信息量的计算

(一)自信息的计算

单符号离散信源的自信息的计算，依据其定义$\mathrm{I}(a_i)=\log\frac{1}{P(a_i)}=-logP(a_i)$便可完成。图 2.5 是计算单符号离散信源的自信息量结果。

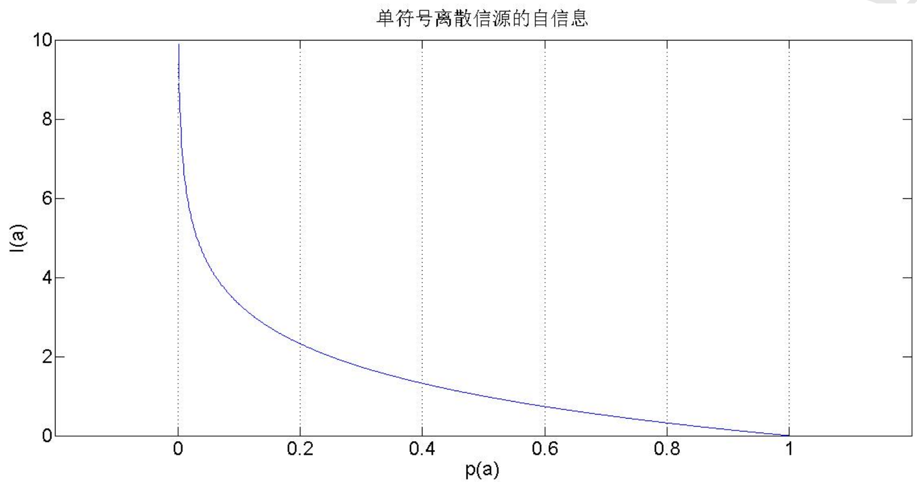


图 2.5 单符号离散信源的自信息与输入概率的关系图

从图 2.5 运行结果来看，自信息量与信源输入符号出现概率呈负相关，事件发生的概率越大则该事件发生所含有的信息量越小。当某一信号出现的概率为 1，即该事件必然会发生，则此时的自信息量为 0，表示为 i）事件发生以前，该事件发生的不确定性为 0；ii）事件发生以后，信源无法提供任何信息量。相反地，当某一信号出现的概率为 0，即该事件必然不发生，则此时的自信息量为无穷大，表示为 i）事件发生以前，该事件发生的不确定性为无穷大；ii）事件发生以后，信源所提供的信息量为无穷大。

(二)信源熵的计算

信源的信息熵计算的基本公式为$\mathrm{H(X)}=-\sum_{i=1}^{r}P(a_i)logP(a_i)$。依据信源的输入概率分布计算出自信息的统计平均值，即为信源熵。其中信源的输入概率分布是由已知条件直接给出或者间接分析得出的，输入概率分布必须满足完备性。

图 2.6 给出了计算最简单的单符号离散信源，即二元信源的信息熵的运行结果图。二元信源发出的消息符号只有 0 和 1 两个，其中一个符号的出现概率为 p 则另一符号的出现概率为（1-p）。二元数字是二元信源的输出。从图 2.6 可以得出，如果二元信源的输出是确定的（即 p 为 0 或 1），则该信源不提供任何信息。

反之，当二元信源输出的两个消息符号等概发生时，信源熵达到最大，此时输出的二元数字序列中，每个二元数字平均提供的信息量为 1 比特。

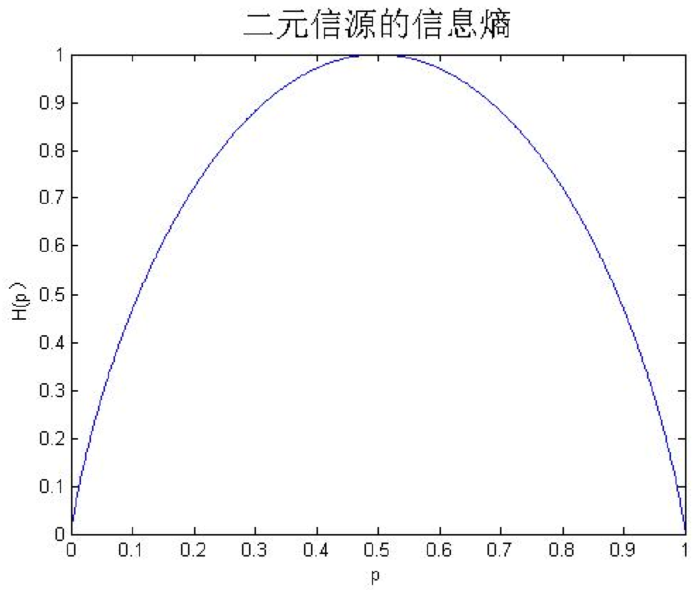


图 2.6 二元信源的信息熵与信源输入概率分布的关系

在计算离散信源时，首先需要判断输入信源的概率分布是否完备，再逐一对自信息进行加权求和。

例如，设某地的天气预报为：晴（可能性 1/2）、阴（可能性 1/4）、雨（可能性 1/8）、雪（可能性 1/8），则信源空间为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 晴 & 阴 & 雨 & 雪 \\ 1/2 & 1/4 & 1/8 & 1/8 \end{bmatrix}$$

则输入概率分布为 p=[1/2,1/4,1/8,1/8]。运行结果如下图 2.7 所示，即该地天气预报所能提供的平均信息量即信源的信息熵为 1.75 比特。

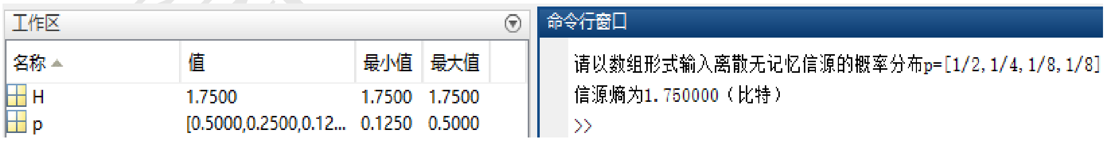


图 2.7 某地天气预报提供的信息量

(三)平均互信息的计算

平均互信息定义为信宿接收到输出符号后平均每个符号所获得的关于输入变量 X 的信息量。由式（2-11）可知$\mathrm{I(X;Y)=H(X)-H(X|Y)}$，而通常情况下已知的量是信源输入分布与信道转移概率$\mathrm{P}(y_j|x_i)$，因此无法直接计算后验的条件熵

H(X|Y)。

为了计算后验条件熵 H(X|Y)，需要由全概率公式求出 P(y)，由乘法公式求出联合概率 P(xy)，再根据条件概率公式求出后验概率 P(x|y)，进而得到后验条件熵。最后在前面计算过信源熵的基础上，由式（2-11）计算得出平均互信息。

例如，设信源 X 的信源空间为

$$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ \frac{1}{3} & \frac{2}{3} \end{bmatrix}$$

信源与图 2.8 所示的二元删除信道（BEC）相接。则信道转移概率矩阵为 $P=\begin{bmatrix} 3/4 & 1/4 & 0 \\ 0 & 1/2 & 1/2 \end{bmatrix}$。运行结果如图 2.9 所示。

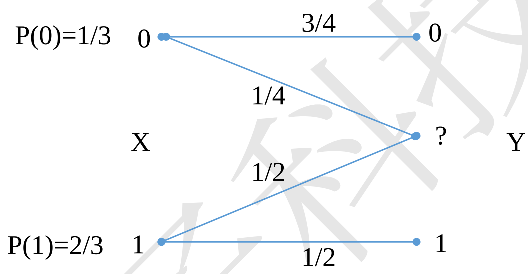


图 2.8 二元删除信道

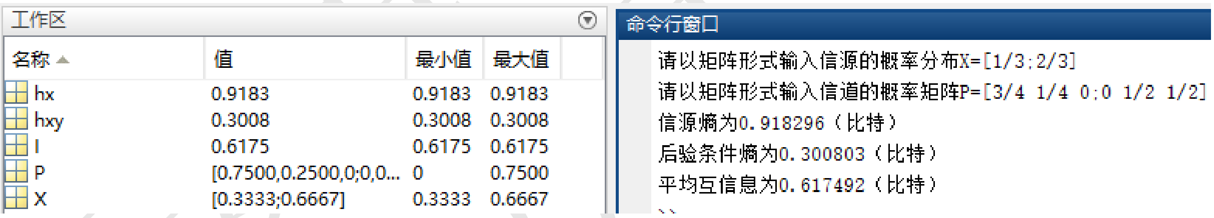


图 2.9 平均互信息计算结果

2.4.2 信道容量的计算

对于一般信道，要计算其信道容量是相当复杂的。从数学上来说，就是对平均互信息 I(X;Y)求极大值的问题。

(一)二元对称信道的信道容量

对于简单的二元对称信道，平均互信息是信道传递概率的下凸函数，极大值为 I(X;Y)＝1－H(p)。因此，二元对称信道的信道容量为

C＝1－H(p)（2-14）

图 2.10 是二元对称信道的信道容量计算的运行结果图。从图 2.10 可以看出，二元对称信道的信道容量只是信道传输概率的函数，不同的二元对称信道其信道容量不同。当信道传输概率 p 为 1/2 时，即信道有一半的可能传对信号，一半的可能传错信号，此时信道所传输的消息对接收端而言将没有任何意义，接收端无法从中获取信息量，因此此时的信道容量为零。

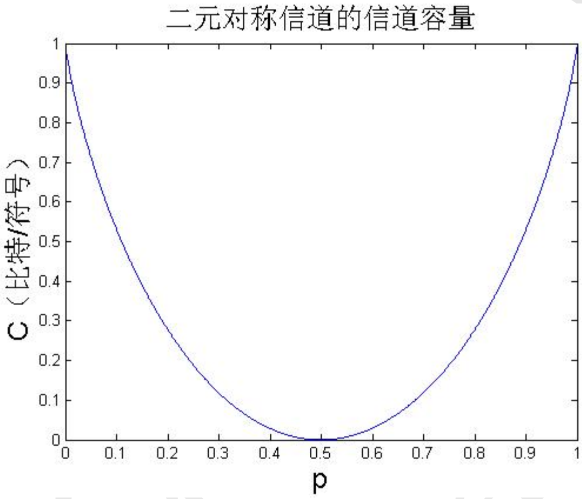


图 2.10 二元对称信道的信道容量

(二)几种特殊信道的信道容量

1) 无噪信道

单符号离散无噪信道，是特殊信道中最基本并且最简单的一种。由于信号经过无噪信道传输后不确定性没有增加，因此信宿接收到消息后，每个符号平均所能获得的信息量就等于信源发出的每个符号所能提供的信息量，即符号包含的信息量。因此，无噪信道的信息传输首要问题可等价成信源熵问题。进而，无噪信道的信道容量问题，也就可以等价为信源熵的最大值问题。

- 无噪无损信道：离散无噪无损信道的输入和输出符号之间的关系是一一对应的，其信道矩阵是 $r\times r$ 阶单位矩阵。这类信道的噪声熵 $H(Y|X)$ 和损失熵 $H(X|Y)$ 都为 0，平均互信息为 $I(X;Y)=H(X)=H(Y)$。当输入信源等概时（匹配信源），信息传输率达到信道容量

$$C=\log r \quad (2\text{-}15)$$

图 2.11 是 3 阶无噪无损信道的信道容量的计算结果截图。

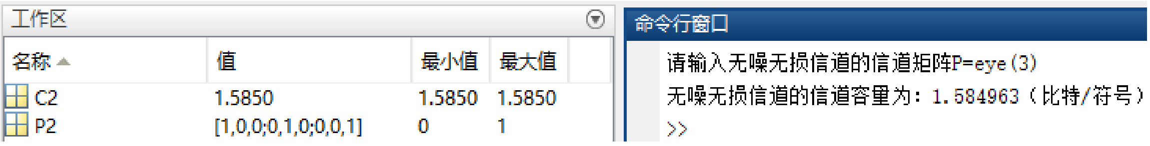


图 2.11 无噪无损信道的信道容量计算结果

- 无噪有损信道：无噪有损信道的输出是输入的确定函数，因而先验概率 P(y|x)等于 0 或 1，噪声熵为 0。但输入输出不是一一对应关系，而是多对一，因而后验概率 P(x|y)不等于 0 或 1，损失熵不等于 0。其平均互信息 I(X;Y)=H(Y)<H(X)，其信道容量为

  C=max(H(Y))=logs（比特/符号）（2-16）

  匹配信源是使输出 Y 达到等概分布的任何输入随机变量。

  图 2.12 是无噪有损信道的信道容量测试结果图。

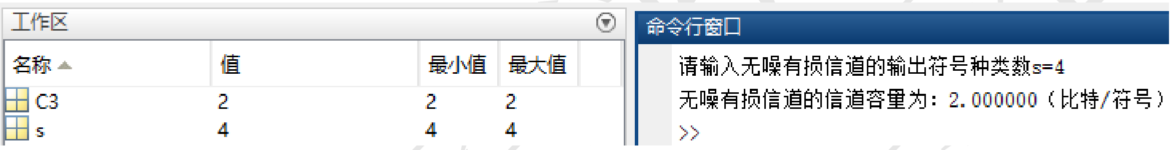


图 2.12 无噪有损信道的信道容量计算结果

2) 对称信道

单符号离散对称信道是一种常用的特殊信道。其特点是信道矩阵具有很强的对称性。

- 强对称离散信道：强对称信道的输入符号集 X：$\{a_1,a_2,\cdots,a_r\}$，输出符号集 Y：$\{a_1,a_2,\cdots,a_r\}$，两个集合相同，每一输入符号的正确传递概率为（1－ε），总的错误传递概率 ε 均匀分配在其它（r－1）个错误传递概率上。其信道容量为

  C=logr－H(ε)－εlog(r－1)（2-17）

  匹配信源是 r 元等概信源。图 2.13 是测试的强对称离散信道信道容量的计算结果截图。

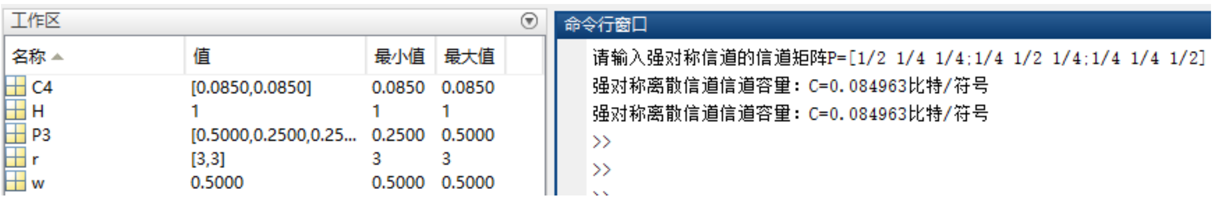


图 2.13 强对称离散信道的信道容量计算结果

- 对称离散信道：对称离散信道的每一行都是另一行的转置，每一列也都是另一列的转置，但与强对称信道不同，其信道矩阵不一定是（r×r）阶方阵，r 与 s 不一定相等，并且，其行集合与列集合不一定是同一集合。因此对称离散信道的信道矩阵不一定是对称矩阵。设信道矩阵[P]的行集合{p}:$\{p_1, p_2, \cdots, p_s\}$含有 s 个元素、列集合{q}:$\{q_1, q_2, \cdots, q_r\}$含有 r 个元素，其信道容量为

$$C = \log s - H(p_1, p_2, \cdots, p_s) \qquad (2\text{-}18)$$

匹配信源是 r 元等概信源。

例如，若对称离散信道的信道矩阵为$P = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 1/6 & 1/2 & 1/3 \\ 1/3 & 1/6 & 1/2 \end{bmatrix}$，代入进行计算后的结果如图 2.13 所示。

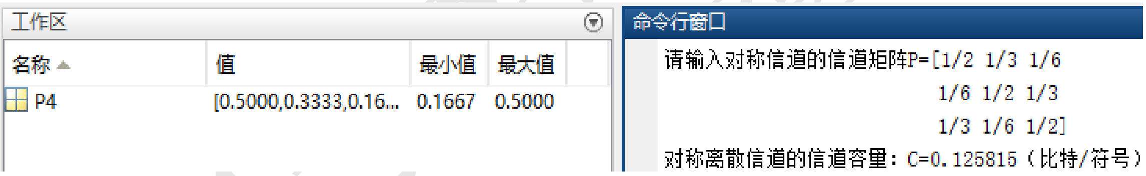


图 2.13 对称离散信道的信道容量计算结果

- 准对称离散信道：准对称离散信道的信道矩阵中的每一行都是其它行的转置，但列方向不具有转置特性，将其分为几个不相交的子集后，各子矩阵都具有行和列的转置特性。准对称信道的信道容量的计算涉及子矩阵的划分，这在 MATLAB 中难以实现，因此此处本文采用的是原始定义的方法，即先求出信源熵和噪声熵，二者相减求出平均互信息，再对其取最大值。准对称信道的匹配信源是 r 元等概信源。此处本文仅对二元信源及其所接信道进行分析计算，更高元的计算读者可以依次类推。

例如，若信道输入符号集为 X:{0,1}，输出符号集为 Y:{0,1,2,3}准对称信道的信道矩阵为$P = \begin{bmatrix} 1/2 & 1/4 & 1/8 & 1/8 \\ 1/4 & 1/2 & 1/8 & 1/8 \end{bmatrix}$，代入进行计算后的结果如图 2.14 和图 2.15 所示。图 2.14 给出了信道容量的计算结果数值，图 2.15 给出了平均互信息与信源输出概率分布的关系，由图中可知，当信源等概分布即输出符号 0 和 1 的概率都是 0.5 时，平均互信息达到最大值 0.06128 比特，此值也就

是信道容量的大小。

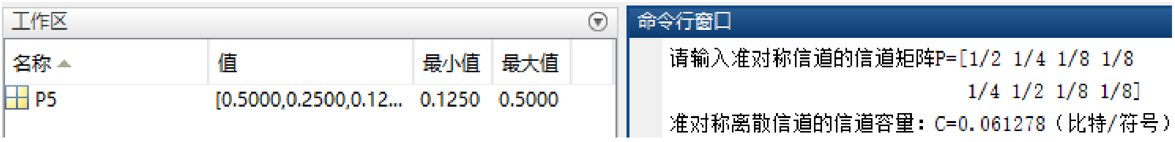


图 2.14 准对称离散信道的信道容量计算结果

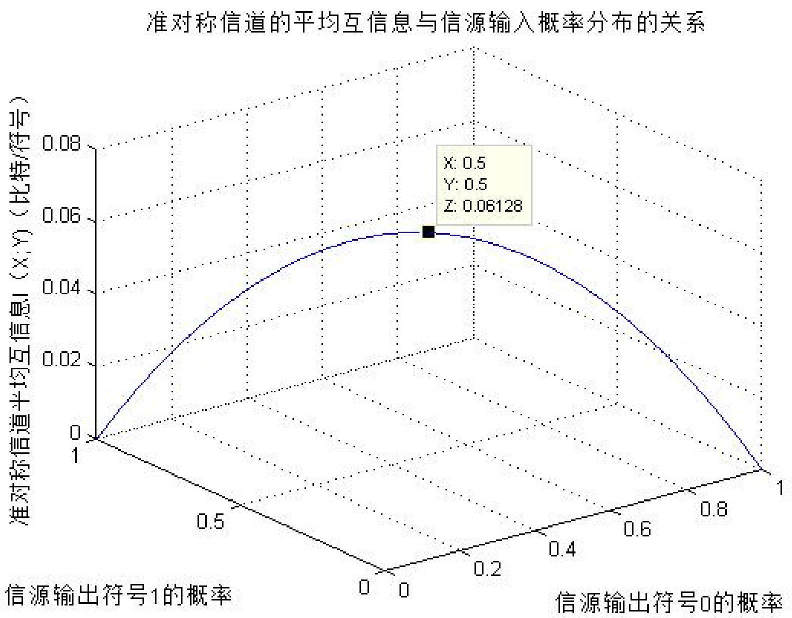


图 2.15 准对称信道的平均互信息

(三)一般离散信道的信道容量

对于一般离散信道来说，计算其信道容量是一项相当复杂的工作。信道容量的迭代算法是一种常用的近似算法，它能以任意给定的精度在有限的步数内算出任意离散无记忆信道的信道容量。迭代算法的原理是，反向传递概率可由输入概率、信道转移概率经贝叶斯定理得到，信道容量是平均互信息的极大值，而平均互信息是输入概率分布和反向传递概率分布的函数[8]。因此，可以交替地以输入概率分布和反向传递概率分布为自变量，使平均互信息最大化。可以证明，信道容量迭代计算的逐级逼近是收敛的，当迭代次数 n 足够大时，可以近似认为 C(n+1,n)＝C。如图 2.16 是信道容量的迭代算的运算流程图。

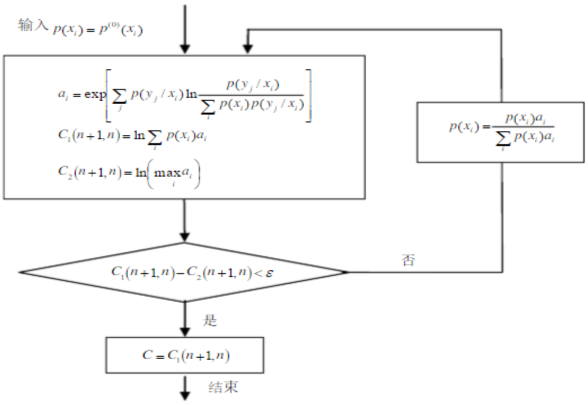


图 2.16 信道容量的迭代算法流程图

例如，若要求信道容限即可容忍的误差为 0.0001，测试不同信道转移概率矩阵所对应的信道容量，测试结果如下图 2.17 所示。

```
>> e=0.0001;
>> P6=[1 0 0;0 1/2 1/2;0 1/2 1/2];
>> channel_dd(P6,e);
迭代次数：n=2
信道容量：C=1.000000（比特/符号）
>>
>> P6=[1 0 0;0 1 0;0 0 1];
>> channel_dd(P6,e);
迭代次数：n=1
信道容量：C=1.584963（比特/符号）
>>
>> P6=[1/2 1/2  0 0;0 1/2 1/2 0;0 0 1/2 1/2;1/2 0 0 1/2];
>> channel_dd(P6,e);
迭代次数：n=1
信道容量：C=1.000000（比特/符号）
>>
>>  P6=[0.3 0.2 0.5;0 0.6 0.4;1 0 0];
>> channel_dd(P6,e);
迭代次数：n=6
信道容量：C=21.246601（比特/符号）
```

图 2.17 迭代算法计算信道容量测试结果

(四)串联信道的信道容量

在完成一些特殊信道及一般信道的信道容量的计算后，本章还要讨论串联信道信道容量的计算。根据信道容量的定义，串联信道的信道容量为

$$C_{串}(\mathrm{I},\mathrm{II},\cdots,n)=\max_{P(x)} I(X;X_n) \qquad （2\text{-}19）$$

由上式可知，串联的无源数据处理信道越多，总的信道容量越有可能减小。

当串联的信道数目趋于无穷大时，如数据被经过无穷多次处理，信道容量可能就会趋于 0。以下以简单的二元对称信道的串联信道来说明。

设离散的二元对称信道的输入符号的概率空间为$\begin{bmatrix} X \\ P(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1/2 & 1/2 \end{bmatrix}$，信道矩阵为$P_1 = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$，则两个这样的信道串联后的信道矩阵为$P_2 = P_1 * P_1 = \begin{bmatrix} (1-p)^2+p^2 & 2p(1-p) \\ 2p(1-p) & (1-p)^2+p^2 \end{bmatrix}$，三个这样的信道串联后的信道矩阵为$P_3 = P_2 * P_1 = \begin{bmatrix} (1-p)^3+3(1-p)p^2 & p^3+3p(1-p)^2 \\ p^3+3p(1-p)^2 & (1-p)^3+3(1-p)p^2 \end{bmatrix}$。由前面对二元对称信道信道容量的讨论可知，C＝1-H(p)，其中 p 为传输错误概率。因此用 MATLAB 计算后的结果如下图 2.18 所示。

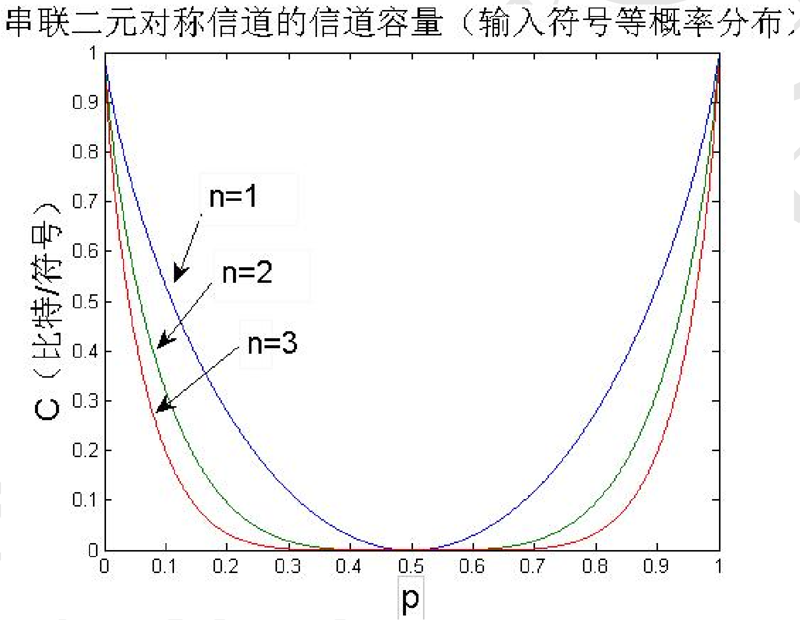


图 2.18 串联二元对称信道的信道容量

在图 2.18 中，当 n=1 时信道即为单个的二元对称信道，n=2 时为两个信道串联后的信道容量，此时信道容量减少了一部分，以此类推。由此表明，串联的二元对称信道相比单个信道，信息的损失会增加，当串联级数越高，信息损失越多。当二元对称信道中的错误概率 p 很小时（如 p=10^{-4}），计算所得的信道容量减少量较小，所以，当串接数很大时，串联信道的信道容量虽然减小，但仍然有相对大的数值。在实际的通信系统中，二元对称信道的错误概率一般在10^{-6}以下，因此经过若干次串联后，信道容量的减少并不明显。

2.5 多符号离散信源与信道

由单符号离散信源与信道所构成的单符号离散通信系统是最基本的通信系统。但实际上，信源发出的消息往往不是单个的信源符号，而是由多个信源符号组成的时间或空间序列构成。这样由多个信源符号构成的序列来代表一个完整消息的信源，称为多符号离散信源。多符号离散信源与多符号离散信道相接，构成多符号离散通信系统。

2.5 多符号离散信源与信道的数学模型

与单符号离散信源相比，多符号离散信源 X 不只在某一时刻发出符号，而是随着时间的推移，每一个单位时间连续不断地发出符号，构成信源符号的时间序列，每一个序列代表一条消息。多符号离散信道在每一个单位时间中的所有 N 个时刻，都相应地对输入消息进行了传输，它是在单位时间中运行了 N 次的扩展信道。

2.5.1 多符号离散信源的数学模型

多符号离散信源 **X** 在每一个单位时间中，以一定的概率$\mathrm{p}(a_{il})(\mathrm{l}=1,2,3,\cdots)$发出信源符号集 X：$\{a_1, a_2, \cdots, a_r\}$中某一符号，随着时间推移，在时间域上形成无限长的随机符号序列。因此，多符号离散信源可表示为 **X**=$X_1X_2\cdots X_N\cdots$。若多符号离散信源的 **X**= $X_1X_2\cdots X_N\cdots$的各维联合概率分布都不随时间的推移发生变化，并且与时间起点无关，则该多符号离散信源 **X** 为 N 维离散平稳信源。因此，从概率空间来说，由无限长的随机变量和时间序列表示的多符号离散平稳信源可由有限长 N 的随机变量时间时间序列 **X**= $X_1X_2\cdots X_N$来表示，该序列组成的信源称为离散平稳无记忆信源 **X** 的 N 次扩展信源。

信源 X 的 N 的次扩展信源$\mathbf{X}^N$，是由q^N个符号组成的离散信源，其 N 重概率空间为

$$\begin{bmatrix} X^N \\ P(\alpha_i) \end{bmatrix} = \begin{bmatrix} \alpha_{1,} & \alpha_{2,} & \cdots, & \alpha_{q^N} \\ P(\alpha_1) & P(\alpha_2) & \cdots, & P(\alpha_{q^N}) \end{bmatrix} \quad (3\text{-}1)$$

式中每个符号α_i是对应于某一个由 N 个a_i组成的序列，$P(\alpha_i)$对应的是由 N

个a_i组成的的序列的概率。

2.5.2 多符号离散信道的数学模型

离散平稳信源 X 的 N 次扩展信源 $\mathbf{X}=X_1X_2\cdots X_N$发出的消息通过信道(X-Y)的过程，可以做如下分析。从运行机制上来看，在时刻t_1，随机变量X_1通过信道，对应输出随机变量Y_1；在时刻t_2，随机变量X_2通过信道，对应输出随机变量Y_2；⋯；在时刻t_N，随机变量X_N通过信道，对应输出随机变量Y_N。在 N 个单位时刻，信道(X-Y)连续运作了 N 次。从整体上来看，输入一个由 N 个随机变量组成的随机矢量 $\mathbf{X}=X_1X_2\cdots X_N$，相应地输出一个由 N 个随机变量组成的随机矢量 $\mathbf{Y}=Y_1Y_2\cdots Y_N$。

若把信道(X-Y)的 N 次扩展信道(**X-Y**)看成一个信道，则它的数学模型可表示为图 3.1。在扩展信道中，信道传递概率已由信道(X-Y)的随机变量之间的条件概率 P(Y|X)扩展为随机矢量之间的条件概率 P(**Y**|**X**)=$P(Y_1Y_2\cdots Y_N|X_1X_2\cdots X_N)$，相应的信道矩阵$[P]_N$也由[P]的(r×s)阶扩展为$(r^N\times S^N)$阶。

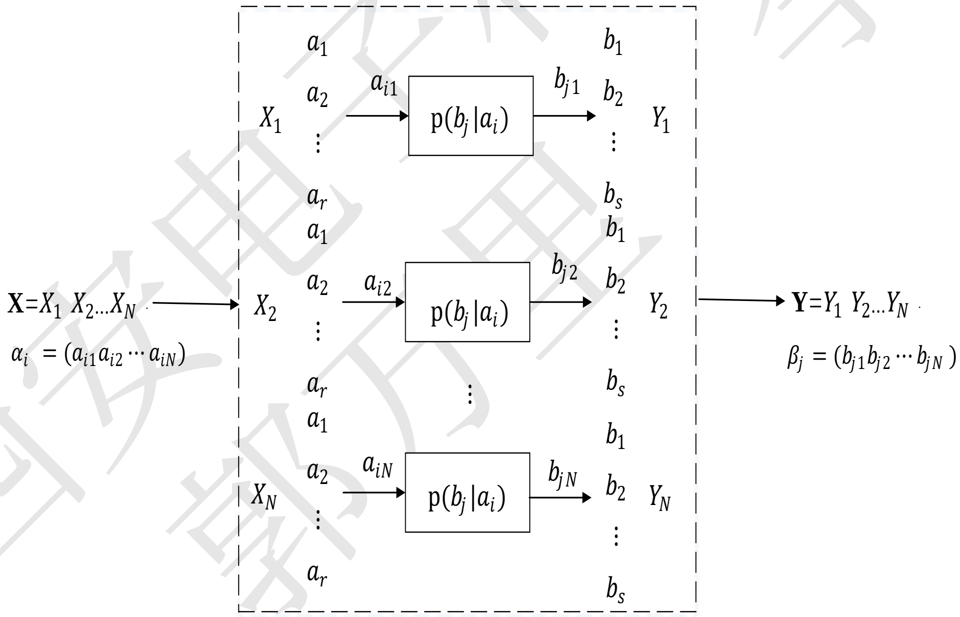


图 3.1N 次扩展信道的数学模型

只要知道扩展的次数 N，就可由单符号离散信道的 X 和 Y 得出 N 次扩展信道的 **X** 和 **Y**。但在一般情况下，无法由单符号离散信道的转移概率得到 N 次扩展信道的转移概率，而只能用测定的方法得到。

另外，对于扩展信道，若其传递概率 P(**Y**|**X**)等于 N 个时刻单符号离散信道（X-Y）的传递概率的连乘，则该信道称为单符号离散无记忆信道(X-Y)的 N 次扩

展信道。

单符号离散信道(X-Y)的 N 次扩展信道(**X-Y**)是同一个单符号离散信道(X-Y)在 N 个单位时间内相继运行 N 次所形成的总体信道。独立并联信道是指由 N 个信道组成，这 N 个信道的输入随机变量X_1，X_2，$\cdots X_N$构成并联信道的输入随机矢量 $\mathbf{X}=X_1X_2\cdots X_N$，N 个信道的输出随机变量$Y_1$，$Y_2$，$\cdots Y_N$构成并联信道的输出随机矢量 $\mathbf{Y}=Y_1Y_2\cdots Y_N$。独立并联信道的数学模型如下图 3.2 所示。

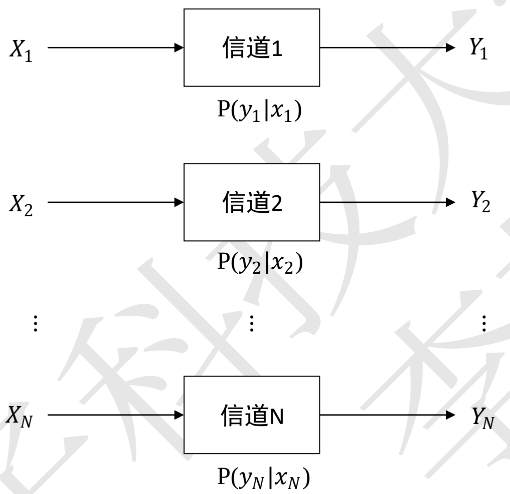


图 3.2 独立并联信道的数学模型

由N个不相同的信道构成的独立并联信道与无记忆信道的N次扩展信道之间的根本区别在于，独立并列信道是由 N 个输入符号集、输出符号集和转移概率都可能是互不相同的信道构成的，而无记忆信道的 N 次扩展信道是同一信道在 N 个单位时间相继运行 N 次，可以等价为由 N 个输入符号集、输出符号集和传递概率都相同的信道构成的。此二者也有相同之处，它们的信道转移概率 P(**Y**|**X**)=P($Y_1Y_2\cdots Y_N|X_1X_2\cdots X_N$)都等于 N 个信道自身的转移概率的连乘。由此可见，当 N 个信道的输入符号集、输出符号集和信道转移概率都相同时，由它们构成的独立并联信道和无记忆信道的 N 次扩展信道是等价的。也就是说，在信源为离散平稳信源的条件下，离散无记忆信道的 N 次扩展信道可以看成是由 N 个相同信道构成的独立并联信道的特例。

2.6 多符号离散信源与信道的信息量

对于多符号离散信源，关注的是离散平稳无记忆信源，而不对离散平稳有记

忆信源及非平稳信源进行研究，这是因为其信息量与信道容量的计算都涉及多维条件概率，具体的计算十分复杂，目前对此二者的分析大多还是研究其性质。多符号离散信源与信道的信息量与信道容量的计算都与单符号离散信源与信道有关。

2.6.1 离散平稳无记忆信源的信息熵

若平稳信源 X 的 N 次扩展信源 **X** 在各个时刻发出的随机变量之间相互统计独立，则信源 X 称为离散平稳无记忆信源，信源 **X** 称为离散平稳无记忆信源 X 的 N 次扩展信源。根据信息熵的定义可以证明离散无记忆信源 X 的 N 次扩展信源的熵等于离散信源 X 的熵的 N 倍，即

$$\mathrm{H}(\mathbf{X}) = \mathrm{H}(X^N) = \mathrm{NH}(\mathrm{X}) \text{（3-2）}$$

该结论表明，离散平稳无记忆信源 X 的 N 次扩展信源 **X** 每发一条信息，所提供的平均信息量等于离散平稳无记忆信源 X 每发送一个符号所提供的平均信息量的 N 倍。因此，可以利用该关系来计算离散平稳无记忆信源 **X** 的信息熵。

需要注意的是，扩展信源的信源熵的单位应是扩展信源每发送一条信息提供的平均信息量，扩展次数 N 代表每条消息的信源符号组成数，因此扩展信源的信源熵的单位是（比特/符号（N 信符）），N 为确切的扩展次数。

2.6.2 N 次扩展信道的平均互信息量

离散无记忆信道的 N 次扩展信道的平均互信息量 I(**X;Y**)，小于或等于信源 X 各时刻通过无记忆信道的平均互信息量之和。只有当信源是离散无记忆信源 X 的 N 次扩展信源 **X**=$X_1X_2\cdots X_N$时，二者才相等。

离散无记忆信源 X 的 N 次扩展信源 **X**=$X_1X_2\cdots X_N$通过扩展信道的平均互信息量 I(**X;Y**)，大于或等于各时刻信源 **X**=$X_1X_2\cdots X_N$的随机变量X_k(k=1,2, $\cdots, N$)相继通过信道的平均互信息量 I($X_k;Y_k$)(k=1,2, $\cdots, N$)之和。只有当信道是无记忆信道的 N 次扩展信道时，二者才相等。

总之，只有当信源是离散无记忆信源，信道是离散无记忆信道时，通过扩展信道的平均互信息量 I(**X;Y**)才等于 N 次扩展信源 **X**=$X_1X_2\cdots X_N$在 N 个时刻分别相继通过无记忆信道的平均互信息量之和。即

$$\mathrm{I}(X^N;Y^N) = \sum_{k=1}^{N} I(X_k;Y_k) \text{（3-3）}$$

2.6.3 独立并联信道的平均互信息量

由 N 个信道组成的独立并联信道的联合平均互信息量 I(**X;Y**) = $I(X_1X_2\cdots X_N;Y_1Y_2\cdots Y_N)$，小于或等于各信道自身的平均互信息量之和。只有当输入随机变量X_1, X_2，$\cdots X_N$之间统计独立时，二者才相等。

输入随机变量之间统计独立的独立并联信道的平均互信息量为

$I(X_1X_2\cdots X_N;Y_1Y_2\cdots Y_N)=\sum_{k=1}^{N}I(X_k;Y_k)$（3-4）

2.7 多符号离散信道的信道容量

2.7.1 无记忆扩展信道的信道容量

对离散无记忆信道来说，每传递一个长度为 N 的消息所携带的平均互信息量的最大值，即为 N 次扩展信道的信道容量。设无记忆信道(X-Y)的匹配信源为离散无记忆信源 X，其信道容量为C_0，则信道(X-Y)的 N 次扩展信道(**X-Y**)的信道容量C_N是C_0的 N 倍，即

$C_N=NC_0$（3-5）

此时扩展信道的匹配信源为信源 X 的 N 次扩展信源 **X**=$X_1X_2\cdots X_N$，并且信源 X 是离散无记忆信道本身的匹配信源，这个信源在 N 个单位时间内所发出的符号是相互统计独立的，形成无记忆信源的 N 次扩展信源。独立并联信道的信道容量

上一节讨论了独立并联信道的平均互信息量，由平均互信息量与信道容量的关系，可以推导出独立并联信道的信道容量的计算公式如下

$C_N=\sum_{k=1}^{N}C_k$ （3-6）

即由 N 个信道构成的独立并联信道的联合信道容量C_N，在数值上等于各信道的信道容量C_k(k = 1,2,$\cdots$, N)之和，其匹配信源 **X**=$X_1X_2\cdots X_N$中的随机变量X_k(k = 1,2,$\cdots$, N)是相应组成信道(k)的匹配信源，并且这些信源之间相互统计独立。

2.8 MATLAB 实现

与单符号离散信源与信道相比，多符号离散信源与信道的信息量与信道容量的计算往往要考虑信源的消息长度及信道的扩展次数，并且往往都是建立在对单

符号离散信源与信道的信息量与信道容量的计算的基础上的。

2.8.1 信息量的计算

对于多符号离散信源与信道的信息量，本节关注以下两方面，分别是信源熵与平均互信息的计算。其中，对于信源熵，只讨论无记忆信源，有记忆信源的信源熵和极限熵较为复杂，本文不做讨论。

(一)信源熵的计算

离散平稳无记忆信源 X 的 N 次扩展信源 $\mathbf{X}=X_1X_2\cdots X_N$的信息熵 H($\mathbf{X}$)，是离散平稳无记忆信源 X 的信息熵 H(X)的 N 倍。因此，为了计算扩展信源的信息熵 H($\mathbf{X}$)，需要先计算出单符号离散信源的信息熵 H(X)，后者在第二章已介绍了其计算方法。

例如，当离散平稳无记忆信源 X 的输入概率分布为$\begin{bmatrix} X \\ P(X) \end{bmatrix}=\begin{bmatrix} a_1 & a_2 & a_3 \\ 1/4 & 1/2 & 1/4 \end{bmatrix}$，则信源 X 的 N=2 次扩展信源的信息熵的计算结果如图 3.3 所示。

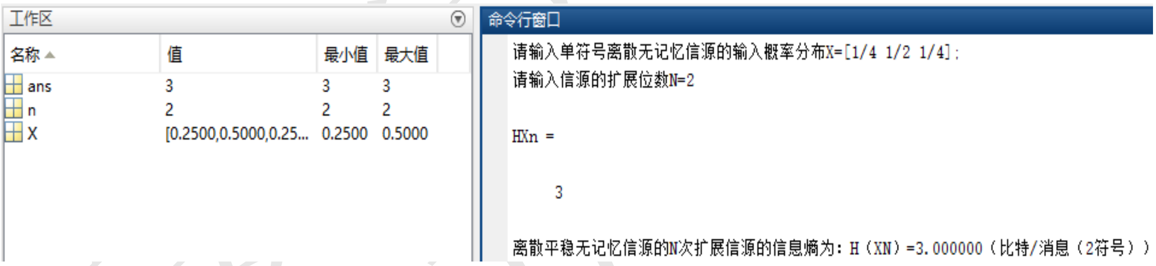


图 3.3 离散平稳无记忆信源的 N 次扩展信源的信息熵计算结果

(二)平均互信息的计算

以下将从离散无记忆信道的 N 次扩展信道的和独立并联信道两方面来实现平均互信息量的计算。

- N 次扩展信道的平均互信息

对于离散无记忆信道的 N 次扩展信道，当且仅当输入信源是离散无记忆信源 X 的 N 次扩展信源、信道是无记忆信道的 N 次扩展信道时，平均互信息量为 N 次扩展信源中每一时刻的随机变量相继单独通过无记忆信道的平均互信息量之和。

假设N次扩展信源**X**在各个时刻的输入概率分布都一样，即为离散平稳无记忆信源，其输入概率分布为X=[1/4;1/2;1/4]，且离散无记忆信道的信道矩阵为P =

$\begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \end{bmatrix}$，信道的扩展次数为N，则平均互信息量为每个时刻的平均互信息量的N倍，而每个时刻的平均互信息可以由I＝H(X)－H(X|Y)求得。计算结果如图3.4所示。

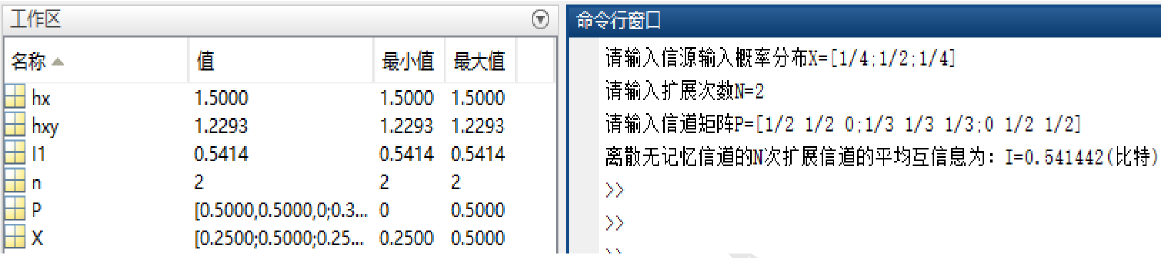


图 3.4 N 次扩展信道的平均互信息量计算结果

- 独立并联信道的平均互信息

由 3.2 节中对独立并联信道的平均互信息的介绍，可以知道 $I(X_1X_2\cdots X_N;Y_1Y_2\cdots Y_N)\le\sum_{k=1}^{N}I(X_k;Y_k)$，当且仅当输入随机变量$X_1X_2\cdots X_N$之间统计独立时，等式成立。

假设输入随机变量之间统计独立，并联的信道数目为 3，依次在 MATLAB 的命令行窗口输入相应的信源概率分布和信道转移概率矩阵，分别求出各个信道的平均互信息，并进行累加，最后结果即为并联信道的平均互信息。测试及计算结果如图 3.5 所示。

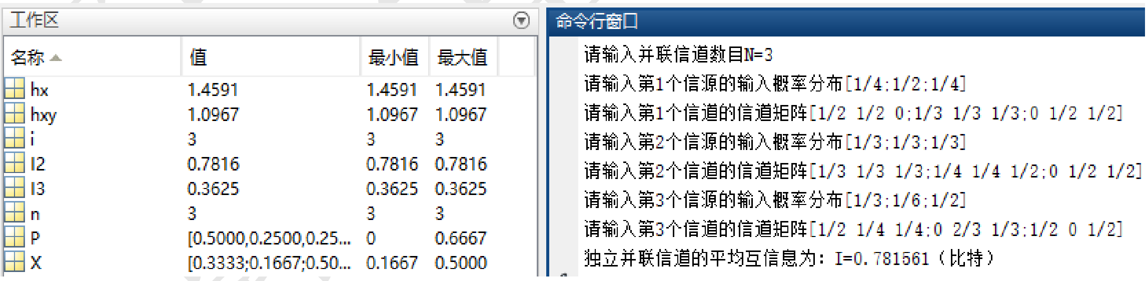


图 3.5 独立并联信道的平均互信息计算结果

2.8.2 信道容量的计算

(一)无记忆扩展信道的信道容量

由上一节中对无记忆扩展信道的信道容量的介绍，可以知道，离散无记忆信道的 N 次扩展信道的信道容量C_N，是离散无记忆信道本身容量C_0的 N 倍。因此，在计算无记忆扩展信道的信道容量时，可以先使用计算一般离散信道的信道容量

的迭代算法，计算出离散无记忆信道的信道容量C_0，再将其乘以 N，就可以得到扩展信道的信道容量C_N。

例如，若离散无记忆信道 X 的信道转移矩阵为$P=\begin{bmatrix}1/2 & 1/4 & 1/4\\ 1/4 & 1/2 & 1/4\\ 1/4 & 1/4 & 1/2\end{bmatrix}$，将其进行 N=2 次扩展，则扩展后的信道的信道容量值及相应变量的值如图 3.6 所示。

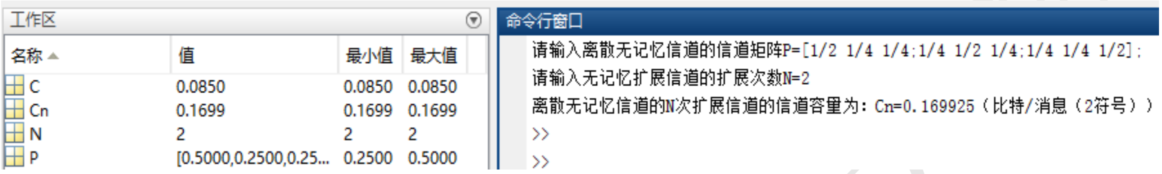


图 3.6 无记忆扩展信道的信道容量计算结果

(二)独立并联信道的信道容量

由上一节对独立并联信道及其信道容量的介绍，可知当输入符号X_i的概率分布为各信道的匹配信源分布时，独立并联信道的信道容量等于各组成信道的信道容量之和。利用这一结论，可以利用迭代算法先求出各组成信道的信道容量再将其进行叠加求和，就可以得到并联信道的信道容量。

例如，假设独立并联信道由信道 1 与信道 2 并联构成，随机变量X_1与X_2统计独立，它们是两个组成信道的输入变量，信源空间分别为

$$\begin{bmatrix}X_1\\ P(X_1)\end{bmatrix}=\begin{bmatrix}0 & 1\\ \frac{1}{2} & \frac{1}{2}\end{bmatrix}$$

$$\begin{bmatrix}X_2\\ P(X_2)\end{bmatrix}=\begin{bmatrix}YES(y) & NO(n)\\ 1/2 & 1/2\end{bmatrix}$$

信道矩阵分别为$P_1=\begin{bmatrix}1/2 & 1/2\\ 1/4 & 3/4\end{bmatrix}$和$P_2=\begin{bmatrix}2/3 & 1/3\\ 0 & 1\end{bmatrix}$。经过计算，该并联信道的信道容量值及相应变量值如图 3.7 所示。

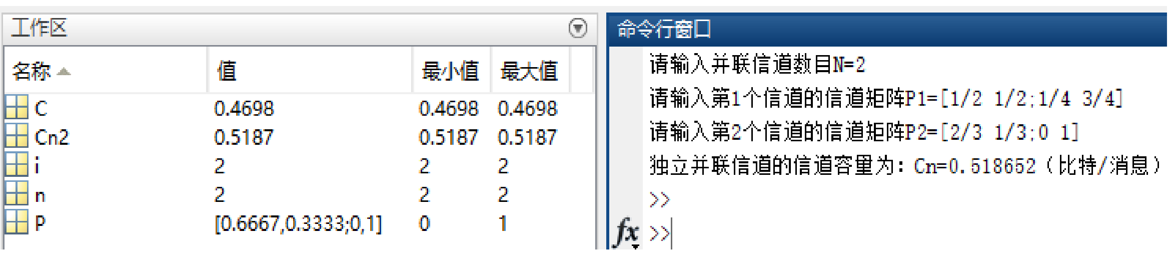


图 3.7 独立并联信道的信道容量计算结果

第三章　单维连续信源与信道

在前面第二章和第三章中，研究对象都是取值为有限或可数的离散信源，其输出消息是时间上离散、取值为有限或可数的随机序列。而实际应用中，有些信源的输出消息常常在时间 t 和取值上都是连续的随机函数 x(t),这样的信源为波形信源，如语音信号 X(t)、电视信号 X(x_0, y_0, t)等。对波形信源通过取样可以变换成连续信源来处理。当信道的输入和输出都是模拟信号时，该信道称为波形信道或模拟信道。波形信道在时间上离散化后就转化成时间离散、取值连续的的连续信道。

本章将在介绍单维连续信源与信道的数学模型的基础上，分析其信息量与信道容量的计算方法，并用 MATLAB 完成其仿真计算。

3.1　单维连续信源与信道的数学模型

3.1.1　单维连续信源的数学模型

若信源的输出消息是由单个符号组成的，每个符号的可能取值是随机且无限不可数的，即输出消息的取值是连续的区间或实数集（-∞，∞）。这种信源十分常见，如语音信号、热噪声信号等。这种信源可以用一维的连续型随机变量 X 来描述，称为连续信源。由一维随机变量组成的单维连续信源 X 是最简单、最基本且最重要的一类连续信源。

单维连续信源可以由一个取值连续的随机变量 X 及其出现概率表示。设 X 的取值区间为$[a,b]$（或实数轴 R），X 在取值区间内的概率密度函数为 p(x)。则单位连续信源 X 的信源空间为

$$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} [a,b]\text{或实数集}R \\ p(x) \end{bmatrix} \qquad (4-1)$$

并满足$\int_a^b p(x)\,dx = 1$ 或$\int_R p(x)\,dx = 1$。这就是描述单维连续信源的数学模型。

3.1.2　单维连续信道的数学模型

输入和输出都是单个连续型随机变量的信道称为单维连续信道，这是最基本的连续信道。假设单维连续信道的输入区间为$[a,b]$或实数域 R、输出区间为$[a',b']$或实数域 R，信道的传递概率密度函数p(y|x)(x ∈ $[a,b], y \in [a',b']$)，对所有的x ∈ $[a,b]$，都有

$\int_{a'}^{b'} p(y|x)\,dy = 1 \qquad x \in [a,b]$（4-2）

则单维连续信道的数学模型可用图 4.1 来表示。

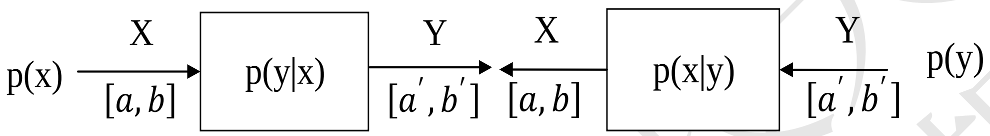


图 4.1 单维连续信道的数学模型图 4.2 单维连续信道的反向信道

结合单维连续信源的概率空间，我们可以知道，因信源 X 的取值区间与信道的输入区间 X:$[a,b]$相同，故信源 X 的任一可能取值 x(a≤x≤b)都能通过信道，以概率密度p(y|x)在信道输出端出现某一值 y(a'≤b')。也就是说，信源 X 在信道的输出端形成连续随机变量 Y:$[a',b']$。

根据概率论的一般理论，在已知单位连续信源 X 的概率密度函数 p(x)和信道的传递概率密度函数p(y|x)的条件下，可以求得p(xy)、p(y)、p(x|y)等概率密度函数。因此图 4.1 所示的“正向信道”可以转换成“反向信道”，如图 4.2 所示。“正向信道”和“反向信道”是同一通信系统的两种不同的表达形式。

3.2　单维连续信源与信道的信息量

3.2.1　单维连续信源的相对熵

为了求连续信源的信源熵，可以利用离散化原则，将连续变量 X 量化分层后用离散变量表示。当量化区间个数 n→∞，量化单位Δ→0，连续信源的信息熵 H(X)就可以由离散信源的熵H(X_n)逼近得到。即

$$\mathrm{H(X)} = \lim_{n\to\infty} H(X_n) = -\lim_{\Delta\to 0}\sum_i p(x_i)\Delta\log p(x_i) - \lim_{\Delta\to 0}(\log\Delta)\sum_i p(x_i)\Delta$$

$$= -\int_a^b p(x)\log p(x)\,dx - \lim_{\Delta\to 0}\log\Delta \text{（4-3）}$$

其中，$[a,b]$是连续信源的取值区间，p(x)为概率密度函数。一般而言，上式第一项积分值是定值，定义为连续信源的相对熵 h(X)（或差熵），第二项极限值是趋于无穷大的常数。所以连续信源 X 的信息熵 H(X)趋于无穷大，等于连续信源 X 的相对熵 h(X)与一个无限大的常数项之和，H(X)就称为连续信源的绝对熵。

$$\mathrm{h(X)} = -\int_a^b p(x)\log p(x)\,dx \qquad (4\text{-}4)$$

绝对熵趋于无穷大的原因是，连续信源 X 在取值区间$[a,b]$中的可能取值是连续的，有无限多个，则信源的不确定性为无穷大。得到输出变量后所能提取的信息量也将为无穷大。因此，单维连续信源 X 的相对熵 h(X)不能用来衡量信源平均不确定性的大小，即不能衡量信源每发送一个符号所包含的平均信息量。相对熵 h(X)不具有信息的内涵。

虽然相对熵不能作为信源的信息测度，但为了与离散信源的熵在形式上统一起来，并且为了计算连续信道的平均互信息量，有必要定义相对熵 h(X)这样一个确定数值的量。

下面根据文献[9]中介绍的连续信源相对熵及其性质来讨论几种信源。

(一)均匀分布的连续信源

设均匀分布的连续信源 X 的取值区间为$[a,b]$，信源空间可以描述为$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} [a,b] \\ p(x) \end{bmatrix}$，其中概率密度函数$\mathrm{p(x)} = \begin{cases} \frac{1}{b-a} (\mathrm{a \le x \le b}) \\ 0 \qquad (x > b, x < a) \end{cases}$。则由式（4-4）可以给出信源 X 的相对熵为

$$\mathrm{h(X) = \log(b-a)} \quad (4\text{-}5)$$

当对数取 2 为底时，相对熵单位为比特/自由度。

该结论表明，均匀分布的连续信源 X 的相对熵 h(X)大小只与取值区间长度有关。与离散信源的信息熵不同，由于连续信源的相对熵 h(X)只是信源熵 H(X)中的定值部分，因此取值不具有非负性。

(二)高斯分布的连续信源

设均值为 m，方差为σ^2的满足高斯分布的连续信源 X 的信源空间为$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} R \\ p(x) \end{bmatrix}$，其中$\mathrm{p(x)} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m^2)}{2\sigma^2}}, -\infty < x < \infty$。则由式（4-4）可以得出信源 X 的相对熵为

$h(X) = \frac{1}{2}\log(2\pi e\sigma^2)$（4-6）

该结论表明，高斯分布的连续信源 X 的相对熵 h(X)只与信源的方差σ^2有关，而与均值无关。当均值为 0 时，方差σ^2就是信源 X 输出的平均功率 P，此时信源相对熵只取决于平均功率 P。

(三)指数分布的连续信源

设指数分布的连续信源 X 的均值为 m，则信源空间为$\begin{bmatrix} X \\ P(X) \end{bmatrix} = \begin{bmatrix} [0,\infty] \\ p(x) \end{bmatrix}$，其中

$p(x) = \begin{cases} \frac{1}{m}e^{-\frac{x}{m}} x \geq 0 \\ 0 \qquad x < 0 \end{cases}$。则由式（4-4）可以得出信源 X 的相对熵为

$h(X) = \log(em)$（4-7）

该结论表明，指数分布的连续信源 X 的相对熵 h(X)只与信源的均值 m 有关。

在讨论相对熵的概念及计算方法后，下面对相对熵的性质进行讨论。

可以证明，在单维连续信源的取值区间$[a,b]$内，相对熵是概率密度函数 p(x)的上凸函数，而且存在最大值。在一般的通信系统中，连续信源 X 的取值区间、平均值和平均功率都可能受到限制，此时的最大相对熵是以概率密度函数 p(x)为自变量的条件极大值。阐明这一性质的是最大相对熵定理。

最大相对熵定理表明：

- 在输出消息的峰值功率受限的单维连续信源中，服从均匀分布的信源的相对熵取值最大，并且最大值仅与限定的峰值功率有关。
- 在输出非负消息且其均值受限的单维连续信源中，服从指数分布的信源的相对熵取值最大，并且最大值仅与限定均值有关。
- 在方差受限的单维连续信源中，满足高斯分布的信源的相对熵取值最大，并且最大值仅与限定的方差有关。
- 在均值为零、平均功率受限的单维连续信源中，满足高斯分布的信源的相对熵取值最大，并且最大值仅与限定的平均功率有关。

3.2.2 单维连续信道的平均互信息

与计算连续信源的信息熵的一样，计算平均互信息时也可以将连续信道量化为离散信道，先计算此离散信道的平均互信息，再将量化间隔Δ向零逼近，就可

以将结果作为连续信道的平均互信息。因此，进行类似的推导后可以得到由单维连续信源 X 和单位连续信道（X-Y）所构建的通信系统中，信源熵 H(X)和信道疑义度 H(X|Y)的值，最后再根据定义得到单维连续信道（X-Y）的平均互信息量

$$\begin{aligned} \mathrm{I(X;Y)} &= \mathrm{H(X)} - \mathrm{H(X|Y)} \\ &= \left[h(X) - \lim_{\Delta\to 0}\log\Delta\right] - \left[h(\mathrm{X|Y}) - \lim_{\Delta\to 0}\log\Delta\right] \\ &= h(X) - h(X|Y) \end{aligned} \quad (4\text{-}8)$$

另外，平均互信息量还有其它两种表达方式：

$$\begin{aligned} \mathrm{I(X;Y)} &= \mathrm{h(Y)} - \mathrm{h(Y|X)} \\ &= h(X) + h(Y) - h(XY) \end{aligned} \quad (4\text{-}9)$$

每一种表达方式都可以用 p(x)、p(y|x)（或由这两个概率密度函数求得的 p(xy)、p(y)、p(x|y)等概率密度函数）的二重积分求得。

式（4-8）表明，单维连续信道的平均互信息量 I(X;Y)，即两个具有信息含义、其值为无限大的信息熵 H(X)和 H(X|Y)之差。也就是说，在单维连续信源与单维连续信道构成的连续通信系统的信息传输问题中，即两个信息熵的“熵差”的问题中，相对熵 h(X)和相对条件熵 h(X|Y)分别取代了信息熵 H(X)和信息条件熵 H(X|Y)的作用。可以发现，计算单维连续信道平均互信息的关系式与离散信道的关系式十分相似，这就是引入相对熵的原因。

总而言之，熵差具有信息度量的意义。

平均互信息有以下重要的性质：1）非负性；2 凸状性。因为连续信道平均互信息量是两个值有限的相对熵之差，其值必为有限值，又有非负性，因此，连续信道平均互信息量是一个非负的有限值。凸状性描述的是平均互信息 I(X;Y)是输入连续信源 X 的概率密度函数 p(x)的上凸函数，这意味着，对于给定信道，通过变动输入连续信源，总可以找到一种概率密度函数，使得连续信道平均互信息量达到极大值，这个极大值就是连续信道平均互信息量的最大值。

3.3 单维连续信道的信道容量

连续信道和离散信道一样，平均互信息 I(X;Y)的最大值即为信道容量，其匹配信源是使 I(X;Y)达到最大值的连续信源 X。信道容量代表信道传输信息的最大能力，是信道性能的重要指标，也是连续信道固有的特征参量。信道容量的取值

与信源无关，只是其表现条件为输入信源是匹配信源。

对于普通的连续信道，其信道容量的计算是较为复杂的。本节重点关注一种特殊的连续信道——高斯加性信道，并计算它的信道容量。

如图 4.3 是加性信道的数学模型。其中 X 是信源输入的连续随机变量，概率密度函数为 p(x)，N 是噪声，也是连续随机变量，概率密度函数为 p(n)，并且信源与噪声之间统计独立，故 p(x n)=p(x)p(n)。输出连续随机变量 Y 是噪声 N 与信源 X 的线性叠加，故 Y=X+N。

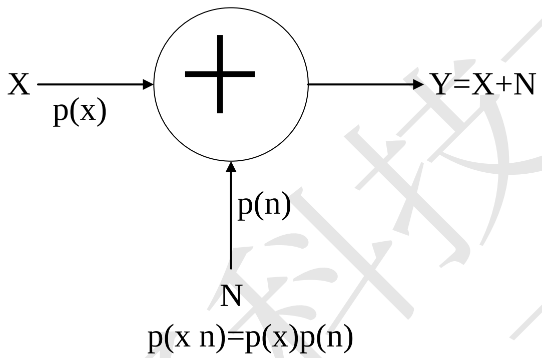


图 4.3 加性信道模型

可以证明，输入信源 X、输出随机变量 Y 之间的联合概率密度函数 p(xy)与输入信源 X、噪声 N 之间的联合概率密度函数 p(xn)相等，由 p(xy)=p(x)p(y|x)及 p(xn)=p(x)p(n)，可以得到

$$\mathrm{p(y|x) = p(n)} \qquad \text{（4-10）}$$

这表明，加性信道的传递概率密度函数 p(y|x)就是噪声 N 的概率密度函数 p(n)，这是加性信道的一个重要特征。概率密度函数的特征必然导致信息熵具有相应的特征。即

$$\mathrm{h(Y|X) = h(N)} \qquad \text{（4-11）}$$

又因为加性信道的输入信源 X 与噪声 N 统计独立，因此，

$$\mathrm{C} = \max_{p(x)}\{h(Y) - h(Y|X)\} = \max_{p(x)}\{h(Y) - h(N)\} = \max_{p(x)}\{h(Y)\} - h(N) \qquad \text{（4-12）}$$

这就是加性信道的信道容量 C 的一般表达式。

由式（4-12），对于实际的高斯加性信道，利用信源的概率密度函数和噪声的概率密度函数，经过推导可以得出信道容量的计算公式。特别地，当噪声 N 是均

值为 0、方差为σ^2的高斯随机变量时，噪声熵为$\mathrm{h(N)} = \log\sqrt{2\pi e\sigma^2}$，高斯加性信道的信道容量

$$\mathrm{C} = \frac{1}{2}\log(1 + \frac{\sigma_X^2}{\sigma^2}) \qquad （4\text{-}13）$$

其匹配信源 X 是均值为零、方差为σ_X^2的高斯连续信源。当信源 X 均值为零时，其方差σ_X^2等于信源 X 的平均功率P_X，是信道对匹配信源 X 的平均功率的限定值，这是信道自身的一个特征参量。对于零均值加性高斯噪声 N，其方差σ^2等于噪声 N 的平均功率P_N。因此式（4-13）又可以表示为

$$\mathrm{C} = \frac{1}{2}\log(1 + \frac{P_X}{P_N}) = \frac{1}{2}\log(1 + \beta) \qquad （4\text{-}14）$$

其中信源平均功率与噪声平均功率的比值 β 称为信道的信噪功率比，简称为信噪比。信道容量仅与高斯加性信道的信噪比 β 有关，因此信噪比是表征信道本身特性的一个重要参量。

对于平均功率受限的非高斯加性信道，要对其信道容量进行精确的计算存在一定难度，只有信道容量的上下限较容易计算得到。其下限正是高斯加性信道的信道容量，因此可以说，当噪声的平均功率给定后，高斯加性信道的信道容量 C 是最小的，高斯噪声是最坏的噪声，高斯加性信道是平均功率受限条件下的最差信道。

3.4 MATLAB 实现

单维连续信源的信息熵为无穷大的数值，等于一个有确定数值的相对熵加上一个无穷大量，因此本节仅对相对熵进行 MATLAB 仿真计算。而相对熵的定义正是为了便于计算“熵差”即平均互信息量，本节对平均互信息量的计算是从定义出发的。最后计算了典型的高斯加性信道的信道容量。

3.4.1 信息量的计算

单维连续信源与信道的信息量的计算涉及相对熵，虽然相对熵不代表信源的平均不确定性，也不代表信源每取一个数值所提供的平均信息量，但相对熵解决了计算平均互信息量时两个无穷大量的相减问题，使之称为两个有确定数值的有限量之差，即“熵差”问题。

(一)信源的相对熵

下面根据 4.2 节中讨论的几种单维连续信源来计算其相对熵。

- 均匀分布的连续信源

对取值区间为$[a,b]$的满足均匀分布的单维连续信源，由式（4-5）可知其相对熵为$\mathrm{h(X)=log(b-a)}$。因此只要知道其取值区间，就能计算出信源的相对熵。

例如，若均匀分布的连续信源 X 的取值区间为[-10,10]，则信源的相对熵计算结果如图 4.4 所示。

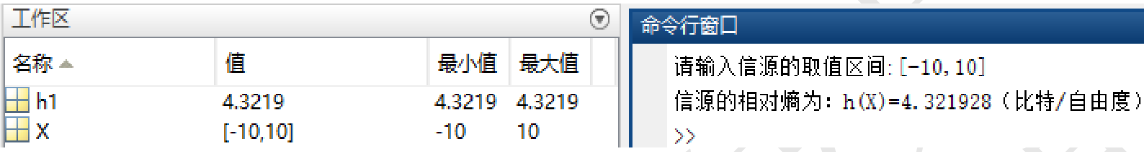


图 4.4 均匀分布的连续信源的相对熵计算结果

- 高斯分布的连续信源

对均值为 m、方差为σ^2的满足高斯分布的单维连续信源，由式（4-6）可知其相对熵为$\mathrm{h(X)}=\frac{1}{2}\log(2\pi e\sigma^2)$。因此只要知道信源的方差，就能计算出信源的相对熵。

例如，若高斯分布的连续信源 X 的方差为 4，则信源的相对熵计算结果如图 4.5 所示。

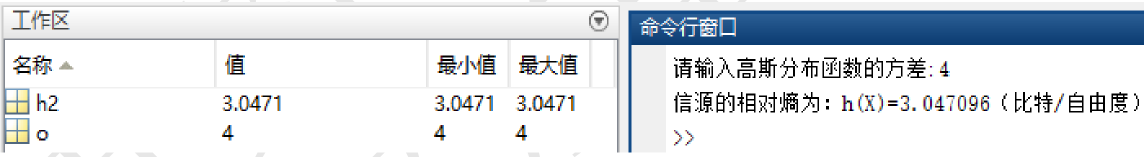


图 4.5 高斯分布的连续信源的相对熵计算结果

- 指数分布的连续信源

对均值为 m 的满足指数分布的连续信源，由式（4-7）可知其相对熵为 $\mathrm{h(X)}=\log(em)$。因此只要知道信源的均值，就能计算出信源的相对熵。

例如，若指数分布的连续信源 X 的均值为 3，则信源的相对熵计算结果如图 4.6 所示。

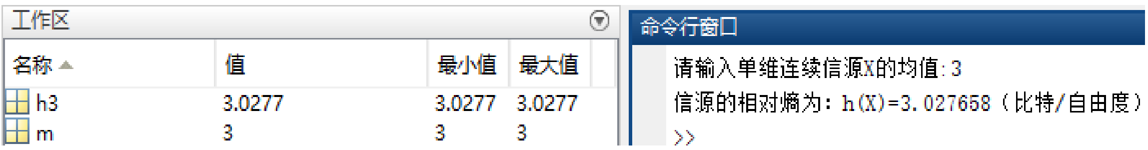


图 4.6 指数分布的连续信源的相对熵计算结果

(二)平均互信息量

单维连续信道的平均互信息量可以用三种形式的“熵差”来表示，每种表示形式都可以用概率密度函数和条件概率密度函数的积分来求得。这里运用式(4-15)来计算。首先需要根据给定的输入随机变量 X 的概率密度函数 p(x)和信道的传递概率密度函数 p(y|x)，来求得输入与输出之间的联合概率密度函数 p(xy)。再对 p(xy)在 x 的取值范围内积分得到输出随机变量 Y 的概率密度函数 p(y)。最后根据式（4-15）进行二重积分，就可以得到输入 X 与输出 Y 之间的平均互信息量。

$$\mathrm{I(X;Y)} = \int_a^b \int_c^d p(xy) \log \frac{p(y|x)}{p(y)} dx\, dy \qquad (4\text{-}15)$$

其中，[a,b]是输入随机变量 X 的取值范围，[c,d]是输出随机变量 Y 的取值范围。

例如，有一信源发出恒定宽度、不同幅度的脉冲 x，幅度值的取值范围为 $[\mathrm{a,b}] = [-4,4]$。此信源与某信道相接，脉冲经过信道传输后，接收端接收脉冲 y 的幅度取值范围为$[\mathrm{c,d}] = [-5,5]$。已知信道的传递概率密度函数为$\mathrm{p(y|x)} = \frac{1}{d-c}$。分析可以知道输入变量为满足均匀分布的信源，$\mathrm{p(x)} = \frac{1}{b-a} = \frac{1}{8}$，输出也是满足均匀分布的随机变量 Y，$\mathrm{p(y)} = \frac{1}{d-c} = \frac{1}{10}$，信道传递概率密度函数$\mathrm{p(y|x)} = \frac{1}{10}$。在 MATLAB 中进行运算后结果如图 4.7 所示。求得的平均互信息量为 0，也就是说，脉冲信号经过该信道传输后，接收端获得输出消息后，无法从输出消息中得到任何关于输入的信息量。

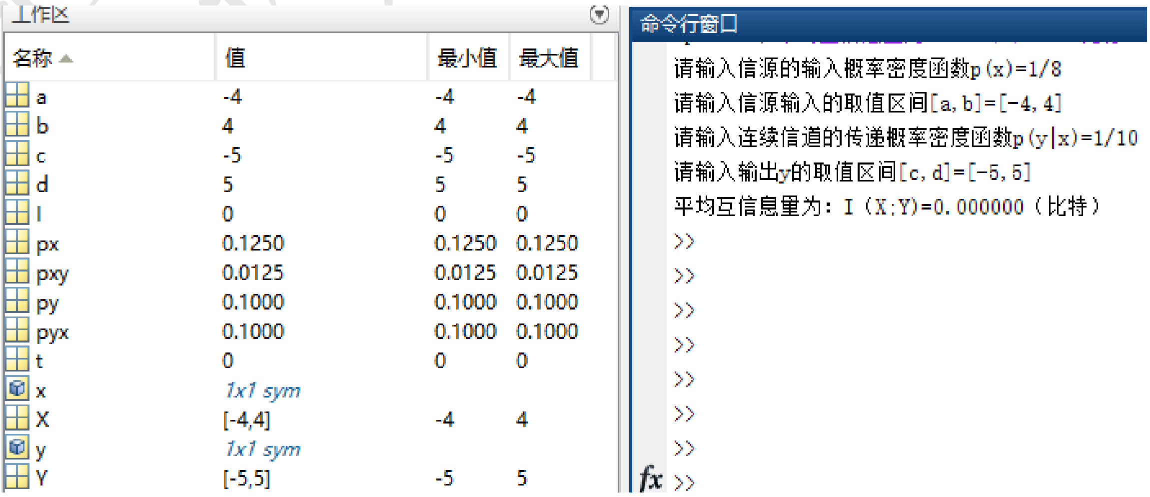


图 4.7 单维连续信道的平均互信息量计算结果

3.4.2　信道容量的计算

与离散信道一样，连续信道的信道容量也定义为平均互信息的最大值。而连续信道的平均互信息的计算涉及相对熵，因此利用最大相对熵定理可以求得连续信道的信道容量。根据上一节中对一种常见的特殊信道——高斯加性信道的信道容量的分析，在已知信源输入的平均功率和噪声的平均功率，或已知信噪比的条件下，可以利用 MATLAB 完成式（4-13）的仿真计算。

假设零均值高斯连续信源的平均功率为 10W，零均值高斯加性噪声的平均功率为 1W，则高斯加性信道的信道容量计算结果如图 4.8 所示。

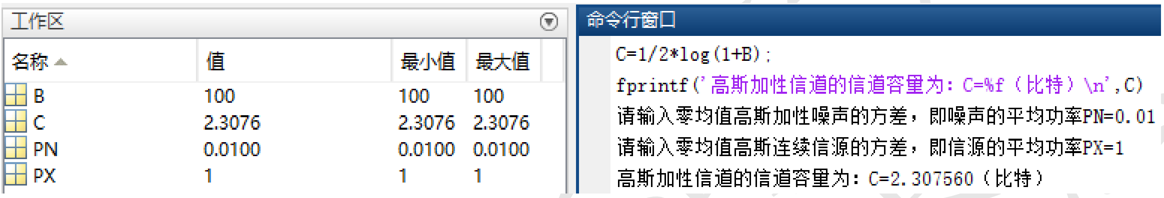


图 4.8 高斯加性信道的信道容量计算结果

3.5 多维连续信源与信道

在实际通信系统中，连续信源输出的消息往往是时间和取值都连续的随机函数 x(t)。这种信源称为波形信源，可以用随机过程{x(t)}来表示。一般随机过程的分析较为复杂，但在通信工程中，由于观察信号的时间、通信设备的通频带都是有限的,因此通信系统的随机过程一半可以近似地认为是限时、限频的随机过程。而限时、限频的随机过程在一定条件下可以转换成时间域或频率域上离散、取值连续的多维连续随机变量序列。因此随机过程通过限时、限频的连续通信系统的信息传输问题，可以转换成多维连续信源与多维连续信道问题加以讨论。

3.5 多维连续信源与信道的数学模型

在解决信息传输问题之前，首先把时间和取值都连续的随机过程变换成时间离散而取值连续的多维连续随机变量序列。这一过程是以时间域上的抽样定理为理论依据的。由抽样定理可以推论，若随机过程{x(t)}不包含 F 赫兹以上频率，则在限定的观察时间 T 内，随机过程{x(t)}可转换为时间域上间隔Δ=1/2F 的 N=2FT 个{x(t)}样点的随机变量序列$\mathrm{X}=X_1X_2\cdots X_N$。

3.5.1 多维连续信源的数学模型

根据上述抽样定理及其推论可以知道，对于限时 T、限频 F 的时间和取值都连续的波形信源{x(t)}，可以将它转换为时间间隔Δ=1/2F 的 N=2FT 个{x(t)}样点的随机变量序列，则构成 N 维连续信源

$$\mathbf{X}=X_1X_2\cdots X_N \qquad (5\text{-}1)$$

其中随机变量X_i的取值都是连续的。

多维连续信源的概率密度函数可以用 N 维联合概率密度函数$\mathrm{p}(x_1x_2\cdots x_N)$来表达。则多维连续信源的数学模型是概率空间

$$\begin{bmatrix}\boldsymbol{X}\\ p(\boldsymbol{x})\end{bmatrix}$$

其中 $\mathbf{X}=X_1X_2\cdots X_N$ ， $\mathrm{p}(\mathbf{x})=\mathrm{p}(x_1x_2\cdots x_N)=\prod_{i=1}^{N}p(x_i)$， $x_i\in X_i(i=1,\cdots,N)$，并且$\int_{\boldsymbol{X}}p(\boldsymbol{x})\,d\boldsymbol{x}=\int_{X_1X_2\cdots X_N}\cdots\int p(x_1x_2\cdots x_N)dx_1dx_2\cdots dx_N=1$。

3.5.2 多维连续信道的数学模型

设连续信道的输入区间为 X: $[a,b]$，输出区间为 Y:$[a',b']$，传递概率密度函数为$p(y|x)$,并且传递概率密度函数满足完备性，即

$$\int_{a'}^{b'} p(y|x)\,dy = 1 \quad x \in [a,b] \qquad (5\text{-}2)$$

信道能通过的最高频率为 F，连续信源{x(t)}的取值区间为 X: $[a,b]$，观察时间为$[0,T]$。现将连续信源{x(t)}与信道(X-Y)相接组成一个限时 T、限频 F 的通信系统，如图 5.1 所示。

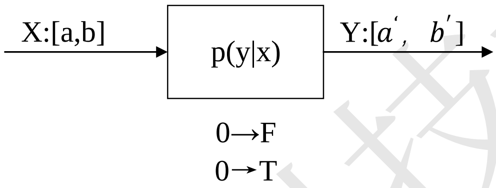


图 5.1 限时 T 限频 F 的连续系统

由时间域的抽样定理及其推论可知，波形信源{x(t)}可以转化为时间间隔Δ=1/2F 的 N=2FT 个连续随机变量$X_i(i=1,\cdots,N)$组成的随机矢量$\mathbf{X}=X_1X_2\cdots X_N$。每一时刻的随机变量$X_i(i=1,\cdots,N)$的取值区间均为$[a,b]$，均适合信道传输，并在信道的输出端输出相应的随机变量$Y_i(i=1,2,\cdots N)$，其取值区间为$[a',b']$。从第一时刻到第 N 时刻，在信道输出端相继出现随机变量$Y_1Y_2\cdots Y_N$，组成随机矢量$\mathbf{Y}=Y_1Y_2\cdots Y_N$。在这种情况下，从总体传递效果来看，相当于形成了一个输入随机变量序列为$\mathbf{X}=X_1X_2\cdots X_N$、输出随机变量序列为$\mathbf{Y}=Y_1Y_2\cdots Y_N$的新的信道。这个信道即为单维连续信道(X-Y)的 N 次扩展信道，亦即 N 维连续信道，如图 5.2 所示是 N 维信道的信道模型。

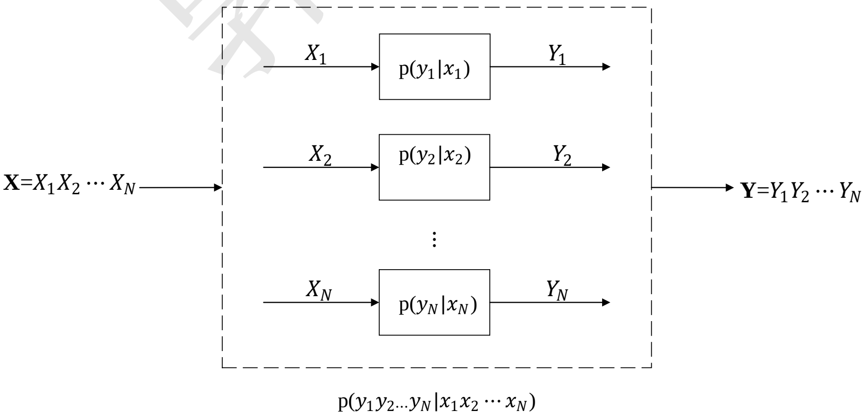


图 5.2N 维信道的信道模型

3.6 多维连续信源与信道的信息量

与从单维离散信源与信道到多维离散信源与信道的推导相似，根据上一章对单维连续信源与信道的介绍，可以得出多维连续信源与信道的相关信息测度。

3.6.1 多维连续信源的相对熵

由相对熵的定义，N 维联合相对熵

$$\mathrm{h}(\mathbf{X})=\mathrm{h}(X_1X_2\cdots X_N)=-\int_R p(\boldsymbol{x})\log p(\boldsymbol{x})\,d\boldsymbol{x} \qquad (5\text{-}3)$$

当 N=2 时，即得二维联合相对熵

$$\mathrm{h}(X_1X_2)=-\iint_R p(x_1x_2)\log p(x_1x_2)\,dx_1dx_2 \qquad (5\text{-}4)$$

与单维连续信源的相对熵一样，以下将根据文献[9]对满足均匀分布、高斯分布的连续信源的相对熵进行讨论。

(一)均匀分布的连续信源

设 N 维连续信源$\mathbf{X}=X_1X_2\cdots X_N$服从均匀分布，随机变量$X_i$取值的区间为$[a_i,b_i](\mathrm{i}=1,2,\cdots,\mathrm{N})$，此时信源的相对熵为

$$\mathrm{h}(\mathbf{X})=\mathrm{h}(X_1X_2\cdots X_N)=\log\left\{\prod_{i=1}^{N}(b_i-a_i)\right\}=\sum_{i=1}^{N}\log(b_i-a_i)$$

$$=\sum_{i=1}^{N}h(X_i) \qquad (5\text{-}5)$$

上式表明，N 维均匀分布的连续信源$\mathbf{X}=X_1X_2\cdots X_N$的相对熵$\mathrm{h}(\mathbf{X})=\mathrm{h}(X_1X_2\cdots X_N)$，等于 N 维区域体积$\prod_{i=1}^{N}(b_i-a_i)$的对数，也等于$\mathbf{X}=X_1X_2\cdots X_N$中各变量$X_i$在各自区间$[a_i,b_i](\mathrm{i}=1,2,\cdots,\mathrm{N})$内均匀分布时的相对熵$h(X_i)=\log(b_i-a_i)(\mathrm{i}=1,2,\cdots,\mathrm{N})$之和。这说明维均匀分布连续信源$\mathbf{X}=X_1X_2\cdots X_N$中各变量$X_i(\mathrm{i}=1,2,\cdots,\mathrm{N})$之间统计独立，各变量$X_i$在各自取值区间$[a_i,b_i](\mathrm{i}=1,2,\cdots,\mathrm{N})$中也服从均匀分布。

(二)高斯分布的连续信源

- 协方差矩阵为[M]的 N 维高斯连续信源

设 N 维连续信源$\mathbf{X}=X_1X_2\cdots X_N$服从高斯分布，协方差矩阵为[M]，则信源的相对熵

$$\mathrm{h}(\mathbf{X})=\mathrm{h}(X_1X_2\cdots X_N)=\frac{1}{2}\log|M|+\frac{N}{2}\log(2\pi e) \qquad (5\text{-}6)$$

上式表明，N 维高斯信源$\mathbf{X}=X_1X_2\cdots X_N$的相对熵$\mathrm{h}(\mathbf{X})=\mathrm{h}(X_1X_2\cdots X_N)$只取决于协方差矩阵[M]。

- 无记忆 N 维高斯连续信源

设 N 维高斯连续信源是无记忆信源，即随机变量X_i之间不相关，此时协方差矩阵[M]的行列式$|M|$为 0，因此信源的相对熵

$$\mathrm{h}(X^N)=\sum_{i=1}^{N}h(X_i)=\sum_{i=1}^{N}\frac{1}{2}\log(2\pi e\sigma_i^2) \qquad (5\text{-}7)$$

上式表明，N 维高斯信源$\mathbf{X}=X_1X_2\cdots X_N$是由 N 个相互统计独立的、均值为$m_i$，方差为$\sigma_i^2(\mathrm{i}=1,2,\cdots,\mathrm{N})$的高斯连续随机变量$X_i(\mathrm{i}=1,2,\cdots,\mathrm{N})$组成的 N 维无记忆高斯信源$X^N=X_1X_2\cdots X_N$，其中每个高斯随机变量$X_i$的相对熵为$\mathrm{h}(X_i)=\frac{1}{2}\log(2\pi e\sigma_i^2)\ (\mathrm{i}=1,2,\cdots,\mathrm{N})$。

单维连续信源的最大相对熵定理表明，在一定限制条件下，相对熵存在最大值。对于 N 维连续信源$\mathbf{X}=X_1X_2\cdots X_N$，最大相对熵定理同样成立。

最大多维相对熵定理表明：

- 在取值区间限定为 N 维区域体积$\prod_{i=1}^{N}(b_i-a_i)$的条件下，均匀分布的 N 维连续信源$\mathbf{X}=X_1X_2\cdots X_N$的相对熵$\mathrm{h}(\mathbf{X})$达到最大值，其最大值仅取决于取值区间。
- 在协方差矩阵[M]限定的条件下，高斯分布的 N 维连续信源$\mathbf{X}=X_1X_2\cdots X_N$的相对熵$\mathrm{h}(\mathbf{X})$达到最大值，其最大值仅取决于协方差矩阵[M]。
- 在随机变量$X_i(\mathrm{i}=1,2,\cdots,\mathrm{N})$的方差$\sigma_i^2(\mathrm{i}=1,2,\cdots,\mathrm{N})$限定的条件下，无记忆高斯信源$X^N=X_1X_2\cdots X_N$的相对熵$h(X^N)$达到最大值，其最大值仅取决于方差$\sigma_i^2$。

3.6.2 多维连续信道的平均互信息量

N 维连续信道的传递特性取决于传递概率密度函数$\mathrm{p}(\mathbf{y}|\mathbf{x})$。一般情况下，N 次扩展信道(**X-Y**)的传递概率密度函数$\mathrm{p}(\mathbf{y}|\mathbf{x})$与单维连续信道(X-Y)的传递概率密度函数p(y|x)没有直接关联。但若 N 维信道(**X-Y**)的传递概率密度函数$\mathrm{p}(\mathrm{y}|\mathrm{x})=\mathrm{p}(y_1y_2\cdots y_N|x_1x_2\cdots x_N)$等于各时刻单维连续信道(X-Y)的传递概率密度函数$\mathrm{p}(y_i|x_i)\ (\mathrm{i}=1,2,\cdots,\mathrm{N})$的连乘，则此二信道分别为单维无记忆连续信道和 N 维无

记忆连续信道。

在某一时刻，N 维无记忆连续信道(**X-Y**)的输出只依赖于该时刻的输入，与该时刻之前的输入序列和输出序列无关，也与下一时刻的输入无关。因此，要分析 N 维连续信道(**X-Y**)的平均互信息量 I(**X;Y**)的特性，就需要分析 I(**X;Y**)与单维连续信道(X-Y)在各时刻传递的平均互信息量$\mathrm{I}(X_i;Y_i)$ $(\mathrm{i}=1,2,\cdots,\mathrm{N})$之和$\sum_{i=1}^{N}\mathrm{I}(X_i;Y_i)$之间的关系。

可以证明，N 维连续无记忆信道(**X-Y**)的平均互信息量为

$$\mathrm{I}(\mathbf{X};\mathbf{Y})\le \mathrm{NI}(\mathrm{X};\mathrm{Y}) \tag{5-8}$$

当且仅当输入信源是 N 维连续无记忆信源$X^N=X_1X_2\cdots X_N$时，等式成立。

上式表明，N 维连续无记忆信道的平均互信息量一般不会超过连续随机变量 X 通过单维无记忆信道的平均互信息量的 N 倍。当且仅当输入信源是 N 维连续无记忆信源时，N 维信道的平均互信息量才能达到单维信道平均互信息量的 N 倍。这一结论是研究无记忆信道的信道容量问题的必要理论依据。

对于一般的多维连续信道，与多符号离散信道的平均互信息量的计算类似，其平均互信息量可以由概率密度函数与条件概率密度函数之间的双重积分得到，只是连续信道的表达式中用概率密度函数替代了离散信道的概率分布函数、用积分号替代了求和号。

3.7 多维连续信道的信道容量

在研究了 N 维无记忆连续信道的平均互信息量后，以下将开始讨论平均互信息量的最大值即信道容量问题。其中重点分析了限带高斯白噪声加性信道(AWGN 信道)的信道容量。

白噪声{n(t)}是指功率谱密度均匀分布于整个频率域的各态历经的平稳随机过程，其功率谱密度在整个频率与内是一个常数。在限时 T、限频 F 的限制条件下，均值为 0、功率谱密度为$|H(f)|^2=\frac{N_0}{2}(-F\le f\le F)$的高斯白噪声加性信道可以由高斯加性无记忆信道进行 N 次扩展得到。又由式（5-8）可知当信源是 N 维无记忆连续信源时，N 维无记忆连续信道的平均互信息最大值为单维无记忆连续信道的平均互信息量的 N 倍。因此 N 维连续信道的平均互信息量的最大值也应是单维连续信道平均互信息量的最大值的 N 倍。也就是说，N 维无记忆连续信道的

信道容量C_N应是单维信道信道容量C_0的 N 倍。其中，由式（4-13）可知，当信道中的噪声是均值为 0、方差$\sigma_N^2 = N_0 F$的高斯随机变量 N，这样的高斯加性信道的信道容量为$C_0 = \frac{1}{2}\log(1 + \frac{\sigma_X^2}{\sigma^2})$。

因此，在限时 T、限频 F 的条件下，满足均值为 0、功率谱密度$|H(f)|^2 = \frac{N_0}{2}(-F \le f \le F)$的平稳高斯白噪声加性信道的信道容量为

$$C_N = NC_0 = \frac{N}{2}\log(1 + \frac{\sigma_X^2}{\sigma^2}) = \frac{N}{2}\log\left(1 + \frac{P_{X_0}}{P_{N_0}}\right) = FT\log\left(1 + \frac{S_0}{N_0}\right)\text{（比特）} \quad (5\text{-}9)$$

进而，在单位时间内，高斯白噪声加性信道的信息传输率的最大值即信道容量

$$C_t = \frac{C_N}{T} = F\log\left(1 + \frac{S_0}{N_0}\right)\text{（比特/秒）} \quad (5\text{-}10)$$

式（5-10）就是著名香农公式。它表明，在限时 T、限频 F 的条件下，高斯白噪声加性信道的信道容量C_N的大小与信道的通频带宽度 F、观察信号的持续时间 T 以及信道的信噪比有关。并且，实际中为了达到系统要求，可以对通频带宽度、信噪比和信道容量这三者进行互换，变换某一者的取值来调节其它值的大小。

在实际情况中，信道常常是非高斯连续信道。在第四章已经讨论过，高斯加性信道的信道容量是非高斯加性信道的信道容量的下限，因此平均功率受限条件下高斯加性信道是最差的信道。因而，香农公式可用于计算其它一切非高斯加性连续信道的信道容量，其信道容量的值会比由香农公式得到的下限值大。

3.8 MATLAB 实现

与计算单维连续信源与信道的信息量与信道容量相同，对信息量计算的是信源的相对熵而不是绝对熵，虽然相对熵不具有信息测度，但它能很好地解决平均互信息量的计算问题。对平均互信息量，本节实现了 N 维无记忆连续信道的平均互信息量的计算。对信道容量，完成了高斯白噪声加性信道的最大信息传输率即信道容量的计算。

3.8.1 信息量的计算

与第四章计算单维连续信源的信息量类似，首先实现几种常见连续信源的相对熵的计算，再利用概率密度函数和条件概率密度函数的积分来实现平均互信息

量的计算。

(一)N 维连续信源的相对熵

下面根据文献[7]中介绍的几种多维连续信源相对熵及其性质来讨论几种信源。

- 均匀分布的连续信源

由式（5-5）可知，当随机变量X_i的取值区间为$[a_i, b_i](\mathrm{i}=1,2,\cdots,\mathrm{N})$，满足均匀分布的 N 维连续信源的相对熵为$\mathrm{h}(\mathbf{X}) == \sum_{i=1}^{N} h(X_i) = \sum_{i=1}^{N} \log(b_i - a_i)$。因此只要知道各随机变量的取值区间，分别求出各个变量X_i在各自区间内均匀分布时的相对熵，再将这些相对熵进行累加求和，就能计算出多维连续信源的相对熵。

例如，满足均匀分布的 3 维连续信源$\mathbf{X} = X_1X_2X_3$，输入随机变量X_i的取值区间分别为[-10,10]、[-8,8]、[-5,5]。则该均匀分布的 3 维连续信源的相对熵计算结果如图 5.3 所示。

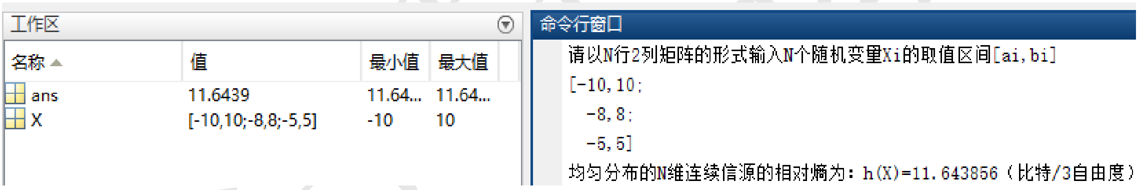


图 5.3 均匀分布的 N 维连续信源的相对熵计算结果

- 高斯分布的连续信源

若协方差矩阵为[M]，则满足高斯分布的 N 维连续信源的相对熵可以利用式（5-6）$\mathrm{h}(\mathbf{X}) = \frac{1}{2}\log|M| + \frac{N}{2}\log(2\pi e)$求得。当知道 N 维连续信源的各输入随机变量的均值后，就能直接利用 MATLAB 提供的 cov 函数求得协方差矩阵，再利用 det 函数求得矩阵的行列式，就能很容易地完成高斯分布连续信源的相对熵的计算。

例如，若满足高斯分布的 4 维连续信源，其 4 个输入随机变量X_1、X_2、X_3、X_4的均值分别为 2、3、4、5，则该连续信源的相对熵计算结果如图 5.4 所示。

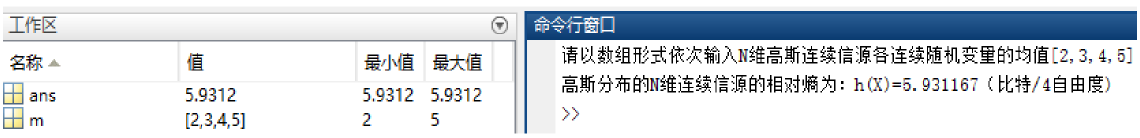


图 5.4 高斯分布的 N 维连续信源的相对熵计算结果

- 无记忆高斯连续信源

对于 N 维无记忆高斯连续信源，由式（5-7）可以得知其相对熵等于各随机变量X_i的相对熵 h(X_i)之和，而各个随机变量X_i都是方差为$\sigma_i^2(\mathrm{i}=1,2,\cdots,\mathrm{N})$的高斯随机变量，其相对熵只与随机变量的方差有关。因此，只要知道各随机变量的方差，根据$\mathrm{h}(X_i)=\frac{1}{2}\log(2\pi e\sigma_i^2)$分别求出各个变量$X_i$的相对熵，再将这些相对熵进行累加求和，就能计算出多维连续信源的相对熵。

例如，若 3 维无记忆高斯连续信源的 3 个输入随机变量X_1、X_2、X_3的方差分别为 1、2、3，则该连续信源的相对熵计算结果如图 5.5 所示。

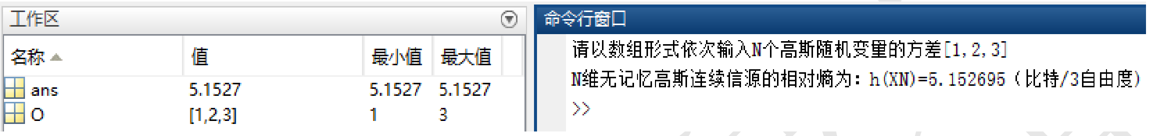


图 5.4N 维无记忆高斯连续信源的相对熵计算结果

(二)N 维连续信道的平均互信息量

由式（5-8）可以知道当输入信源为 N 维无记忆连续信源时，N 维无记忆连续信道(**X-Y**)的平均互信息量$\mathrm{I}(\mathbf{X};\mathbf{Y})=\mathrm{NI}(\mathrm{X};\mathrm{Y})$，等于输入随机矢量 **X** 中任一时刻随机变量$X_i(\mathrm{i}=1,2,\cdots,\mathrm{N})$通过单维无记忆信道后，与输出随机变量$Y_i(\mathrm{i}=1,2,\cdots,\mathrm{N})$之间的平均互信息的 N 倍大小，也就是构成 N 次扩展信道的单维连续信道(X-Y)的平均互信息量的 N 倍大小。因此，只要根据第四章计算单维连续信道的平均互信息量的方法求出输入随机变量 X 的通过无记忆信道(X-Y)的平均互信息量，再乘以信道扩展次数 N，就能得到 N 维无记忆连续信道的平均互信息量。

例如，已知 4 维无记忆连续信道的输入随机变量 X 的概率密度函数为 p(x)及其输入区间。信道的传递概率密度函数及信道输出随机变量 Y 的输出区间都给出。则该连续信源的相对熵计算结果如图 5.5 所示。

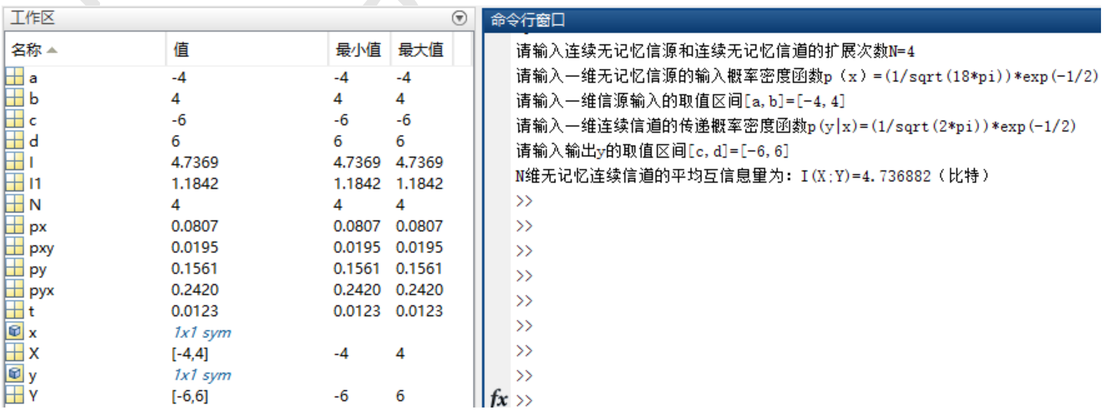


图 5.5 多维连续信道平均互信息量的计算结果

3.8.2 信道容量的计算

香农公式是计算多维连续信道及波形信道的信道容量的最重要、最简洁的公式。对于限带高斯加性白噪声加性信道而言，只要知道带宽 F、输入信号的平均功率及噪声的平均功率或平均功率谱密度（或信噪比）就能利用香农公式计算出此信道的最大信息传输率即信道容量。对于有色高斯加性信道，原则上只要知道输入信号的平均功率和有色噪声的平均功率普密度，就能计算出信道容量，但一般情况下其计算十分复杂。因此在同样限带、限平均功率的条件下，可以用香农公式计算的信道容量来作为真实信道容量的下限值。

例如，设高斯白噪声加性信道(AWGN 信道)的通频带宽度为 3000Hz，输入信号的平均功率为 3W，噪声平均功率为 0.03W，此时的信噪功率比为 100 即 20dB。则该信道的最大信息传输率即信道容量的计算结果如图 5.6 所示。

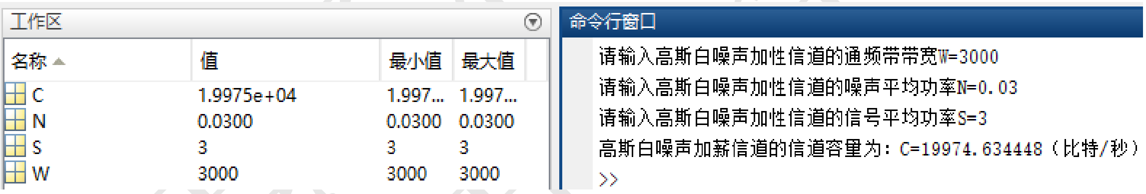
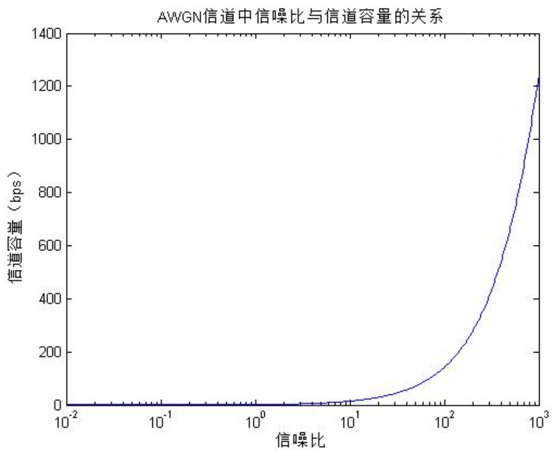


图 5.6 AWGN 信道的信道容量计算结果

以下将分析信噪比、带宽与信道容量这三者之间的关系。如图 5.7 和图 5.8 分别是信噪比与信道容量的关系图、带宽与信道容量的关系图。



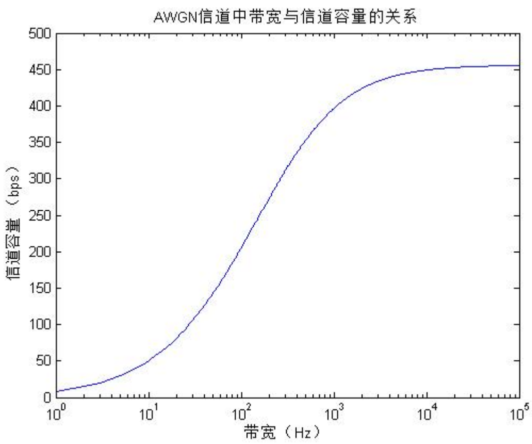


图 5.7 信噪比与信道容量的关系图 5.8 带宽与信道容量的关系

从图 5.7 可看出提高信噪比能增加信道的信道容量，当噪声功率趋于 0 时信道容量趋于无穷，这意味着无干扰连续信道的信道容量为无穷大。从图 5.8 可看出增加带宽可增加信道容量，但并不能无限制地使信道容量增大。

把这三者结合起来，可以得到如图 5.9 所示的带宽、信噪比、信道容量三者的关系图。从图中可以看出，当信道容量一定时带宽与信噪比之间可以互换，这在实际中非常有用。

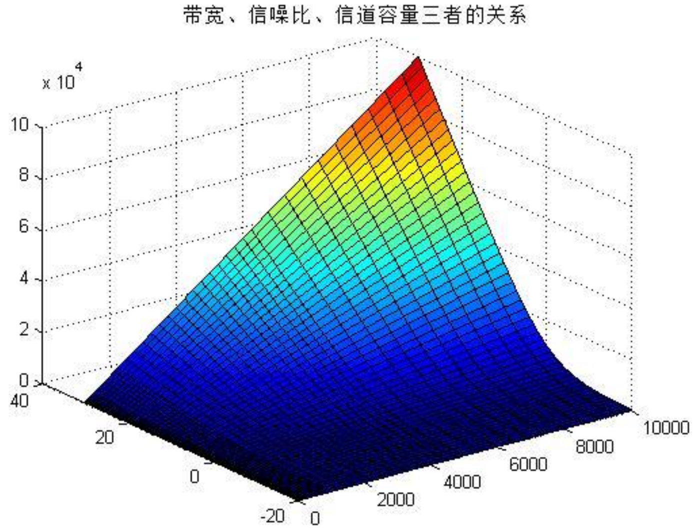


图 5.9 带宽、信噪比与信道容量三者的关系

第四章　信息量的应用

前面四章分别讨论了离散信道和连续信道的信息量与信道容量的含义、性质及计算方法，并运用 MATLAB 实现了仿真计算。本章将在已有理论的基础上联系实际，探讨信息量在一些领域中的意义。最后联系随机信号分析课程中的互相关函数，比较互信息量与互相关函数在描述变量之间相关性的能力，得出互信息量优于互相关函数这一结论。

4.1　文本

通信完成后接收者从消息中获取的信息量，在数值上等于通信前后不确定性的减少量。研究信息量是从统计意义上讲的平均信息量，包括信源的自信息量和发端与收端之间的互信息量。

对于一篇由字符组成的文本，其信息量的计算需要先统计各字符出现的频次在文本总字符数中占的比例，以此作为字符出现的概率$p(i)$。如果将一篇文本看成信源，则其平均自信息量，即信源熵为$\mathrm{H}=-\sum_i p(i)\log p(i)$。

此时信源熵的实际意义是，读者看完一篇文章后，所能获得的信息的多少，也就是阅读前后对文章所叙述内容的平均不确定性的减少量。在这里，本文讨论的信息是指语法信息，而非语义信息，因为一段文字甚至一个词语的含义，在不同情境下都可以不同。信息也非语用信息，因为一篇文章所包含的内涵对不同读者而言都是不一样的，不同的读者阅读完同一篇文章，从文章中提取出的信息量是不一样的。语用信息有很大的主观性。

互信息量是衡量两种信号之间相关性的统计量。对于两篇文本，互信息量可以用来衡量它们之间的相似性，如果互信息量非常大，则这两篇文本相似度非常高；反之，如果互信息量非常小，则阅读完其中一篇文本后，将很难获得关于另一篇文本的信息，也就是说，这两篇文本的相似度非常低。这可以应用于文本相似度检测中。

4.2 图像

4.2.1 静止图像

对于图像，可以使用图像像素上的灰度值来充当一个离散的符号。

图像熵是一种特征的统计形式，它反映了图像包含的平均自信息量：

①. 当图像为纯色图时(白或黑)，只有一个灰度值，说明图像不包含任何地目标，此时熵最小，H=0，图像包含的信息量为零；

②. 当图像包含 N 个灰度值时，即图像中的每个像素的灰度值不相同，可以认为图像每个单一像素都是一个独立地物目标，类似于地图充满了物体，因此此时熵最大，为 H=logN，图像的信息量最大。

因此，图像的熵 H 越大，则图像包含的像素灰度种类越丰富，灰度分布越均匀，图像的地物目标越多，图像的信息量越大，反之则反。

图像熵可以分为一维熵和二维熵。一维熵表示图像中灰度分布的聚集特征所含有的信息量。设图像中灰度值为 i 的像素占总像素的比例为 p(i)，该值代表某个灰度在该图像中出现的概率，可以根据灰度直方图得到。定义灰度图像的一元灰度熵为:

$$\mathrm{H} = -\sum_{i=0}^{255} p(i) log_2 p(i) \qquad (6\text{-}1)$$

图像的一维熵可以表示图像灰度分布的聚集特征，却不能反映灰度分布的空间特征。二维熵正是能反映会度分布空间特征的特征量。图像灰度分布的空间特征量可以由图像邻域灰度均值来表示，将它与图像的像素灰度构成特征二元组(i,j)，其中 i 表示像素的灰度值(0≤i≤255)，j 表示邻域灰度(0≤j≤255)，f(i,j)为特征二元组(i,j)的出现频数，N 为图像的尺度，则离散的图像二维熵可以定义为：

$$P_{ij} = \frac{f(i,j)}{N^2} \qquad (6\text{-}2)$$

$$\mathrm{H} = -\sum_{i=0}^{255} P_{ij} \log P_{ij} \qquad (6\text{-}3)$$

上式能表示某个像素位置上的灰度值与周围像素的灰度分布的综合特征。也就是说，在图像所包含信息量的条件下，图像二维熵能够综合反映图像中像素位置的灰度信息和像素邻域内灰度分布的特征。

图像熵可以表示接收者在收到一张图像后所能获得的信息量。一张色彩鲜艳、五彩斑斓的图像，对观察者而言势必造成极大的视觉冲击，因此获得的信息量非

常大；而一张空白的图像会使观察者感觉平淡无味，将无法使观察者得到任何信息。因此，在现实情境中，可以尝试将图像熵定义为图像对观察者造成的视觉冲击的大小。

例如，日常生活中拍照时可以给图片加上各种效果的滤镜，不同的滤镜其实是给图片蒙上了不同的色调。对与同一张图片，加上复古风格的滤镜后，它的图像熵将比加上清新风格的滤镜后的图像熵更小。

对于两张静态图像，可以用互信息量来衡量图像之间的相似性。两张图像之间的互信息量越大，说明它们之间的相似度越大，也就是说，在已经获得一张图像的条件下，对另一张图像存在的不确定越小。因此，图像之间的互信息量可以应用在图像相似性检测中，例如人脸识别等身份验证的场合，这在日常生活中有重要的实际意义。

4.2.2 动态图像

对于动态图像，可以将它看成是多张静态图像的组合。因此可以认为，动态图像中所包含的平均自信息量，等于多张静态图像之间的联合平均自信息量。一个色彩明艳的视频，其所包含的平均自信息量会比一个色彩暗淡的视频所包含的信息量大。据此，可以得出结论，动态图像的平均自信息量代表其色调鲜艳程度。

若关注的是动态图像中前后两帧的关联性，即关注收到一帧静态图像后，能从中获得的对另一帧图像的信息量，则此时可以用两帧图像，即两张静态图像之间的平均互信息量来衡量。当两帧之间的互信息量较大时，两图像之间较为相似；而当两帧之间的互信息量值变化不大时，则说明后一图像与前一图像相比未发生较大变化。

例如，天气预报中，气象专家收到的是动态的气象图，若将动态的气象图看成由一帧一帧的静态图像组成的动态图像，则每两张静态图像之间的互信息量，可以表示两张静态图像之间的关联性。如果气象图中前后两帧的关联性很小，则表示云层变化非常快，导致前后两帧的变化非常大，因此，此时可以推断，天气将会发生大变化。在此，可以用平均互信息量来衡量天气变化的剧烈程度。

4.3 语音

语音的信息量的衡量，可以从时域和频域来看。

从时域来看，衡量对象是语音的振幅。将一段连续语音信号进行离散量化，统计各个振幅段出现的频率，以此作为概率来计算平均自信息量。对于一段语音，若各个时刻的声音振幅都一样，此时振幅只有一种，出现概率为 1，则平均自信息量为 0；若各个时刻的声音振幅都不一样，此时振幅的种类等于离散化时间区间的个数，在这种情况下，振幅相等于是呈现均匀分布，平均自信息量最大。因此，可以说，时域上语音的平均自信息量代表语音的抑扬顿挫程度。

从频域来看，语音的频率可以反应一个人说话的音调高低，体现一个人的发音特征。因此，可以用频率域上两段语音之间的互信息量来衡量两段语音之间的发音相似程度。注意，这里的发音相似程度是指音调高低的相似程度，而不是声音内容的相似程度。进而，在实际场景中，可以利用频域上语音的互信息量来表示两段语音的说话者是否为同一人的可能性高低。平均互信息量越高，则两段语音出自同一人之口的可能性越高；反之越小。也就是说，可以将频域上的平均互信息量用于语音识别。

4.4 互信息量与互相关函数

互信息量是度量随机变量之间相关程度高低的有效手段，这可以让人联想到随机信号课程中学习过的互相关函数，二者都有衡量变量间相关性大小的功能。互相关函数被广泛应用于实际工程中，而互信息量却不被发掘。实际上，互信息量能克服互相关函数无法度量非线性相关的变量之间的相关程度这一缺点，并且其应用不限于实值变量，因此能更好地满足实际需要。

4.4.1 互信息量与互相关函数的含义

(一)互信息量

互信息量是信息论课程中的重要信息度量，它表示一个随机变量中包含的关于另一个随机变量的信息量，也表示在得知其它随机变量的前后，对这个随机变量存在的不确定性的减少量。两个随机变量的互信息可以作为变量间相互依赖性

或相关联性的量度。

不同于相关系数，互信息并不局限于实值随机变量，也不局限于线性系统，它更加一般且实用。互信息最常用的单位是 bit。

考虑到前文已经对互信息量的概念、性质及计算方法做过详细的讨论，此处不再赘述。

(二)互相关函数

互相关函数是随机信号分析课程中的重要信息度量，被广泛应用于实际工程。互相关函数代表的是两个时间序列之间的相关程度，它描述了在任意两个不同时刻t_1、t_2，随机信号 x(t)与 y(t)的取值之间的相关程度，能用来确定输出信号在多大程度上由输入信号得来。在工程中能有效修正测量中因接入噪声源而带来的误差。

设两个函数分别是 f(t)和 g(t)，则互相关函数定义为 R(u)=f(t)*g(-t)，它表示两个函数在不同的相对位置上互相匹配的程度大小。

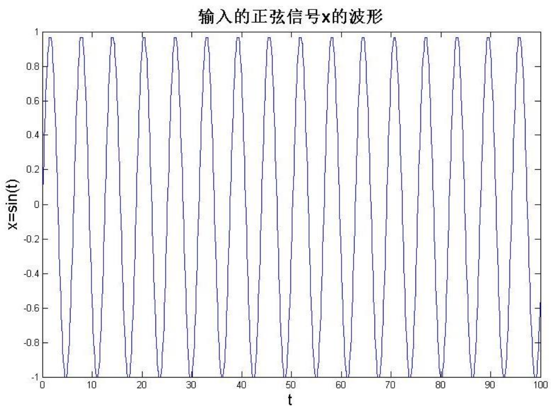
对于连续信号公式表示为$R(\tau)=\frac{1}{T}\int f(t)g(t+\tau)dt$，积分限为 0 至 T。

对于离散信号公式表示为$R(n)=\frac{1}{N}\sum x(m)y(m+n)$，其中 m 从 0 到 N-1 变化。

4.4.2 互信息量与互相关函数的性能对比

利用 MATLAB 分析正弦函数先后通过线性系统和非线性系统的输出与输入之间的相关程度高低，即分别计算正弦信号经过两种系统后的互信息量与互相关函数值。

设输入信号为 x=sin(t)，t 在[0,100]区间每隔 0.1 取一个点。线性系统的系统函数为$y_1=5x+10$，非线性系统的系统函数为$y_2=x^2+x^3$。图 6.1 和图 6.2 分别是输入信号和输出信号的波形图。



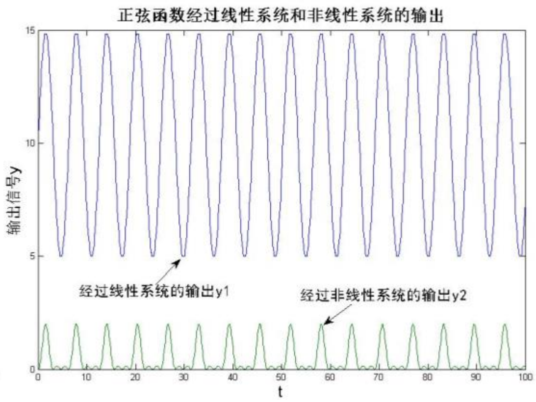


图 6.1 输入正弦信号 x 的波形　图 6.2 输出信号y_1和y_2的波形

要计算正弦信号经过系统的互信息量，就需要计算信源熵、输出 y 的平均自信息量、输入 x 与输出 y 之间的联合熵。计算信源熵和 y 的平均自信息量时，对输入信号、输出信号这样的连续信号进行离散、量化，分别统计落入各个量化区间的时间区间个数，进而计算出频率，最后得到自信息量。而计算联合熵时，则需要将 x 和 y 这两个变量张成二维空间，进行离散量化，统计落入各个小空格的频率，最后得到联合熵。

要计算正弦信号通过系统的互相关函数值，可以利用 MATLAB 提供的 xcorr 函数，在函数中标明“coeff”，就可以得到无偏的互相关函数值。

由于此处平均互信息量的计算较为复杂，以下将给出计算互信息量的详细过程。

①. 计算信源熵 h(X)。先将输入 x 的取值范围分成 30 小块，并统计落入各个区间的样本数，即为频次。如图 6.3 是统计结果的直方图，所有区间的频次总和为 x 离散化后的点数 1000。再根据频次算出频率 p(x)，就能利用信源熵公式算出 h(X)。

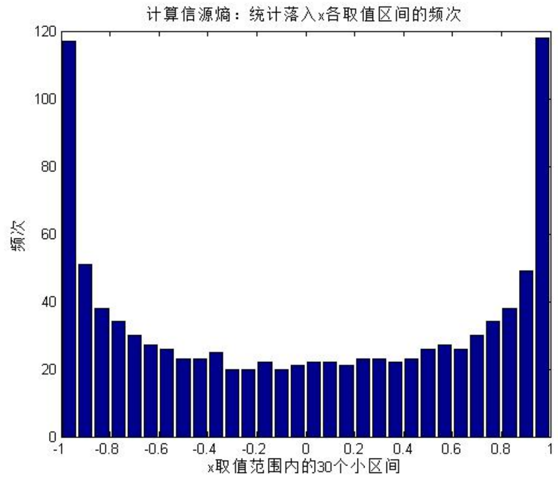
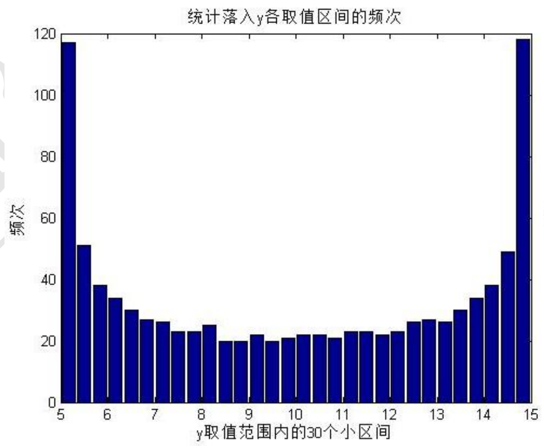


图 6.3 统计落入 x 各取值区间的频次

从图 6.3 可以看出，将正弦函数 x 的取值范围等分为 30 个小区间后，这 1000 个组成 x 的样点分别落入不同区间。其中，x 取值范围的两端有更多的样点落入，这是因为正弦函数在越接近峰值时，数据变化越迅速，因此 MATLAB 对正弦函数默认的离散化规则不是均匀分布的，零点附近分布的抽样点较稀疏，峰值附近分布的抽样点较密集。

②. 计算输出 y 的平均自信息量。与计算信源熵的方法一样，计算出平均自信息量 h(Y)。如图 6.4 为统计落入线性输出y_1各取值区间的频次直方图，图 6.5 是统计落入非线性输出y_2各取值区间的频次直方图。



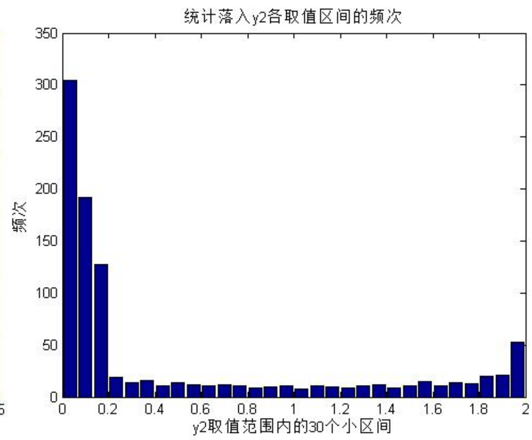


图 6.4 统计落入y_1各取值区间的频次图图 6.5 统计落入y_2各取值区间的频次图

③. 计算 x 与 y 的平均互信息。只要能计算出联合概率p(xy)，就能算出平均互信息量。要计算联合概率，需要将 x 和 y 张成二维空间，等分成 30×30 个小区间，统计落入这些小区间的样本数，就可以算出联合概率分布，