**多媒体技术实验报告**

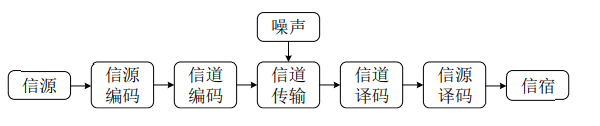
龙晓怡 2018302100026

**香农编码MATLAB实现**

# 1实验背景

## 1.1信源编码

### 1.1.1通信系统模型

一般的通信系统模型如图所示：

可以发现，信源编码处于通信系统模型的最前端，是信息进行传递时的第一个步骤，对信息传输具有重要意义。

### 1.1.2信源编码的定义

在计算机科学和信息论中，信源编码是按照特定的编码机制用比未经编码少的数据比特（或者其它信息相关的单位）表示信息的过程。信源编码是一种以提高通信有效性为目的而对信源符号进行的变换，或者说为了减少或消除信源冗余度而进行的信源符号变换。具体说，就是针对信源输出符号序列的统计特性来寻找某种方法，把信源输出符号序列变换为最短的码字序列，使后者的各码元所载荷的平均信息量最大，同时又能保证无失真地恢复原来的符号序列。

### 1.1.3信源编码的作用

信源编码的作用之一是，即通常所说的数据压缩；作用之二是将信源的模拟信号转化成数字信号，以实现模拟信号的数字化传输。

### 1.1.4信源编码的分类

根据信源的性质进行分类，则有信源统计特性已知或未知、无失真或限定失真、无记忆或有记忆信源的编码；按编码方法进行分类可分为分组码或非分组码、等长码或变长码等。然而最常见的是讨论统计特性已知条件下，离散、平稳、无失真信源的编码，消除这类信源剩余度的主要方法有统计匹配编码和解除相关性编码。比如香农码、哈夫曼码，它们属于不等长度分组码，算术编码属于非分组码；预测编码和变换编码是以解除相关性为主的编码。对限定失真的信源编码则是以信息率失真函数𝑅(𝐷)为基础，最典型的是矢量量化编码。对统计特性未知的信源编码称为通用编码。

# 2 实验目的

1.理解信源编码的意义。

2.掌握香农编码的过程。

3.对给定的信源使用MATLAB进行香农编码。

# 3 实验原理

## 3.1香农第一定理及相关概念

香农第一定理指出了平均码长与离散信源概率之间的关系，同时也指出了可以通过编码使平均码长达到极限值，这是一个很重要的极限定理。根据香农第一定理，当选择每个码字的长度𝑙𝑖为满足下式的一个整数时，这种编码方法就是香农编码。

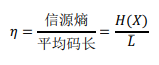
𝐼(𝑥𝑖)≤𝑙𝑖≤𝐼(𝑥𝑖)+1

其中𝐼(𝑥𝑖)表示信源符号𝑥𝑖的信息量，即，𝐼(𝑥𝑖)=−𝑙𝑜𝑔2𝑝(𝑥𝑖)

经过香农编码之后，码字的平均码长𝐿为：

其中，𝑝(𝑥𝑖)表示信源符号𝑥𝑖的统计概率，𝑙𝑖表示信源符号𝑥𝑖经过香农编码得到

的码字的长度。

香农编码的效率为：

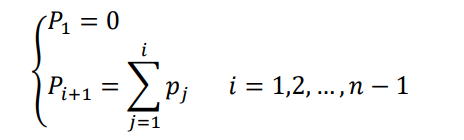
## 3.2香农编码流程

香农编码严格意义上来说不是最佳码，它是采用信源符号的累计概率分布函数来分配码字。

### 3.2.1编码具体步骤

1）将n个信源符号按概率从大到小顺序进行排列，即：𝑝1≥𝑝2≥⋯⋯≥𝑝𝑛

2）根据不等式−log2(𝑝(𝑥𝑖))≤𝑙𝑖≤−log2(𝑝(𝑥𝑖))+1，计算信源符号𝑥𝑖对应的码字的码长（𝑙𝑖取在此范围内的整数），其中，𝑝(𝑥𝑖)表示信源𝑥𝑖的2统计概率。

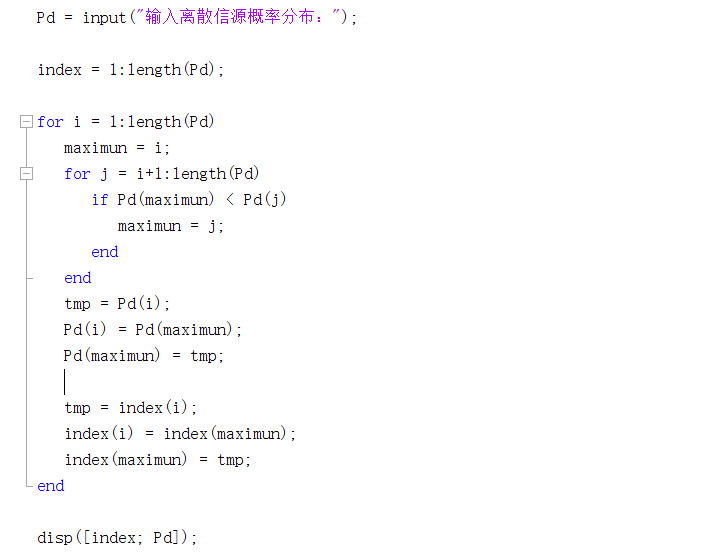
3）根据下式计算排序后信源符号𝑥𝑖的累加概率𝑃𝑖：

4）将累加概率𝑃𝑖变换成二进制小数，取小数点后的𝑙𝑖位二进制数作为信源𝑥𝑖的码字。香农编码的效率不高，实用性不大，但对其他编码方法有很好的理论指导意义。一般情况下，按照香农编码方法编出来的码，其平均码长不是最短的，即不是紧致码（最佳码）。只有当信源符号的概率分布使（2）中不等式左边的等号成立时，编码效率才能达到最高。

### 3.2.2关键函数

#### 双层For循环降序排列离散信源概率

输出离散信源符号（值存在数组index中）和其出现概率（值存在pd）



#### 求二进制数

用排序后信源符号𝑥𝑖的累加概率𝑃𝑖乘以2后再次赋值给𝑃𝑖，如果整数部分有进位，则𝑃𝑖转化为二进制后的小数点后第一位为1，否则为0，将新的𝑃𝑖的小数部分再次乘以2，同样，如果小数部分有进位，则转化为二进制后的小数点后第二位为1，否则为0，从而得到小数点后的第二位，依此类推，直到得到了满足要求的位数，或者没有小数部分了为止，最终即可获得所有信源符号对应的码字。

#### 打印输出

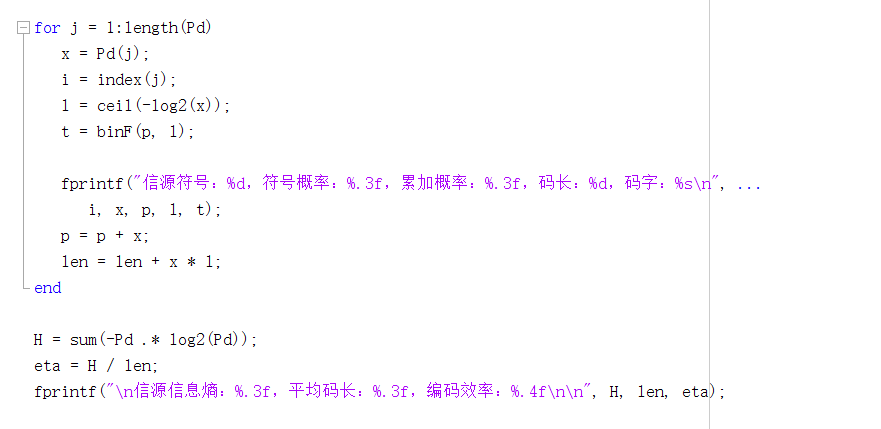
ceil()向上取整

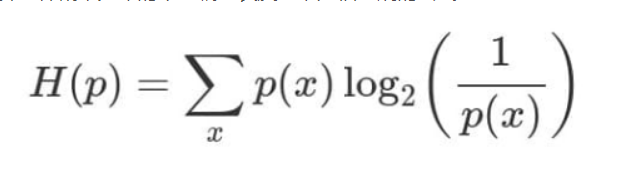
floor()向下取整

l=ceil(-log2(x)) 计算每个码字的长度

p存放累加概率

len计算平均码长=每个码字长度\*符号概率的和



信源信息熵H

编码效率=信源信息熵H/平均码长len

# 4 实验结果

运行程序，输入一组离散信源的分布概率，可以看到对应码字、码长，信源信息熵、平均码长和编码结果正确输出。

