

现实中,有些时候不得不进行某些不可逆编码的处理,或者有时候没有必要要求编码完全没有损失。例如,在实际的通信中,信息在信道的传输过程中总会受到噪声和干扰的影响,一般不可能完全保持原样发送,而或多或少总会产生一些失真。此外,随着科学技术的发展,数字系统的应用越来越广泛,这就需要传送、存储和处理大量数据。为了提高传送和存储的效率,往往需要压缩数据,这样也会带来一定的信息损失。信道编码定理虽然告诉我们有噪声信道的无失真编码似乎是可能的,但是这里的无失真只能是无限逼近 0,而无法达到 0,除非编码分组的长度无穷大。因此从这个角度讲,有噪声信道的无失真要求也是不可能达到的。而在实际生活中,人们一般并不要求完全无失真地恢复信息,通常要求在保证一定质量(一定保证度)的条件下再现原来的消息,也就是说允许失真的存在。例如,音频信号的带宽是 20~20000Hz,但只要取其中一部分即可保留主要的信息;在公用电话网中只选取带宽中的 300~3400Hz 即可使通话者较好地获取主要的信息;在要求有现场感的语音传输中取 50~7000Hz 的频带即可较好地满足要求。还有诸如连续信源、无理数这样的编码,在完全无失真的需求下,编码的长度将会是无穷大。

由此可见,不同的要求允许不同大小的失真存在,完全无失真的通信既不可能也无必要,而有必要进行将失真控制在一定限度内的压缩编码,将其称为限失真编码。

信息率失真理论是进行量化、数模转换、频带压缩和数据压缩的理论基础。本章主要介绍信息率失真理论的基本内容及相关的编码方法。

如何进行这种限失真编码呢?考虑我们前面提出的问题,如果要将有 10 万位小数的 1~100 的数字进行压缩,可以采取四舍五入的方法,将这个数转换为只有 10 位小数的数值。由于小数点 10 位之后的数值都是微不足道的,所以这种压缩带来的失真并不大。可以理解为将一个集合中的元素映射为另外一个集合中的压缩,或者是映射为原集合中的一部分的元素。

## 5.1 失真测度

### 5.1.1 系统模型

通过前面的例子和讨论,可以建立研究限失真信源编码(有损压缩)的系统模型,如图 5-1 所示。信源发出的消息  $X$  通过有失真的信源编码输出为  $Y$ ,由于是有失真的编码,所以  $X$  和  $Y$  的元素之间不存在一一对应的关系,可以假设  $X$  通过一个信道输出  $Y$ ,

这种假想的信道称为试验信道。这样,就可以通过研究信道的互信息来研究限失真编码,而  $X$  和  $Y$  的关系也可以用转移概率矩阵(信道矩阵)来表示。

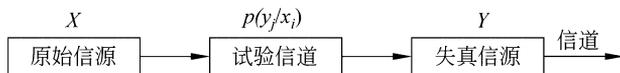


图 5-1 限失真编码模型

除了描述输入输出的关系外,我们还关心如何才能限制失真的问题,因为这一切都是建立在限失真的要求之上的。既然要限制失真,就需要有关于失真的度量。

### 5.1.2 失真度和平均失真度

如何来度量失真呢?我们先从最简单的单个符号的信源的失真度量(distortion measure)开始,然后以此为基础来建立更多符号的失真度量。

#### 1. 单个符号失真度

设有离散无记忆信源:

$$\begin{bmatrix} X \\ p(x_i) \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ p(x_1) & p(x_2) & \cdots & p(x_n) \end{bmatrix}$$

信源符号通过信道传送到接收端  $Y$ :

$$\begin{bmatrix} Y \\ p(y_j) \end{bmatrix} = \begin{bmatrix} y_1 & y_2 & \cdots & y_m \\ p(y_1) & p(y_2) & \cdots & p(y_m) \end{bmatrix}$$

信道的转移概率矩阵为:

$$[p(Y|X)] = \begin{bmatrix} p(y_1|x_1) & p(y_2|x_1) & \cdots & p(y_m|x_1) \\ p(y_1|x_2) & p(y_2|x_2) & \cdots & p(y_m|x_2) \\ \vdots & \vdots & & \vdots \\ p(y_1|x_n) & p(y_2|x_n) & \cdots & p(y_m|x_n) \end{bmatrix}$$

对于每一对  $(x_i, y_j)$ , 指定一个非负的函数  $d(x_i, y_j)$  为单个符号的失真度或失真函数(distortion function), 用它来表示信源发出一个符号  $x_i$  并在接收端再现  $y_j$  所引起的误差或失真。

失真函数是根据人们的实际需要以及失真引起的损失、风险和主观感觉上的差别大小等因素人为规定的。有时候未必能够证明为什么采用这个函数是合理的, 而其他的函数没有它好。假设发出一个符号, 如果收到的也是它, 则失真为 0; 如果收到的不是它, 而是其他的符号, 则存在失真。失真函数大于 0, 有:

$$d(x_i, y_j) = \begin{cases} 0, & x_i = y_j \\ \alpha, & x_i \neq y_j, \alpha > 0 \end{cases}$$

**注意:** 这里的  $x_i = y_j$  从表面上看, 两者的符号是相同的,  $x_i, y_j$  的符号集也应该是相同的。但是实际上没有必要要求符号相同, 只需要符号代表的值相同就行了, 比如圆周率  $\pi$  经过定义可以代表 3.1415926...

失真度还可表示成矩阵的形式:

$$[\mathbf{D}] = \begin{bmatrix} d(x_1, y_1) & d(x_1, y_2) & \cdots & d(x_1, y_m) \\ d(x_2, y_1) & d(x_2, y_2) & \cdots & d(x_2, y_m) \\ \vdots & \vdots & & \vdots \\ d(x_n, y_1) & d(x_n, y_2) & \cdots & d(x_n, y_m) \end{bmatrix} \quad (5-1)$$

$[\mathbf{D}]$ 称为失真矩阵,它是  $n \times m$  阶矩阵。

常用的失真函数有以下 4 个。

#### (1) 误码失真函数

$$d(x_i, y_j) = \delta(x_i, y_j) = \begin{cases} 0, & x_i = y_j \\ a, & \text{其他} \end{cases}$$

这种失真函数表示:当  $i=j$  时,  $X$  与  $Y$  的取值是一样的,用  $Y$  来代表  $X$  就没有误差,所以定义失真度为 0。当  $i \neq j$  时,用  $Y$  代表  $X$  就有误差,所有不同的  $i$  和  $j$  引起的误差都一样,所以定义失真度为常数  $a$ 。通常规定  $a=1$ ,此时失真称为汉明失真,失真矩阵为汉明失真矩阵,该矩阵的特点为主对角线上的元素全部为 0,其他全为 1。

$$\mathbf{D} = \begin{bmatrix} 0 & 1 & \cdots & 1 \\ 1 & 0 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 0 \end{bmatrix}_{r \times r} \quad (5-2)$$

#### (2) 均方失真函数

$$d(x_i, y_j) = (x_i - y_j)^2$$

这种失真函数称为平方误差失真函数,相应的失真矩阵称为平方误差失真矩阵。假如信源符号代表信源输出信号的幅度值,则意味着较大的幅度失真要比较小的幅度失真引起的错误更为严重,严重的程度用平方表示。

#### (3) 绝对失真函数

$$d(x_i, y_j) = |x_i - y_j|$$

#### (4) 相对失真函数

$$d(x_i, y_j) = |x_i - y_j| / |x_i|$$

上述 4 种失真函数的第一种适用于离散信源,后 3 种适用于连续信源。

## 2. 序列失真度

许多情况下,需要处理的信源为一个序列,可以将序列的每一个符号对应的失真求和进行平均。设  $x = (x_1, x_2, \dots, x_N)$ , 其中  $x_i$  取自符号集  $A$ , 而  $y = (y_1, y_2, \dots, y_N)$ ,  $y_j$  取自符号集  $B$ , 则序列的失真度定义为:

$$d(x, y) = \sum_{i=1}^N d(x_i, y_j) \quad (5-3)$$

序列的单个符号的失真度定义为:

$$d_N(x, y) = \frac{1}{N} \sum_{i=1}^N d(x_i, y_j) \quad (5-4)$$

在一些教材中,将序列的单个符号的失真度(求平均后的失真度)定义为序列的失真度,而实际上以上两种失真度量各有其不同的应用场合和意义。

这种定义必然合理吗?从直观角度看,它们各自有什么样的适用场合?

### 3. 平均失真度

失真度  $d(x_i, y_j)$  只能表示两个特定的具体符号  $x_i$  和  $y_j$  之间的失真,为了能在平均意义上表示信道每传递一个符号所引起的失真大小,我们定义平均失真度为失真函数的数学期望,即  $d(x_i, y_j)$  在随机变量  $X$  和  $Y$  的联合概率分布  $P(X, Y)$  中的统计平均值。

$$\bar{D} = E[d(x_i, y_j)] \quad (5-5)$$

由数学期望的定义可得:

$$\begin{aligned} \bar{D} &= \sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) d(x_i, y_j) \\ &= \sum_{i=1}^n \sum_{j=1}^m p(x_i) p(y_j | x_i) d(x_i, y_j) \end{aligned} \quad (5-6)$$

对于连续随机信源,定义平均失真度为:

$$\bar{D} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p_{x,y}(x, y) d(x, y) dx dy \quad (5-7)$$

对于长度为  $L$  的离散信源序列,平均失真度为:

$$\begin{aligned} \bar{D}(L) &= \sum_{i=1}^{n^L} \sum_{j=1}^{m^L} p(x_i, y_j) d(x_i, y_j) \\ &= \sum_{i=1}^{n^L} \sum_{j=1}^{m^L} p(x_i) p(y_j | x_i) d(x_i, y_j) \end{aligned} \quad (5-8)$$

信源序列的单个符号的平均失真度(有时也称为信源的平均失真度)为:

$$\begin{aligned} \bar{D}_L &= \frac{1}{L} \sum_{i=1}^{n^L} \sum_{j=1}^{m^L} p(x_i, y_j) d(x_i, y_j) \\ &= \frac{1}{L} \sum_{i=1}^{n^L} \sum_{j=1}^{m^L} p(x_i) p(y_j | x_i) d(x_i, y_j) \end{aligned} \quad (5-9)$$

对于试验信道的信源和信道均无记忆的长度为  $L$  的离散信源序列,单个符号的平均失真度为:

$$\bar{D}_L = \frac{1}{L} \sum_{l=1}^L E[d(x_{il}, y_{jl})] = \frac{1}{L} \sum_{l=1}^L \bar{D}_l \quad (5-10)$$

以上小写的  $x$  和  $y$  均表示一个具体的值,而不是随机变量,  $d$  代表针对具体符号的失真度,而  $\bar{D}$  则代表平均值。

如果信源和失真度一定,就只是信道统计特性的函数。信道传递概率不同,平均失真度也随之改变。

如果规定其平均失真度  $\bar{D}$  不能超过某一限定的值  $D$ ,即

$$\bar{D} \leq D \quad (5-11)$$

则  $D$  就是允许失真的上界。上式称为保真度准则(fidelity criteria)。把保真度准则作为对信道传递概率的约束,再求信道信息率  $R = I(X; Y)$  的最小值就有实用意义了,即在可以接受的失真范围内进行压缩。

## 5.2 信息率失真函数及其性质

要进行压缩,必须考虑信息压缩造成的失真是在一定的限度内的,因此这个编码长度应该在允许的失真范围内尽量小。从直观感觉可知,若允许失真越大,信息传输率可以越小;若允许失真越小,信息传输率需要越大。所以信息传输率与信源编码所引起的失真(或误差)是有关的,对信息进行压缩的效果与失真也是相关的。

### 5.2.1 信息率失真函数的定义

在允许一定失真  $D$  的情况下,信源可以压缩的极限应该是一个与失真相关的函数,可以定义信息率失真函数(information rate distortion function)为这一极限,简称率失真函数,记为  $R(D)$ 。

在给定信源的概率分布  $P(X)$  和失真度  $D$  以后,  $P_D$  是满足保真度准则  $\bar{D} \leq D$  的试验信道集合,即如果把  $X$  和  $Y$  当作信道的输入输出的话,这个信道集合中的信道的决定性参数就是信道传递(转移)概率  $p(y_j | x_i)$ 。在给定信源和失真度以后,信道的输入和输出的平均互信息  $I(X; Y)$  是信道传递概率  $p(y_j | x_i)$  的下凸函数,所以在这些满足保真度准则的  $P_D$  集合中一定可以找到某个试验信道,使信宿的信息量达到最小(可以证明为  $I(X; Y)$ ),而这个最小值可以从直观上理解为并且可以被证明为在保真度准则下的信源压缩极限,即信息率失真函数  $R(D)$ ,所以:

$$R(D) = \min_{P(y_j | x_i) \in P_D} \{I(X; Y)\} \quad (5-12)$$

或者直接表述为:

$$R(D) = \min_{P(y_j | x_i); \bar{D} \leq D} \{I(X; Y)\}$$

其中,  $R(D)$  的单位是奈特/信源符号或比特/信源符号。

上面定义的式子为限失真编码的压缩极限,可以利用渐进等分性来证明,本章后面部分会给出证明。

信息率失真函数这一命名也体现了信息的压缩极限是与允许的失真  $D$  相对应的一个函数,所以下面将会讨论这个函数的性质。

如果说试验信道的说法可能难以理解的话,可以将试验信道理解为限失真信源编码器的输入  $X$  和输出  $Y$  之间的一种概率上的映射关系,或者直接理解为概率  $p(y_j | x_i)$ 。

在离散无记忆平稳信源的情况下,可证得序列信源的信息率失真函数为:

$$R_N(D) = NR(D) \quad (5-13)$$

从数学上来看,平均互信息  $I(X; Y)$  是输入信源的概率分布  $P(x)$  的  $\cap$  型上凸函数,而平均互信息  $I(X; Y)$  是信道传递概率  $p(y_j | x_i)$  的  $U$  型下凸函数。因此,可以认为信道容量  $C$  和信息率失真函数  $R(D)$  具有对偶性。

研究信道容量  $C$  是为了解决在已知信道中传送最大的信息量。充分利用给定的信道,使传输的信息量最大而错误概率任意小,就是一般信道编码问题。研究信息率失真函数是为了解决在已知信源和允许失真度  $D$  的条件下,使信源必须传送给用户的信息量最

小。这个问题就是在可以接受的失真度  $D$  的条件下,尽可能用最少的码符号来传送信源消息,使信源的信息尽快地传送出去以提高通信的有效性。这是信源编码问题。

信息容量  $C$  和信息率失真函数  $R(D)$  之间的对应关系如表 5-1 所示。

表 5-1 信道容量  $C$  和信息率失真函数  $R(D)$  的比较

	信道容量 $C$	信息率失真函数 $R(D)$
研究对象	信道	信源
给定条件	信道转移概率 $p(y_j x_i)$	信源分布 $p(x)$
选择参数(变动参数)	信源分布 $p(x)$	试验信道转移概率或者信源编码器的映射关系 $p(y_j x_i)$
结论	求 $I(X;Y)$ 最大值	求 $I(X;Y)$ 最小值
概念上(反映)	固定信道,改变信源,使信息率最大(信道传输能力)	固定信源,改变信道,使信息率最小(信源可压缩程度)
通信上	在使得错误概率 $P_e \rightarrow 0$ 的限制下,使传输信息量最大——信道编码	在给定 $D$ 的限制下,用尽可能少的码符号传送——信源编码
对应定理	有噪信道编码定理	限失真信源编码定理

## 5.2.2 信息率失真函数的性质

下面讨论函数  $R(D)$  的性质,作为一个函数,其函数值取决于自变量,所以首先讨论关于它的自变量的取值范围,即定义域。

### 1. 信息率失真函数的定义域

$R(D)$  的自变量是允许平均失真度  $D$ ,它是人们规定的平均失真度  $\bar{D}$  的上限值。这个值不可以任意选取,这是因为平均失真度的值是受制约的,而且失真度与平均失真度均为非负值,显然满足下式:

$$0 \leq D_{\min} \leq D \quad (5-14)$$

$$D_{\min} = \sum_x p(x) \min_y d(x,y) \quad (5-15)$$

以上最小值的计算方法都是直接求各个失真度的最小值,然后按照概率加权平均,这是否正确?为什么?

(1) 最小值。对于离散信源,在一般的情况下可以采用前面的定义,当  $X$  和  $Y$  一一对应时,平均失真度为 0,而平均失真度显然不可能小于 0,所以  $D_{\min}$  为 0,此时,  $R(D_{\min}) = R(0) = H(X)$ 。

对于连续信源,  $D_{\min}$  趋向于 0 时,  $R(D_{\min}) = R(0) = H_c(X) = \infty$ 。

连续信源无失真的时候,传输的信息量是无穷大,实际信道容量总是有限的,无失真传送这种连续信息是不可能的。只有当允许失真( $R(D)$ 为有限值)时,传送才是可能的。

(2) 最大值。信源最大平均失真度  $D_{\max}$ : 必需的信息率越小,容忍的失真就越大。当  $R(D)$  等于 0 时,对应的平均失真最大,也就是函数  $R(D)$  定义域的上界值  $D_{\max}$  最大。

由于信息率失真函数是平均互信息量的极小值,平均互信息量大于或等于0,当 $R(D)=0$ 时,即平均互信息量的极小值等于0。满足信息率为0的 $D$ 值可能存在多个,但是鉴于我们总是希望失真度最小,存在多种选择时,总是选择最小值,所以这里定义当 $R(D)=0$ 时, $D$ 的最小值为 $R(D)$ 定义域的上限,即 $D_{\max}$ 是使 $R(D)=0$ 的最小平均失真度。

$R(D)=0$ 时, $X$ 和 $Y$ 相互独立,所以有:

$$p(y_j | x_i) = p(y_j) \quad (i=1,2,\dots,n)$$

满足 $X$ 和 $Y$ 相互独立的试验信道有许多,相应地可求出许多平均失真值,这类平均失真值的下界就是 $D_{\max}$ 。

$$\begin{aligned} D_{\max} &= \min_{p(y_j)} \sum_{i=1}^n \sum_{j=1}^m p(x_i) p(y_j) d(x_i, y_j) \\ &= \min_{p(y_j)} \sum_{j=1}^m p(y_j) \sum_{i=1}^n p(x_i) d(x_i, y_j) \end{aligned} \quad (5-16)$$

令

$$\sum_{i=1}^n p(x_i) d(x_i, y_j) = D_j$$

则

$$D_{\max} = \min_{p(y_j)} \sum_{j=1}^m p(y_j) D_j \quad (5-17)$$

上式是用不同的概率分布 $\{p(y_j)\}$ 对 $D_j$ 求数学期望,取数学期望中最小的一个作为 $D_{\max}$ 。实际上是用 $p(y_j)$ 对 $D_j$ 进行线性分配,并使线性分配的结果最小。

当 $p(x_i)$ 和失真矩阵已给定时,必可计算出 $D_j$ 。 $D_j$ 随 $j$ 的变化而变化。 $p(y_j)$ 是任选的,只需满足非负性和归一性。若 $D_s$ 是所有 $D_j$ 当中最小的一个,可取 $p(y_s)=1$ ,其他 $p(y_j)=0$ ,此时 $D_j$ 的线性分配值(或数学期望)必然最小,即有:

$$\begin{aligned} p(y_j) &= \begin{cases} 1, & j=s \\ 0, & j \neq s \end{cases} \\ D_{\max} &= \min_j D_j \end{aligned} \quad (5-18)$$

通俗地说,在进行最大限度地压缩的时候,极端的情况就是将输出端符号压缩为一个,可以将任意的信源符号 $x_i$ 都转换为一个相同的符号 $y_s$ ,由于对方接收到的符号是确定的,因此无须传递任何信息,或者说传递的信息量为0。对于不同的 $y_s$ ,会带来不同的失真度,我们当然会选择失真度最小的一个。

以上定义体现了在两个目标下的一种理性选择,人们追求对信源的最大压缩,同时也追求最小的失真度。当其中某个条件相同时,就会追求另外一个指标的最优化。比如信息率失真函数定义为最小值就是在相同的失真(或允许失真)的情况下追求最大的压缩(后面给出证明),而最大失真度的定义其实是在最大压缩的情况下追求最小的失真度。

在限失真编码中,还有哪些需求或者目标?

实际上,不是有意去进行理性的选择,平均失真度的值是可以超过这一值的。

由于 $R(D)$ 是非负函数,并且它是用从 $P_D$ 中选出的 $p(y|x)$ 求得的最小平均互信息

量,所以当  $D$  增大时,  $P_D$  的范围增大,所求的最小值不大于范围扩大前的最小值,因此  $R(D)$  为  $D$  的非增函数。当  $D$  增大时,  $R(D)$  可能减小,直至减小到  $R(D)=0$ ,此时对应着  $D_{\max}$ 。当  $D > D_{\max}$  时,  $R(D)$  仍然为 0。

可以得到下面的结论:

(1) 当且仅当失真矩阵的每一行至少有一个零元素时,  $D_{\min}=0$ , 一般情况下的失真矩阵均满足此条件。

(2) 可适当修改失真函数,使得  $\min_y d(x, y)=0$ 。

(3)  $D_{\max}$  和  $D_{\min}$  仅与  $p(x)$  和  $d(x, y)$  有关。

**例 5-1** 设试验信道输入符号集  $\{a_1, a_2, a_3\}$ , 各符号对应概率分别为  $1/3, 1/3, 1/3$ , 失真矩阵如下所示,求  $D_{\max}$  和  $D_{\min}$  以及相应的试验信道的转移概率矩阵。

$$(d_{ij}) = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \\ 3 & 2 & 1 \end{bmatrix}$$

解:

$$\begin{aligned} D_{\min} &= \sum_x p(x) \min_y d(x, y) \\ &= p(a_1) \min(1, 2, 3) + p(a_2) \min(2, 1, 3) + p(a_3) \min(3, 2, 1) \\ &= 1 \end{aligned}$$

令对应最小失真度  $d(a_i, b_j)$  的  $p(b_j | a_i)=1$ , 其他为 0, 可得对应  $D_{\min}$  的试验信道转移概率矩阵为:

$$[p(y | x)] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{aligned} D_{\max} &= \min_y \sum_x p(x) d(x, y) \\ &= \min \{ [p(a_1) \times 1 + p(a_2) \times 2 + p(a_3) \times 3], \\ &\quad [p(a_1) \times 2 + p(a_2) \times 1 + p(a_3) \times 2], [p(a_1) \times 3 + p(a_2) \times 3 + p(a_3) \times 1] \} \\ &= 5/3 \end{aligned}$$

上式中第二项最小,所以令  $p(b_2)=1, p(b_1)=p(b_3)=0$ , 可得对应  $D_{\max}$  的试验信道转移概率矩阵为:

$$[p(y | x)] = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

本例给出的是一种特异的失真矩阵,在输入和输出符号数目相等的时候,这种失真矩阵对应的输出符号是一种理性的选择吗?

**例 5-2** 离散二元信源  $p(x) = \left[ \frac{1}{3}, \frac{2}{3} \right]$ ,  $[D] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ , 求  $D_{\max}$ 。

解:

$$\left. \begin{aligned} D_1 &= \frac{1}{3} \times 0 + \frac{2}{3} \times 1 = \frac{2}{3} \\ D_2 &= \frac{1}{3} \times 1 + \frac{2}{3} \times 0 = \frac{1}{3} \end{aligned} \right\} D_{\max} = \min\left(\frac{2}{3}, \frac{1}{3}\right) = \frac{1}{3}$$

例 5-3 二元信源为  $\begin{bmatrix} x_1 & x_2 \\ 0.4 & 0.6 \end{bmatrix}$ , 相应的失真矩阵为  $\begin{bmatrix} \alpha & 0 \\ 0 & \alpha \end{bmatrix}$ , 计算  $D_{\max}$ 。

解: 先计算  $D_j$ 。由定义得  $D_1 = 0.4\alpha, D_2 = 0.6\alpha$ , 所以  $D_{\max} = \min(D_1, D_2) = 0.4\alpha$ 。

### 2. $R(D)$ 是关于平均失真度 $D$ 的(下)凸函数

设  $D_1$  和  $D_2$  为任意两个平均失真,  $0 \leq \alpha \leq 1$ , 则有:

$$R[\alpha D_1 + (1-\alpha)D_2] \leq \alpha R(D_1) + (1-\alpha)R(D_2) \quad (5-19)$$

证明: 当信源分布给定后,  $R(D)$  可以视为试验信道转移概率  $p(y|x)$  的函数, 即

$$R(D_1) = \min_{p(y|x) \in P_{D_1}} I[p(y|x)] = I[p_1(y|x)]$$

$$R(D_2) = \min_{p(y|x) \in P_{D_2}} I[p(y|x)] = I[p_2(y|x)]$$

且有:

$$\sum_{x,y} p(x)p_1(y|x)d(x,y) \leq D_1 \Rightarrow p_1(y|x) \in P_{D_1}$$

$$\sum_{x,y} p(x)p_2(y|x)d(x,y) \leq D_2 \Rightarrow p_2(y|x) \in P_{D_2}$$

令  $D_0 = \alpha D_1 + (1-\alpha)D_2, p_0(y|x) = \alpha p_1(y|x) + (1-\alpha)p_2(y|x)$ , 那么:

$$\begin{aligned} \sum_{x,y} p(x)p_0(y|x)d(x,y) &= \sum_{x,y} p(x)[\alpha p_1(y|x) + (1-\alpha)p_2(y|x)]d(x,y) \\ &\leq \alpha D_1 + (1-\alpha)D_2 = D_0 \end{aligned}$$

可知  $p_0(y|x)$  满足保真度准则  $D_0$ , 即  $p_0(y|x) \in P_{D_0}$ 。

$$\begin{aligned} R(D_0) &= \min_{p(y|x) \in P_{D_0}} I[p(y|x)] \leq I[p_0(y|x)] \\ &= I[\alpha p_1(y|x) + (1-\alpha)p_2(y|x)] \\ &\leq \alpha I[p_1(y|x) + (1-\alpha)p_2(y|x)] \\ &= \alpha R(D_1) + (1-\alpha)R(D_2) \end{aligned}$$

上式利用了平均互信息量是条件概率的下凸函数的性质。

### 3. $R(D)$ 是 $(D_{\min}, D_{\max})$ 区间的连续和严格的递减函数

证明:  $R(D)$  在定义域内为凸函数, 从而保证了连续性。下面证明在定义域内也是非增函数。由  $D_1 > D_2$  可得  $P_{D_1} \supset P_{D_2}$ , 在较大范围内求得的极小值一定不大于在所含小范围内求得的极小值, 所以  $R(D_1) \leq R(D_2)$ 。由于在定义域内  $R(D)$  不是常数, 而是非增的下凸函数, 可以通过反证法证明  $R(D)$  是严格递减函数。

$R(D)$  的非增性也是容易理解的。因为允许的失真越大, 所要求的信息率可以越小。

图 5-2 为  $R(D)$  函数的一般形式, 连续信源和离

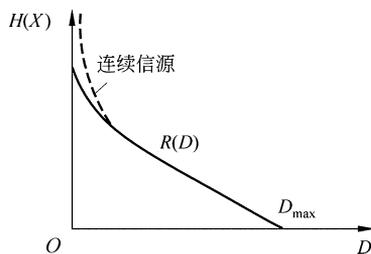


图 5-2  $R(D)$  函数的一般形式

散信源的信息率失真函数有所不同,图中虚线代表连续信源,实线代表离散信源。

综上所述,信息率失真函数  $R(D)$  的定义域为  $(D_{\min}, D_{\max})$ , 这一定义域为一种理性的选择条件下的定义域,实际上,可以让平均失真度超过该值。一般情况下输出端符号足够多且选择合理时,  $D_{\min} = 0, R(D_{\min}) = H(X)$ ; 当  $D \geq D_{\max}$  时,  $R(D) = 0$ ; 而当  $D_{\min} < D < D_{\max}$  时,  $H(X) > R(D) > 0$ 。通过信息率失真函数可以看出信源在允许一定失真的情况下的压缩潜力。

此外,如果将自变量和因变量颠倒过来,可得  $D(R)$ , 称为失真信息率函数,它是  $R(D)$  的逆函数。如何来理解和记忆两个函数的命名呢,信息率失真函数是因变量信息率随着自变量失真度函数关系的表示,失真信息率函数则颠倒过来。

**例 5-4** 若有一个离散、等概率单消息(或无记忆)二元信源:  $p(u_0) = p(u_1) = \frac{1}{2}$ , 且采用汉明距离作为失真度量标准,即  $d_{ij} = \begin{cases} 0, & u_i = u_j \\ 1, & u_i \neq u_j \end{cases}$  时,若有一具体信源编码方案为:  $N$  个码元中允许错一个码元,实现时  $N$  个码元仅传送  $N-1$  个,剩下一个不传送,在接收端用随机方式决定这个码元(为掷硬币方式)。此时,速率  $R'$  为:

$$R' = \frac{N-1}{N} = 1 - \frac{1}{N} \text{ (b/符号)}$$

抛硬币错误的概率为  $1/2$ , 所以:

$$D = \frac{1}{N} \times \frac{1}{2} = \frac{1}{2N}$$

所以信息率为  $R'(D) = 1 - \frac{1}{N} = 1 - 2 \times \frac{1}{2N} = 1 - 2D$ 。

若已知这一类信源理论上的信息率失真函数  $R(D) = H\left(\frac{1}{2}\right) - H(D)$  (后面将进一步给出计算), 则可以对两者进行比较。在图 5-3 中, 阴影范围表示实际信源编码方案与理论值之间的差距, 我们完全可以找到更好即更靠近理论值并且缩小阴影范围的信息源编码, 这就是工程界寻找好的信源编码的方向和任务。

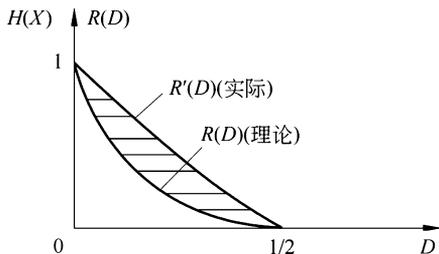


图 5-3 实际信息率与信息率失真函数值的对比