**第 2 次课 学时 2**

|  |  |
| --- | --- |
| **授课章节内容** | 第二章 单符号离散信源 |
|  | 2.1信源的数学模型 |
|  | 2.2信源符号的自信息量 |
|  | 2.3信源的信息熵 |
|  | 2.4信息熵的基本性质 |
| **教学目标** | 教学目标2 |
| **支撑毕业要求** | 毕业要求1-3 |

**教学要求：**

1. 知识目标

* 了解信源的数学模型及其分类
* 要求会计算信源的信息熵

掌握信息熵的基本性质能够正确解释信源的数学模型：能够正确解释自信息和信源熵的概念，能够计算信源熵，叙述并能正确解释信源熵的基本性质。

2. 能力目标

* 具备信源描述和离散信源熵计算的能力。

1. 素质目标

* 通过结合所学知识测算实际信源信息量的大小，培养工程科学素养。

**教学重点与难点**：

* 信源熵计算，平稳信源定义

**教学过程设计：**

* 引出信源熵并建立数学模型

**教学方法及手段：**

PPT为主，板书为辅

| **讲授与指导内容** | | **讲课、互动内容设计** |
| --- | --- | --- |
| 第2章 单符号离散信源  最简通信系统由信源、信道和信宿三部分构成。离散信源可用离散随机矢量来描述，其中每个随机变量都是取值离散的。当随机矢量只含一个离散随机变量时，这种信源就称为单符号离散信源，单符号离散信源是研究各类信源的基础。  2.1 单符号离散信源的数学模型  如果信源***X***符号集为*A*={*a*1,*a*2,…,*ar*}，*r*为符号集大小，信源符号对应某一概率分布， {*p*(*ai*)，*i*=1,2,…,*r*}，称此信源为单符号离散信源，信源空间数学模型可由式（2.1）描述。  (2.1)  其中，，  【例2.1】一个三元无记忆信源***X***，符号集*A*={0,1,2}，概率分布为：*P*(0)=1/2, *P*(1)=1/3, *P*(2)=1/6,写出信源空间模型。  解：信源空间模型如下式所示：    【例2.2】某箱子中共有32个质地均匀的球，其中，红球16个，黄球8个，蓝球和白球各4个。每次拿出一个球，拿出后又放回，各种彩球出现的先验概率分别为：    用随机变量*X*表示这个信源，其信源空间为：    由上述两个例子看出：信源空间包含两个基本要素：随机变量*X*的状态空间和概率空间，而概率空间又是决定性要素。概率可测是香农信息论的基本前提。  2.2 信源符号的自信息量  设信源发出某符号*ai*,由于信道中存在噪声或其它干扰，收端收到的是*ai*的某种变型*bj*，收信者收到*bj*后，从*bj*中获取关于*ai*的信息量，用*I*(*ai*; *bj*)表示，则  （2.2）  若信道是无噪无损信道，即一一对应信道，收信者收到*bj*就是*ai*本身，那么就完全消除了对信源符号*ai*的不确定性，即  （2.3）  把*I*(*ai*; *ai*)定义为信源符号*ai*的自信息量，并用*I*(*ai*)表示，即  （2.4）  信源符号的自信息量度量问题，相当于信源发符号的不确定性度量问题，是先验概率的函数：  （2.5）  根据自信息量定义所需的4个公理性条件，自信息量函数*I*(*ai*)定义如下：  （2.6）  信息量*I*(*ai*)随*P*(*ai*)变化的曲线如图2.1所示。  C:\Users\Galaxylyx\Desktop\2.1.png2.1  图2.1 自信息量*I*(*ai*)随*P*(*ai*)变化的曲线  自信息量的单位与对数的底有关  若以“2”为底，自信息量单位为“比特”（bit-binary unit），即  (比特)  若以“e”为底，自信息量单位为“奈特”（nat-nature unit），即  (奈特)  若以“10”为底，自信息量单位为“哈特”（Hart-Hartley），即  (哈特)  若以正整数“r”为底，自信息量单位为“r进制信息单位”，即  (r进制信息单位)  不同信息单位之间可以进行换算。本书如不加说明，信息度量默认为以“2”为底，且略去底数“2”不写。  香农信息量的度量，是信息论发展史上的里程碑，奠定了信息论发展成为一门学科的基础。  【例2.3】箱中80个红球，20个白球，现从箱中随机取出两个球。求:  (1)求解“两个球中红、白各一个”的不确定性  (2)求解“两个球都是白球”所提供的信息量  (3)求解“两个球都是白球”和“两个球都是红球”相比较，哪个事件发生的更难测？  解：设x为红球数，y为白球数，则：  ，  ，  ，  从上述计算结果来看，1.416bit大于0.195bit,显然,事件“两个球都是白球”要比事件“两个球都是红球”更难测。  【例2.4】设二元随机矢量，矢量中每个变量为独立同分布随机变量，且1符号的概率。求序列的自信息量。  解：  2.3 信源的信息熵  自信息量*I*(*ai*)是针对具体的特定事件*ai*在发生前存在的不确定度的度量，或发生后，提供给接收者的信息量。自信量的度量具有个体差异性，不具备对信源总体不确定度的度量能力。对信源***X***总体进行信息度量，即信源熵，它表征信源平均每个符号所含有的不确定性或平均输出一个符号提供给观测者的信息量，是自信息量的统计均值，即有  （2.7）  熵的单位即自信息量的单位，取决于对数的底。  【例2.5】（a）设有一信源***X***由两个事件*a1*，*a2*组成，其概率空间如下    则其信源熵为    （b）设有另一信源***Y***如下    则    可见有*H*(***Y***)>*H*(***X***)，即信源***Y***的平均不确定性要大于信源***X***的平均不确定性。直观地分析也易得出这样一个结论：信源中各事件的出现概率越接近，则事先猜测某一事件发生的把握性越小，即不确定性越大。从后述信息熵的极值性质，我们可知：当信源各事件等概率出现时具有最大的信源熵，也即信源的平均不确定性最大。  【例2.6】设二元信源***X***的信源空间为，其中，.  试计算当*p*=1/2；*p*=0(或*p*=1)时信息熵*H*(***X***)的值；  信息熵*H*(***X***)是概率分量的函数*H*(***P***)。画出*H*(***P***)的函数曲线，并指明函数*H*(***P***)的特性。  解：信源***X***的信息熵    这表明，二元信源***X***的信息熵*H*(***X***)是概率分量*p*的函数。  (1)当*p*=1-*p*=1/2时，二元信源称为二维等概信源，其信息熵    这表明，二元等概信源***X***，每一个符号（不论是“0”还是“1”）含有的平均信息量是1比特，每发一个符号（不论是“0”还是“1”）提供的平均信息量是1比特；每一个符号（不论是“0”还是“1”）存在的平均不确定性是1比特；收信者能确切无误的收到信源X的“0”或“1”时，收信者每收到信源***X***的一个符号，获取的平均信息量是1比特。  当*p*=0(或*p*=1)时，二元信源***X***是一个确知信源，其信息熵    因为，对任意小的正数，以“e”为底的对数可展开成级数    当时，的极限值等于零，即有    所以    这样，可合理的约定    则当*p*=0(或*p*=1)时，信源***X***的信息熵  这表明，以概率1发符号“0”或“1”的二元确知信源，不存在任何不确定性，不含有任何信息量。  （2）当时，取不同的值*P*，可计算出二元信源***X***的信息熵    相应地，得到*H*(*p*)函数曲线如图2.2所示  C:\Users\Galaxylyx\Desktop\2.2.png2.2  图 2.2 *H*（*p*）函数曲线  从图中可见，若二元信源***X***发符号“0”（或“1”）是确定事件，即*p*=0(或*p*=1)时时，则。这意味着确知信源***X***不提供任何信息量；若二元信源***X***以相同概率0.5发符号“0”或“1”，即*p*=1/2, *H*(*p*)=1。这意味着等概信源***X***每发一个符号提供的平均信息量达到最大值1比特/信源符号。若二元信源***X***发“0”（或“1”）的概率*p*(0<*p*<1)越接近1/2，则发“1”（或“0”）的概率1-*p*亦越接近1/2.若“0”和“1”的概率越接近则*H*(*p*)的值越大。这意味着二元信源***X***的平均不确定性越大，每符号（不论是“0”还是“1”）含有的平均信息量越大；相反，二元信源***X***发“0”（或“1”）的概率*p*(0<*p*<1)越远离1/2，则发“1”（或“0”）的概率1-*p*亦越远离1/2.若发“0”和“1”的概率相差越大，则*H*(*p*)的值就越小。这意味着二元信源***X***的平均不确定性越小，每符号（不论是“0”还是“1”）含有的平均信息量就越小。这体现出*H*(*p*)是概率分量*p*的上凸形函数的基本特性。  【例2.7】在一个箱子中装有*m*个黑球和(*n*-*m*)个白球。设实验***X***：随机的从箱子中取出一个球而不再放回箱子；实验***Y***：从箱子中取出第二个球。  计算实验***X***所获取的平均信息量  若实验***X***摸取的第一个球的颜色不知道，计算实验B所获取的平均信息量。  解：令W代表白球；B代表黑球。  设实验***X***中取出的球是黑球的概率为*P****X***(B)；实验***X***中取出的球是白球的概率为*P****X***(W)，则有      信源***X***的信源空间可表示为    实验***X***所获取的平均信息量，就是信源的信息熵    (2)若实验***X***中取出的第一个球是白球，令实验***Y***中取出的白球的概率为P***Y*** (W/W)；取出黑球的概率为P***Y*** (B/W)，则有      若实验***X***中取出的第一个球是黑球，令实验***Y***中取出白球的概率为P***Y***(W/B)；取出黑球的概率为P***Y***(B/B)，则有      设实验***Y***中出现白球的概率为P***Y***(W)；出现黑球的概率为P***Y***(B)，则有      信源***Y***的信源空间可表示为    实验***Y***所获取的平均信息量，就是信源***Y***的信息熵    这说明，实验***X***和实验***Y***的平均不确定性或获取的平均信息量是相等的。  2.4 信息熵的性质  在介绍熵的性质之前，给出几个在信息论证明中有用的几个定义和定理。并将单个随机变量的熵推广到两个或者多个变量熵的情形。  2.4.1 联合熵与条件熵  定义2.4.1 对于服从联合分布为*p*(*x*,*y*)的一对离散随机变量(*X*,*Y*)，其联合熵*H*(*X*,*Y*)（joint entropy）定义为  （2.8）  上式亦可表示为  （2.9）  也可以定义一个随机变量在给定另一随机变量下的条件熵，先求条件分布熵；然后对条件分布熵再求统计均值既得条件熵。  定义2.4.2 若(*X*,*Y*)~*p*(*x*,*y*)，条件熵（conditional entropy）*H*(*Y*|*X*)定义为：  （2.10）  联合熵和条件熵定义的这种自然性可由一个事实得到体现，它就是一对随机变量的熵等于其中一个随机变量的熵加上另一个随机变量的条件熵。其证明见如下定理。  定理2.1（链式法则）  （2.11）  【证明】  （2.12）  等价地记为：  （2.13）  等式的两边同时取对数期望，即得本定理。  推论  （2.14）  2.4.2相对熵  若和为定义在同一概率空间的两个概率测度，定义相对于的相对熵为  （2.15）  相对熵又称散度、鉴别信息、方向散度、交叉熵、Kullback\_leibler距离等。注意，在上式（2.15）中，概率分布的维数不限，可以是一维，也可以是多维，也可以是条件概率。  在证明下面的定理前，首先介绍一个在信息论中常用的不等式。  对于任意正实数x，下面不等式成立：  （2.16）  实际上，设*f*(*x*)=ln *x*-*x*+1，可求得函数的稳定点为*x*=1，并可求得在该点的二阶导数小于0，从而可得*x*=1为*f*(*x*)取极大值的点，即*f*(*x*)=ln *x*-*x*+1≤0，仅当*x*=1时，上式（2.16）右边等号成立。令*y*=1/*x*，可得1-1/*y*≤ln y,再将*y*换成*x*，就得到左边的不等式。  定理2.2　如果在一个共同有限字母表概率空间上给定两个概率测度*P*(*x*)和*Q*(*x*)，那么  （2.17）  仅当对所有*x*，*P*(*x*)=*Q*(*x*)时，等式成立。  【证明】因*P*(*x*),*Q*(*x*)≥0，，所以根据式（2.16），有  （2.18）  仅当对所有*x*，*P*(*x*)=*Q*(*x*)时，等式成立。  式（2.17）称为散度不等式（divergence inequality），该式说明，一个概率测度相对于另一个概率测度的散度是非负的，仅当两测度相同时，散度为零。散度可以解释为两个概率测度之间的“距离”，即两概率测度不同程度的度量。不过，散度并不是通常意义下的距离，因为它不满足对称性，也不满足三角不等式。  【例2.8】设一个二元信源的符号集为{0,1}，有两个概率分布*p*和*q*，并且*p*(0)=1-*r*, *p*(1)=*r*, *q*(0)=1-*s*, *q*(1)=*s*,求散度*D*(*p*//*q*)和*D*(*q*//*p*)，并分别求当  *r*=s和*r*=2s=1/2时散度的值。  解：根据式（2.15），得      当*r*=s时，有  当*r*=2s=1/2时，有      定理2.3 （熵的不增原理）  （2.19）  【证明】设，那么    上面利用了散度不等式，仅当*X*，*Y*相互独立时，等式成立。  式（2.19）表明，条件熵总是不大于无条件熵，这就是熵的不增原理：在信息处理过程中，已知条件越多，结果的不确定性越小，也就是熵越小。  2.4.3 Jensen不等式及其结果  定义2.4.3　若对于任意的及，满足  （2.20）  则称函数在区间上是凸的（convex）。如果仅当或，上式成立，则称函数f是严格凸的（strictly convex）。  定义2.4.4　如果-*f*为凸函数，则称函数*f*是凹的。如果函数总是位于任何一条弦的下面，则该函数是凸的；如果函数总是位于任何一条弦的上面，则该函数是凹的。  定理2.4（Jensen不等式）若给定凸函数*f*和一个随机变量*X*，则  （2.21）  进一步，若*f*是严格凸的，那么式（2.21）中的等式蕴含的概率为1（即*X*是个常量）。  【证明】我们只证明离散分布情形，且对分布点的个数进行归纳证明。当*f*为严格凸函数时，等号成立条件的证明留给读者。对于两点分布，不等式变为  （2.22）  这由凸函数的定义可直接得到。假定当分布点的个数为*k*-1时，定理成立，此时记，则有  （2.23）  2.4.4信息熵的基本性质  1.对称性  概率矢量中，各分量的次序任意改变，熵不变，即  （2.24）  其中，是1,2，，*n*的任何一种*n*级排列。该性质说明熵仅与随机变量总体概率特性（即概率分布）有关，而与随机变量的取值及符号排列顺序无关。  2.非负性  （2.25）  仅当对某个时，等式成立。  因为自信息量是非负的，熵为自信息的平均，所以也是非负的。不过，非负性仅对离散熵有效，而对连续熵来说这一性质并不成立。  3.确定性  （2.26）  这就是说，当随机变量集合中任一事件概率为1时，熵就为0。这个性质意味着，从总体来看，事件集合中虽含有许多事件，但是如果只有一个事件几乎必然出现，而其他事件几乎都不出现，那么，这就是一个确知的变量，其不确定性为0.  4.扩展性  （2.27）  利用可得到上面的结果，其含义是，虽然小概率事件自信息大，但在计算熵时所占比重很小，可以忽略。  5.极值性  定理2.5（离散最大熵定理）对于有限离散随机变量，当符号集中的符号等概率发生时，熵达到最大值。  【证明】设随机变量有*n*个符号，概率分布为;为等概率分布，即。根据散度不等式有  （2.28）  即，  当且仅当等概分布时取等号。  注意：离散最大熵定理仅适用于有限离散随机变量，对于无限可数符号集，只有附加其他约束求最大熵才有意义。  6.上凸性  是概率矢量的严格的上凸函数。  这就是说，若，那么，其中均为*n*维概率矢量，。该性质可用凸函数性质（1）来证明（提示：先证明是严格上凸的）。  7.一一对应变换下的不变性  离散随机变量的变换包含两种含义，一是符号集中符号到符号的映射，二是符号序列到序列的变换。首先研究第一种情况。设两随机变量*X*,*Y*，符号集分别为*A*,*B*，其中*Y*是*X*的映射，可以表示为*A*→*B*，*x*→*f*(*x*)。因此有  （2.29）  所以；，而另一方面，所以，仅当是一一对应映射时等号成立，此时。应用类似的论证也可推广到多维随机矢量的情况，因此得到如下定理。  定理2.6 离散随机变量（或矢量）经符号映射后的熵不大于原来的熵，仅当一一对应映射时熵不变。  【例2.9】设二维随机矢量***XY***，其中***X***,***Y***为独立同分布随机变量，符号集为*A*={0,1,2}，对应的概率为{1/3,1/3,1/3}，做变换*u*=*x*+*y*，*v*=*x*-*y*，得到二维随机矢量***UV***；求*H*(***U***)，*H*(***V***)，*H*(***UV***)。  解：***U***、***V***取值空间如表1、表2所示。  表1 ***U***取值   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***U*=*X*+*Y*** | | ***Y*** | | | | 0 | 1 | 2 | | ***X*** | 0 | 0 | 1 | 2 | | 1 | 1 | 2 | 3 | | 2 | 2 | 3 | 4 |   ***U***的符号集为{0,1,2,3,4}              表２　***V***取值   |  |  |  |  |  | | --- | --- | --- | --- | --- | | ***V*=*X***-***Y*** | | ***Y*** | | | | 0 | 1 | 2 | | ***X*** | 0 | 0 | -1 | -2 | | 1 | 1 | 0 | -1 | | 2 | 2 | 1 | 0 |   ***V***的符号集为{-2,-1,0,1,2}              因为是一一对应变换，所以    看到，所以***U***，***V***不独立。  本章要点  1.单符号离散信源数学模型    2.定义 自信息量函数    3.定义 信源的信息熵  (比特/信源符号)  4.定义 联合熵H(X,Y)    5.定义 条件熵    6.链式法则    7.定义 P相对于Q的相对熵    8.Jensen不等式  若给定凸函数*f*和一个随机变量*X*，则  9.信息熵的性质  对称性  非负性  确定性  扩展性    极值性  上凸性  是概率矢量***p***的严格上凸函数。  变换不变性 离散信源经过一一对应变换后熵不变 | | 提问：信源是什么？  如何描述？  此部分内容讲解注意让同学与所学概率内容相结合  统计独立的信息量之和怎么求解？与各个独立的信息量有什么关系  注意：公式推导中的符号表示。  提问:提出信息熵的背景是什么？  再次强调： 概率与信息量之间有什么关系？  这是一个典型例题，要精讲。  拓展讲解相对熵的应用场合。  提问：多维情况下，凸函数定义的特点？  注意：讲解此部分信息熵的几种基本性质要将理论推导和感性认识相结合。  简要复习本章内容 |
| **小结** | 信源进行分类，信源数学模型，信息熵的计算 | |
| **复习要点** | 离散信源的信息熵 | |
| **思考题** | 1信源平均每个消息携带多少信息量？  2整个信源能输出多少信息量？ | |
| **作业题** | 2.2 2.4 | |

作者签名：

****