信息論基础

北京科学教育編輯室

1962年5月

信息論基础

*

出版者：北京科学教育編輯室
印刷者：中国人民解放軍535工厂
787×1092毫米 1/16 印張5
1962年5月第二版
定价：0.65元

目　　录

緒　論 …… 3
第一章　信源及信息量 …… 7
§1-1　信源的数学描述 …… 7
§1-2　$H(\ )$的推导 …… 9
§1-3　常数C的决定 …… 18
§1-4　$H(\ )$的一些特性 …… 20
习　题 …… 23
第二章　信道及信道容量 …… 24
§2-1　信道的数学描述 …… 24
§2-2　傳輸率 …… 24
§2-3　$P(e)$与輸入$P(\ )$，A_s的关系 …… 28
§2-4　信道容量 …… 29
习　题 …… 34
第三章　編碼器及編碼問題 …… 36
§3-1　編碼器 …… 36
§3-2　編碼效率 …… 36
§3-3　无干扰編碼定理 …… 38
§3-4　理想二进編碼 …… 39
习　题 …… 41
第四章　无記忆信道 …… 42
§4-1　編碼定理 …… 42
§4-2　最大集与扩充集 …… 42
§4-3　定理的証明 …… 44
习　题 …… 49
第五章　有記忆信道 …… 50
§5-1　信源及其$H(\ )$率 …… 50
§5-2　信道及其容量 …… 52
§5-3　編碼定理 …… 54
§5-4　$A.E.P$的証明 …… 56
第六章　連續信道 …… 60
§6-1　連續分布的$H(\ )$ …… 60
§6-2　$H(\ )$的一些基本特性 …… 61
§6-3　$H(\ )$的最大値 …… 62
§6-4　平均功率受限制的信道容量 …… 65

附录 1，2，3，4，5，6 ……67—74
附表 1 ……78
附表 2 ……78

緒　　論

自从世界上有了人类，我們就有通信，随着科学的发展，通信方法也愈来愈进步。在一百年以前，电通信已經开始被利用，虽然早期的电通信只是极其簡单的有綫电报，但第一部电报机出現20年后，双工电报、电話等通信方法即陆續被提出。1894年 $A.O.$ 波波夫創立第一个雷雨指示器，通信便进入到无綫电时代，加上二极管、三极管于1904年、1906年相继問世，在这50多年中，电信事业有了极大的发展。新的方法不断出現，如微波、长距离波导、对流层、流星、人造卫星……都可以用来进行通信。应用的領域不断扩充，指战員可以在后方看到前方作战情况，病人可以吞下診断器向医生报告病情，紅色火箭可以飞到月球拍摄背面图形再傳回地球，……

在通信发展一日千里，通信种类不胜枚举的今天，能不能找到一个衡量通信系統好坏的标准？可不可以对不同的通信系統作比較？有沒有理想的通信系統？……变成了急切希望解决的問題。信息論企图根据我們已經累积起来的經驗，从理論上研究这些問題。虽然，現在的信息論还远远不能全部解决我們在通信中希望解决的問題，但它的应用却大大超出了通信的范圍。例如在自动計算、自动控制、文字改革、生理学、物理学等方面都得到了初步应用，而且正在不断发展着。

不論那一种通信系統，它的好坏总是可以用它的“数”、“质”两方面来衡量。一般說来：所謂“数”，可以理解为通信系統的效率，也就是单位时間所能做的“工作”量；所謂“质”，可以理解为通信系統的可靠性。但如果进一步追問，例如电話和电視的效率那一个高？电报和广播的可靠性那一个大？我們却很难作出明确的答复。所以这样，有两个主要原因：

1. 以前我們研究通信系統的时候，电話是电話，电視是电視。电話通信是用送話器、受話器，电視却借助于摄像器、显像器。各有一套，自然不易比較。

2. 即使对同一种通信系統，也沒有明确标准可以衡量它的效率和可靠性，因此更談不上这个系統和那个系統的比較了。如果我們能够定出一个系統的效率为5，另一个系統的效率为7，那么就可以說后者的效率高于前者。

信息論研究通信的方法是：

（一）建立一个能够代表各种通信系統的物理模型，作为研究的对象。

（二）从量方面研究这个模型的效率和可靠性，这也就是信息論的主要內容。

以下我們将分别闡述这两个問題。

（一）根据我們一百多年的經驗，希望建立一个能够代表各种类型通信系統的模型，并不是一件很困难的事情。图1就是这样的一个模型。

通信的开始点是信源。信源的輸出是消息，文字、数字、音乐、話音、脉冲等等都可以看成是消息。消息經过編碼器变为信号，莫尔斯电碼的点、划，編碼調制的电碼組，調幅、調頻、調相的脉冲，单边带、双边带的連續波等等都可以看成是信号。明綫、电纜、波导等等都可以看成是信道。信号在信道中傳輸的时候，受到了干扰。电弧产生的火花、太阳黑子

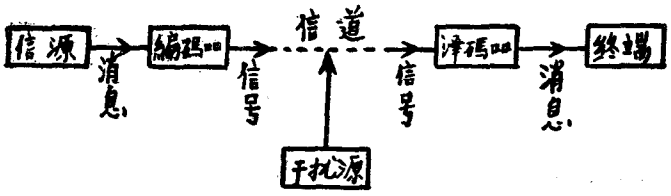

图 1

的影响、波长相近的广播、导綫的热騷动等等都可以看成是干扰。发送的信号連同干扰到达接收端，由譯碼器将它还原为消息（当然，还原的消息很可能已經不是发送的消息了）。送进接收終端后，全部通信过程就算告一段落。由于我們采用了一些含义較广的名詞，不难看出，图 1 的确能够代表各种类型的通信系統。其实它还可以代表不是电的通信系統。例如，消息是某甲的思想，信号是表达这个思想的語言；又例如，消息是中文語言，信号是外国語言等等。我們不准备对构成图 1 的各部分給出物理定义，而将从数学上加以描述，这样可以不失其广泛的代表性。

（二）要从量方面研究这个模型，首先应該解决图 1 的作用，信息論是根据这样一个概念出发的，任何一种通信之所以需要，原因只在于接收終端事前（收到消息之前）不知道发送的消息是什么。如果眞能事先确实知道，通信就大可不必了。例如某甲到达北京后，发一張电报給某乙，电文是“平安到达北京”。所以要有这次通信，是某乙在未收到电报之前不知道某甲是否“平安”，也不知道某甲“到达北京”可以換一种說法；某甲希望通过电报，消除某乙不知道的事情而使他知道，这就是通信的目的。拿数学上的术語来說：接收終端对发送的消息具有先驗不肯定性。

通信的目的就是要解除接收終端的这种先驗不肯定性。

不肯定性如何被解除呢？唯一的办法是用某种方式告訴对方（即进行通信），至于采用何种方式？只是手段問題而已，其目的相同。明确了解了模型的作用，方能进而从数量上加以研究。現在我們已經能看出，如果有一个数量足以反映解除了多少不肯定性，則它自然可以衡量各种通信系統的好坏。还是先让我們看看上面电报的例子，假如电报只有“平安”兩字，虽則部分地解除了某乙的不肯定性，但他对某甲到达何处，依然不清楚，考其原因，可以解釋为电报“內容”不够充足。“內容”究竟指什么而言？一般可以看成是“情报”，不过內容也好，情报也好，它們的眞正含义并非顧名可以思义的，还得有一番解釋。所以信息論中用了另外一个名詞“信息”去代替它們。以后我們对信息量（信息的数量）比較詳細地論述。

綜上所述：接收終端对发送消息具有一定数量的不肯定程度，但当終端从通信系統收到足够的信息量后，不肯定性随之消除；从而能够肯定发送的消息。

現在关鍵問題在于如何找到反映不肯定程度的量。如果我們对任一种随机現象进行观察，虽則事先总是沒有办法完全肯定它出現的情况，但不肯定程度却有所不同。例如有一种随机現象，它全部可能出現的情况共有 n 种 x_1，…，x_n，各 x（情况）出現的概率分别为 $P(x_1)$，…，$P(x_n)$。显然，对那一个 x 出現的不肯定程度决定于（i）各 $P(x)$ 取什么值，(ii) n 的大小。例如同样是 $n=2$，我們对胎儿是否有兩只手的不肯定程度将远远小于是否女孩的不

肯定程度。又例如同样是 $P(x_1)=\cdots=P(x_n)$，一种随机現象是在两个号碼中任意抽出一个，另一种是在 10000 个号碼中任意抽出一个，我們对出現什么号碼的不肯定程度，显然前者小于后者。

随机現象的全部可能情况 x_1，…，x_n 連同这些 x 的出現概率 $P(x_1)$，…，$P(x_n)$ 称为描述这个随机現象的概率空間。怎样找到用以消除对某概率空間的不肯定性所必須收到的信息量？或者簡单地說，怎样找到一个足以反映概率空間不肯定程度的量？是信息論的重要課題之一，我們将在第一章中加以詳細討論。上面举了一些例子只是希望我們从概念上承认对概率空間的不肯定程度的确有量存在，这就够了。有些文献称这个量为“熵”（讀商），它的笔划太多，写在公式里面很不方便，所以我們即使称为熵，也还需另找一个符号去代表它，以便于公式的书写。因而以后我們直接規定代表这个量的符号为 H 而不另用新字命名。

对接收端来說，信源輸出是一种随机現象。例如它可以是这个或那个声音，可以是这个或者那个单字……，所以在信息論中把信源看成是一个概率空間。消息是以某种規律构成的一串、一串的 x。由于我們可以任意規定 x，显見这样定义的信源和消息具有极其广泛的代表性。

同样地对接收終端来說，編碼器輸出也是一种随机現象，因此編碼器的作用只是在它的輸入概率空間和輸出概率空間之間給出两者的对应关系。例如我們用 ·-，··-，···- 等分别去对应 1，2，3 等。当然，編碼不应該狹义地理解成电报中的編制电碼，应該从广义方面解釋，例如以不同的脉冲幅度对应不同的声压，以不同的頻率对应不同濃淡程度等都屬于編碼。第一、第三章将分别对信源及編碼器作进一步的討論。

既然終端对信源的不肯定程度的量是 H，不难想像只要收到的信息量也是 H，那么不肯定性就全部消失，通信目的就被完成。終端收到信息是通过信道来完成的。信道負有傳輸信息的任务，它的好坏对通信系統起决定性的作用。所以我們用了第四、第五、第六章分别討論不同情况的信道，其实以后我們将看到第三章的編碼器也可以說是一种信道。

我們可以进一步說明早先提到过的通信系統的效率和可靠性問題。从信息論观点看，所謂效率就是傳送信息量的效率。粗略地說，单位时間內消除終端的不肯定性愈多，效率愈高。所謂可靠性是指抵抗干扰的能力，干扰能够使終端即使收到信号后，仍然保留某些不肯定性。保留的不肯定性愈少，抗干扰能力就愈强。由于不肯定性能够用数量来反映，所以效率和可靠性也就能够用数量来反映，随之我們可以对各类通信系統进行比較。而且我們还能够找出理想情况的通信系統用以和現行系統比較。不过信息論尚不能解决如何实現理想的通信系統这个問題。

上面我們只着眼于不肯定性，而忽略了，至少不是全面地考虑了信号的特性。因为任意一种信号都具有下列三个特性：

1. 占用一定的頻带 F；
2. 占用一定的时間 T；
3. 要求一定的超扰值 $\log\frac{P_s}{P_n}$ 其中 P_s 为信号功率，P_n 为噪声功率。

如果让 F、T、$\log\frac{P_s}{P_n}$ 分别是欧几里德三度空間的三个量，那么它們的乘积 $FT\log\frac{P_s}{P_n}$ 是一个体积称为信号体。如果我們拿单位信号体所能傳輸的信息量去衡量效率，拿单位

信号体的抗干扰能力去衡量可靠性，就能够兼顾信号特性和不肯定性，这是比較好的办法。不过这里討論的仍然主要是不肯定性的問題。

一言以蔽之，今后各章中我們主要从不肯定性的观点出发，研究通信系統的两个基本問題——效率和可靠性。

第一章　信源及信息量

我們在緒論中已經提到过：信源是信息的来源；信源的輸出是一种随机現象；通信的目的是要使接收者能够收到足够的信息量来或多或少地解除他对信源輸出的不肯定性。不过，单凭这些概念性的描述，很难再作进一步的研究。因此本章首先从数学上給信源一些基本的描述；然后确定一个能够反映信源不肯定程度的量$H(\ \ )$，也就是用以解除对信源不肯定性所需的信息量；最后給出$H(\ \ )$的“单位”和它的一些基本特性。

§1.1　信源的数学描述

在描述之前，先說明一些符号和几个名詞的含义。

对随机現象每做一次試驗，总可以得到一个結果，这个結果我們称之为基本事件。所有基本事件的全体称为基本事件集或者基本空間。今后将用$X=\{x\}$表示集X的元素为x，用$x\in X$表示x屬于集X，更詳細的表示方法是

$$X:\ x_1,\ x_2,\ \cdots,\ x_n。$$

对一个随机現象的研究，固然要知道它全部可能的結果，同时这些結果以什么概率出現也是非常重要的。我們用$P(x_i)$表示x_i这个事件出現的概率，而用$P(\ \ )$表示$P(x_1)$，$P(x_2)$，$\cdots$，$P(x_n)$这几个基本事件出現概率的全体。由于我們規定基本事件集X包括全部可能的結果，显然$\sum_{i=1}^{n}P(x_i)=1$。为了书写方便，有时把它写成$\sum_{X}P(x)=1$，$\sum_{X}P(x)$的意思是要我們将x取遍X所包含的基本事件。$P(\ \ )$虽然可以取这样、那样的一些值，但总不外是如何把1分配給各个x的問題，因此$P(\ \ )$称为概率分布。再明确一点，可以說是X的分布。

基本事件集連同其概率分布称为概率空間，記为X，$P(\ \ )$。詳細一点的表示方法可以是

$$X,\ P(\ \ ):\ \begin{matrix} x_1,\ \cdots,\ x_n \\ P(x_1),\ \cdots,\ P(x_n)。\end{matrix}$$

概率空間是随机現象的描述，因此我們也可以把它看成是对信源的描述，或者說：从数学上看，信源是概率空間。

現在让我們来举一个概率空間作为信源的例子。基本事件集由26个英文字母及1个間隔符号共27个元素构成，它們的出現概率如表1所示。这27个元素連同27个概率所构成的概率空間是一个信源，它以表1的概率发出字母（包括間隔符号）。

显而易見，由上述信源一个接着一个地发出字母而形成的消息，它不成其为英文，至少信源不懂文法，也不懂修辞，它只会以$P(\ \ )$发出字母。这就等于我們用一个布袋，里面放进10000个对人手感覚完全一样的球，其中2000个写上間隔，1050个写上e字，$\cdots$，10个写上Z字。現在我們随手摸一个球出来，把球上的字母記下，然后把摸出的球放回布袋；攪匀后再摸一个出来，挨次記下球上的字母。如此继續下去，虽則可以得到一串字母序列，但在这一串字母序列中，我們会发現很多既不能发音，也不能从字典中找到的“字”。

信源为什么不懂文法、修辞？从数学上应作何解釋？其实，我們忽略了信源一个接着一

英文字母出現概率表　　表 1

英文字母	出現概率	英文字母	出現概率
間隔	0.2	u	0.0225
e	0.105	m	0.021
t	0.072	p	0.0175
o	0.0654	y	0.012
a	0.063	w	0.012
n	0.059	g	0.011
i	0.055	b	0.0105
r	0.054	v	0.008
s	0.052	k	0.003
h	0.047	x	0.002
d	0.035	j	0.001
l	0.029	q	0.001
c	0.023	z	0.001
f	0.0225		

个地发出字母是一种过程，在这过程当中，$P(\ \ )$不是固定不变的，应該根据前一段时間已經发出的字母而改变后一段时間的概率分布。例如：已經发出一个間隔符号后，接着再出現一个間隔的概率就不应該是表 1 中的 0.2，而是 0；已經出現一个 t 再出現一个 t 的概率也不应是 0.072，而是远远小于 0.072。因此，假如我們希望信源发出的消息是实际英文的話，那么可以拿一个随机过程去代表信源，不应拿上述的基本概率空間作为信源。以后我們会談到怎样用一个随机过程来描述信源，現在我們可以換用一种比較簡单的方法，看看如何能够使到消息比較接近实际的英文。

如果我們用英文单字代替字母作为基本事件，那么信源发出的消息，虽然还不是英文，但至少每一个字是英文字，能够发音，也能够在字典中找到。进一步，如果用英文句子作为基本事件，即使消息語无倫次的，但一句、一句单独地看，却是文理通順的。继續下去，不难想像可以得到和英文很接近的消息。

上面的說明，可以用重复空間的概念加以表示。設有两个基本空間

$$X：x_1，\cdots，x_n \text{ 及 } Y：y_1，\cdots，y_m，$$

它們的联合空間記为$X\otimes Y$，

$$X\otimes Y：x_1y_1，\cdots，x_1y_m；\cdots；x_ny_1，\cdots，x_ny_m。$$

其中 xy 表示 x 和 y 同时发生。（也常写成 x，y）联合空間共有 nm 个元素，它的概率分布記为 $P(,)$。联合空間連同它的概率分布称为联合概率空間，記为$X\otimes Y$，$P(,)$。联合事件与基本事件的概率有下面关系，

$$P(x,y)=P(x)P(y/x)。$$

式中的 $P(y/x)$表示 x 已經出現后，y 出現的概率。当而且只当X与Y在概率上无关的时候，$P(y/x)$才等于 $P(y)$，$P(x,y)=P(x)P(y)$才能成立。

如果Y就是X，联合空間变成$X\otimes X$，称为X的重复空間。重复空間可以由两个或者很多个（甚至无穷多个）同一的基本空間自相联合而成。n 个X的重复空間記为$\overset{n}{\otimes}X$。

要是英文单字永不会超过K个字母，X是以字母为其基本事件的空間。（不妨称为字母

空間）K 个 X 的重复空間可以代表以英文单字为其元素的空間。（不妨称为单字空間）因为，虽然重复空間具有 n^k 个元素，可能远远超过英文单字的数目，但如果我們令凡是字典中沒有的元素的概率为 0，則信源发出的消息，自然都是英文单字，能在字典中找到。由此看来，重复概率空間代表的信源，能够发出比較接近英文的消息。要注意的是，这并不意味着我們不应該用字母空間代表信源了。今后我們会看到以字母空間代表信源仍有其重要的作用。

§1.2 H( )的 推 导

有了信源的数学定义之后，現在我們着手研究如何找到一个能够反映信源不肯定程度的量，也就是希望从数量上确定概率空間的不肯定程度。有些文献称这个量为熵，但我們今后将記为 $H(\ )$，而不給予特殊名称。本节准备根据一些对不肯定性的直观想法推导出 $H(\ )$。

先从随机事件开始，一个随机事件所以具有不肯定性，是因为我們事前无法决定它是否出現——除非我們被某种方法所告知。說得詳細一点：除非我們被某种方法供給一定数量的情报（信息），否則无法决定它是否出現。虽然，如何确定这个信息的数量？現在尚知道得很少。但只凭直观的想法，也不难承认这个量与随机事件 x 的出現概率 $P(x)$有关，因此可以将信息量看成是 $P(x)$的函数，記为 $I[P(x)]$。为了书写方便，以后写成 $I(P_x)$。当然，$I(P_x)$就是用以解除我們对 x 是否出現这个不肯定性所需的情报数量。目前所知，仅此而已，甚至当 $P(x)$增加（减小）的时候，$I(P_x)$是增加还是减小？我們也不易回答。初步想来，当$P(x)$很接近于 1 的情况，$I(P_x)$应該很小，因为 x 已經几乎肯定出現，所以不肯定性就几乎沒有了。如果 $P(x)=1$，就是說 x 为必然事件，这时不肯定性应該不存在。不过，让我們来看看另一种情况，当 $P(x)$很小的时候，似乎 $I(P_x)$应該很大，因为 x 的出現很不肯定的。可是当 $P(x)=0$，x 变成不可能事件，所謂不可能事件也就是我們能够肯定它不会出現的事件，这样看来，不肯定性又好像应該等于 0。到底$I(P_x=0)$应該是很大呢？还是等于 0？这个疑問不能单从“$I(P_x)$与 $P(x)$有关”这样一个簡单的想法得到解答，所以我們暫时把 $P(x)=0$ 的情况除外，留待以后对不肯定性的别的想法上加以解决。

为了便于直观地比較各个事件的不肯定程度，自然我們希望 $I(P_x)$ 是个非負实数。至于 x 是什么？在討論 $I(P_x)$的时候，不是我們关心的事情，可以不必过問。

进一步，假設有一基本概率空間

$$X,\ P(\ ):\ \begin{matrix} x_1, & \cdots, & x_i, & \cdots, & x_n \\ P_1, & \cdots, & P_i, & \cdots, & P_n, \end{matrix}$$

其中 $P_1+\cdots+P_n=1$。虽然我們已經想到用 $I(P_i)$来表示事件 x_i 的不肯定程度，但空間中每一个事件的概率未必相等，如何找一个量来反映这个空間的不肯定程度？这是我們接着要做的工作。

以 x_i 对应着的非負实数 $I(P_i)$代替 x_i，可以得到一个随机变数 ξ，

$$\xi:\ \begin{matrix} I(P_1), & \cdots, & I(P_i), & \cdots, & I(P_n) \\ P_1, & \cdots, & P_i, & \cdots, & P_n。 \end{matrix}$$

我們采用 ξ 的数学期望（即概率平均）来反映概率空間 X，$P(\ )$ 的平均不肯定程度，記为 $H(X)$。于是

$$H(X)=\sum_{i=1}^{n} P_i I(P_i)=\sum_{X} P_x I(P_x) \qquad (1.2\text{-}1)$$

应該注意：x 是事件，不是数，因此 $H(X)$并不是指概率空間 X 的数学期望，而是指 ξ

的数学期望。X，$P(\ \ )$不存在数学期望是显而易見的，因为我們不知道一个数(比方說 P_i)乘一个事件（比方說 x_i）是什么。

从（1-2-1）式可以看出我們关心的只是各个事件的概率。和 $I(P_x)$一样，$H(X)$与 X 的事件代表什么无关，所以 $H(X)$也可以写成 $H(P_1, \cdots\cdots, P_n)$。

我們开始討論 $I(P_x)$的时候，沒有包括 $I(P_x=0)$的情况，故目前空間 X，$P(\ \ )$中的元素只应取那些有可能出現的事件，不可能事件应該略去。（以后将看到，如果我們在空間中放进一个不可能事件，不会影响 $H(\ \ )$。不过，目前我們无法說明这一点，只好暫时将之略去）。

$H(X)$的含义已經搞清楚了，但由于它在信息論中的重要性，我們再着重地把它写一次。

H(X)是用以解除我們对X中那一个事件出現这个不肯定性所应收到的平均信息量

不过，这仍不足以决定 $H(\ \ )$。換句話說：即使給出 $P_1, \cdots\cdots, P_n$，仍无法得到 $H(\ \ )$。我們必須再找到一些 $H(P_1, \cdots\cdots, P_n)$与 $P_1, \cdots\cdots, P_n$ 之間的强有力的关系。下面便是其中比較重要的一个。

設有一概率空間

$$X, P(\ \ ): \begin{matrix} x_1, & x_2, & x_3 \\ P_1, & P_2, & P_3 \end{matrix}。$$

我們有两种方法考虑 $H(P_1, P_2, P_3)$。第一种方法是直接考虑 x_1, x_2, x_3 那一出現，这时的 $H(\ \ )$根据（1-2-1）式为

$$H(P_1, P_2, P_3)=P_1I(P_1)+P_2I(P_2)+P_3I(P_3)。$$

第二种方法，把 x_1 不出現記为，$\bar{x}_1$，显然 x_2 和 x_3 的出現都必須在 x_1 不出現（即 $\bar{x}_1$ 出現）的情况下才能发生。反过来說，如果 $\bar{x}_1$ 已經发生，則 x_2 或者 x_3 必然出現其中的一件。x_2 和 x_3 的条件概率（条件为 $\bar{x}_1$ 已經发生）分别記为 $P(x_2/\bar{x}_1)=\frac{P_2}{P(\bar{x}_1)}$，$P(x_3/\bar{x}_1)=\frac{P_3}{P(\bar{x}_1)}$。現在我們把 x_1, x_2, x_3 那一个出現的考虑分成两步进行。首先考虑 $x_1, \bar{x}_1$ 那一个出現，解决这个問題需要的平均信息量为 $H[P_1, P(\bar{x}_1)]$。如果 x_1 出現，我們就肯定了；如果 $\bar{x}_1$ 出現，我們仍然存在是 x_2 还是 x_3 出現的不肯定性，因此应該再收到一些信息量，$H\left(\frac{P_2}{P(\bar{x}_1)}, \frac{P_3}{P(\bar{x}_1)}\right)$，才能肯定。由于 $\bar{x}_1$ 出現的概率为 $P(\bar{x}_1)$，那就是說只有 $P(\bar{x}_1)\times 100\%$ 我們須要收到 $H\left(\frac{P_2}{P(\bar{x}_1)}, \frac{P_3}{P(\bar{x}_1)}\right)$，其他的部分則不必再收到任何信息量已經能够肯定 x_1 出現。因此，为了解决 x_1, x_2, x_3 那一个出現的不肯定性所需的信息量，平均說来应該是

$$H[P_1, P(\bar{x}_1)]+P(\bar{x}_1)H\left(\frac{P_2}{P(\bar{x}_1)}, \frac{P_3}{P(\bar{x}_1)}\right)。$$

上述两种方法所需要的平均信息量应該相等，因为不管我們是直接考虑 x_1, x_2, x_3 那一个出現；还是先考虑 x，$\bar{x}_1$ 的出現，如果 $\bar{x}_1$ 出現，再考虑 x_2, x_3 的出現。要解决的都是同一个概率空間的不肯定性。所以我們有

$$H(P_1,\ P_2,\ P_3)=H[P_1,\ P(\bar{x})]+P(\bar{x})H\left(\frac{P_2}{P(\bar{x}_1)},\ \frac{P_3}{P(\bar{x}_1)}\right)$$

把这个想法推广一步，不难得到

$$H(P_1,\cdots\cdots,P_{n-1},q_1,q_2)=H(P_1,\cdots\cdots,P_n)+P_nH\left(\frac{q_1}{P_n},\ \frac{q_2}{P_n}\right)$$

我們已經讲过 $P_n\neq 0$，故式中应有 $P_n=q_1+q_2>0$。

除此之外，当然我們还能够找出一些 P_1，……，P_n 与$H(P_1,\ \cdots\cdots,\ P_n)$的关系。例如申农（*C.E.Shannon*），他还想到当 n 增加时，$H\left(\frac{1}{n},\ \cdots\cdots,\ \frac{1}{n}\right)$是个不减函数。即：一个具有 n 个基本事件的空間，如果这些事件的概率都相等，那么空間包含的事件愈多，則它的不肯定程度愈大。这个想法也是很自然的。又例如苏联数学家辛欽（*А.Я.Хинчин*）曾經想到各 P 相等时 $H(\ \ )$最大。他們能根据这些直观的想法推导出 $H(P_1,\ \cdots\cdots P_n)$是 P_1，……，P_n 的什么函数。但是根据苏联另外一位数学家法捷耶夫（*Д.К.Фадеев*）的卓越工作所指出，只要 $H(\ \ )$滿足下列三个条件，它便能唯一地被决定。（除了一个常数外）我們这一节主要是遵循法捷耶夫給出的严格数学方法来推导 $H(\ \ )$。

这三个条件是：

（1）$H(P,1-P)$是 P 的連續函数，$0\leqslant P\leqslant 1$；

（2）$H(P_1,\cdots\cdots,P_n)$对所有 P 来說都是对称的；

（3）如果 $P_n=q_1+q_2>0$，則

$$H(P_1,\cdots\cdots,P_n-1,q_1,q_2)=H(P_1,\cdots\cdots,\ P_n)+P_nH\left(\frac{q_1}{P_n},\ \frac{q_2}{P_n}\right)$$

自然，其中 $P_1+\cdots\cdots+P_n=1$。其实这三个条件总結，健全了我們上面对 $H(\ \ \ \ )$的一些直观想法。条件（2）的意思是$H(\ \ \ \ )$与$(\ \ \ \ )$中各 P 的位置无关。例如$H(P_1,P_2,\cdots\cdots,P_n)$可以写成$H(P_2,P_1,\cdots\cdots,P_n)$，也可以写成$H(P_n,P_1,\cdots\cdots P_{n-1})$等等。

有了这三个条件，我們就可以根据下面定理找到 $H(P_1,\cdots\cdots,P_n)$。

定理：符合条件（1）—（3）的函数$H(P_1,\cdots\cdots,P_n)$一定具有

$$H(P_1,\cdots\cdots,P_n)=-C\sum_{i=2}^{n}P_i\log P_i$$

的形式，其中 C 为正常数。

这个定理的証明将通过一些引理而被得到。

引理（1）：

$$H(1,0)=0$$

証：应用条件（3），

$$H\left(\frac{1}{2},\ \frac{1}{2},\ 0\right)=H\left(\frac{1}{2},\ \frac{1}{2}\right)+\frac{1}{2}H(1,0)$$

$$H\left(0,\ \frac{1}{2},\ \frac{1}{2}\right)=H(0,1)+H\left(\frac{1}{2},\ \frac{1}{2}\right)$$

根据条件（2），上面两式左手边相等，因此右手边也应該相等。我們有

$$\frac{1}{2}H(1,0)=H(1,0)$$

于是引理得到了証明。否則我們得到 $1=2$。

这引理虽然仍旧未解决 $I(0)$ 等于什么，但它已指出概率空間

$$X,\ \bar{x}$$
$$P,\ 1-P$$

的 $H(P,1-P)$，不論 $P=1$ 或者 $P=0$，都应該等于零。

引理（2）：如果 $P_n>0$，則

$$H(P_1,\cdots\cdots,P_n,0)=H(P_1,\cdots\cdots,P_n)$$

証：这个引理的証明可以直接从条件（3）及引理（1）得到。

以前我們暫时規定概率空間不包含不可能事件；現在我們可以說，即使它包含不可能事件；它的不肯定程度并未改变。

引理（3）：

$$H\Big(\overbrace{\frac{1}{mn},\cdots\cdots,\frac{1}{mn}}^{mn个}\Big)=H\Big(\overbrace{\frac{1}{m},\cdots\cdots,\frac{1}{m}}^{m个}\Big)+H\Big(\overbrace{\frac{1}{n},\cdots\cdots,\frac{1}{n}}^{n个}\Big)\quad(1.2\text{-}2)$$

証：不难証明条件（3）可以推广至 m 个 q 的情况。即

$H(P_1,\cdots\cdots,P_{n-1},q_1,\cdots\cdots,q_m)=H(P_1,\cdots\cdots,P_n)+P_nH\Big(\frac{q_1}{P_n},\cdots\cdots,\frac{q_m}{P_n}\Big)$，如果其中 $P_n=q_1+\cdots\cdots+q_m>0$。根据这个推广了的条件（3），直接可以得出下式。

$$H(\overbrace{q_{11},\cdots\cdots,q_{1m_1}}^{m_1个};\cdots\cdots;\overbrace{q_{n1},\cdots\cdots,q_{nm_n}}^{m_n个})=H(q_{11},\cdots\cdots,q_{1m_1};\cdots\cdots;P_n)$$
$$+P_nH\Big(\frac{q_{n1}}{P_n},\cdots\cdots,\frac{q_{nm_n}}{P_n}\Big)\Big(\overbrace{\quad}^{n组}\Big).$$

应用条件（2），将 P_n 移至括号的左边，并将这步驟继續下去，我們有

$$H(q_{11},\cdots\cdots,q_{1m_1};\cdots\cdots;q_{n1},\cdots\cdots,q_{nm_n})$$
$$=H(P_1,\cdots\cdots,P_n)+\sum_{i=1}^{n}P_iH\Big(\frac{q_{i1}}{P_i},\cdots\cdots,\frac{q_{im_i}}{P_i}\Big),$$

如果其中 $P_i=q_{i1}+\cdots\cdots+q_{im_i}>0$。$(i=1,2,\cdots\cdots,n)$

現在假設 $m_1=\cdots\cdots=m_n=m$，$q_{11}=\cdots\cdots=q_{1m_1}=\cdots\cdots=q_{n1}=\cdots\cdots=q_{nm_n}=q$；那么

$q=\frac{1}{mn}$，而且 $P_1=\cdots\cdots=P_n=\frac{1}{n}$。最后

$$H\Big(\frac{1}{mn},\cdots\cdots,\frac{1}{mn}\Big)=H\Big(\frac{1}{n},\cdots\cdots,\frac{1}{n}\Big)+\sum_{i=1}^{n}P_iH\Big(\frac{1}{m},\cdots\cdots,\frac{1}{m}\Big).$$

因为

$$\sum_{i=1}^{n}P_i=1$$

于是証明了引理。

这个引理使我們能够得到一个很重要的結果。現在我們来談談这个結果如下：n 是概率

空間的元素数目，所以它一定是一个正整数$\geqslant 1$。当它等于1（只有一个事件），就是說这空間只有一个必然事件，我們已經知道$H(1,0)=0$；重要的是$n\geqslant 2$的情况。

根据数論知識，我們知道：一个数，如果它只能被两个整数（一个是1，另一个是它自己）除尽，它就叫素数，例如2，3，5，7，……就是素数；最小的素数是2；除2外，所有的素数都是奇数；一个大于1的整数必定能够分解成为素数因子的乘积，即

$$2\leqslant a=p_1^{\alpha_1}\cdots\cdots p_s^{\alpha_s} \qquad (1.2\text{-}3)$$

其中$p_1,\cdots\cdots,p_s$均为素数，$\alpha_1,\cdots\cdots,\alpha_s$为自然数。例如：

$$616=2\times 2\times 2\times 7\times 11=2^3\times 7^1\times 11^1。$$

（1-2-3）式称为a的素数因子分解式；如果不考虑素数因子的次序，分解方法是唯一的。就是說；給定一个a，只有一个素数因子分解式，如果不考虑各p的次序的話。

以上关素数的叙述，今后我們都将用到。現在我們把n的素数因子分解式写出来。

$$2\leqslant n=p_1^{\alpha_1}\cdots\cdots p_s^{\alpha_s}$$

把它代入定理的公式中，

$$H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)=H\left(\frac{1}{p_1^{\alpha_1}\cdots p_s^{\alpha_s}},\cdots\cdots,\frac{1}{p_1^{\alpha_1}\cdots p_s^{\alpha_s}}\right)$$

$$=H\left(\frac{1}{p_1^{\alpha_1}},\cdots\cdots,\frac{1}{p_1^{\alpha_1}}\right)+\cdots\cdots+H\left(\frac{1}{p_s^{\alpha_s}},\cdots\cdots,\frac{1}{p_s^{\alpha_s}}\right)$$

$$=\alpha_1 H\left(\frac{1}{p_1},\cdots\cdots,\frac{1}{p_1}\right)+\cdots\cdots+\alpha_s H\left(\frac{1}{p_s},\cdots\cdots,\frac{1}{p_s}\right)。$$

再令

$$\frac{H\left(\frac{1}{p},\cdots\cdots,\frac{1}{p}\right)}{\log p}=Cp \qquad (1.2\text{-}4)$$

則

$$H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)=\alpha_1 Cp_1\log p_1+\cdots\cdots+\alpha_s Cp_s\log p_s。 \qquad (1.2\text{-}5)$$

由于素数因子分解式的唯一性，（1-2-5）式也是唯一的，除了各項的先后次序外。显而易見，如果我們能够証明（1-2-5）式中的$Cp_1=\cdots\cdots=Cp_s=$常数，那么$H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)$就被規定了。这是我們接着要作的工作，将在引理（5）中加以証明。有一点应該特别注意的是我們用小写的p代表n的素数因子（它不是概率）。因此（1-2-4）式的p无須考虑它是素数以外的数。不过，当$n\to\infty$时，也有可能使得$p_s\to\infty$，（当然不一定）但它仍为素数这一点我們应該肯定的。

引理（4）：当$n\to\infty$时，

$$\frac{1}{n}H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)\to 0;$$

同时

$$H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)-H\left(\frac{1}{n-1},\cdots\cdots,\frac{1}{n-1}\right)\to 0。$$

証：根据推广了的条件（3），不难写出

$$H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)=H\left(\frac{1}{n},\ \frac{n-1}{n}\right)+\frac{n-1}{n}H\left(\frac{1}{n-1},\cdots\cdots,\frac{1}{n-1}\right)\quad(1.2\text{-}6)$$

这里把 p_n 当作 $n-1$ 个 $\frac{1}{n}$ 之和。令

$$\eta_n=H\left(\frac{1}{n},\ \frac{n-1}{n}\right),$$

那么，

$$\eta_n=H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)-\frac{n-1}{n}H\left(\frac{1}{n-1},\cdots\cdots,\frac{1}{n-1}\right)\qquad(1.2\text{-}7)$$

由于条件（1）对 $H(P,1-P)$ 在閉区間〔0,1〕上連續的規定，当 $n\to\infty$ 时

$$H\left(\frac{1}{n},\ \frac{n-1}{n}\right)\to H(1,0)\ \text{也就是}\eta_n\to 0\ \text{故}$$

$$\sum_{k=2}^{n}K\,\eta_k=nH\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)。$$

或者可以写成

$$\frac{1}{n}H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)=\left(\frac{1}{n^2}\sum_{k=2}^{n}K\,\eta_k=\frac{n(n+1)-2}{2n^2}\ \ \frac{2}{n(n+1)-2}\sum_{k=2}^{n}K\eta_k。\right.$$

因为 $\sum_{k=2}^{n}K\eta_k=\overbrace{\eta_2+\eta_2}^{2\text{个}}+\overbrace{\eta_3+\eta_3+\eta_3}^{3\text{个}}+\cdots\cdots+\overbrace{\eta_n+\eta_n+\cdots\cdots+\eta_n}^{n\text{个}}。$

上式右手边一共有

$$\frac{n(n+1)}{2}-1=\frac{n(n+1)-2}{2}$$

项，所以 $\frac{2}{n(n+1)-2}\sum_{k=2}^{n}K\eta_k$ 就是 $\sum_{k=2}^{n}K\eta_k$ 的头 $\frac{n(n+1)-2}{2}$ 项的算术平均。同时因为 $\eta_k\to 0$，故算术平均也趋于 0。这就是引理中第（1）式所需要的証明。

引理中第（2）式的証明可以根据（1.2-6）式及引理中第（1）式直接得到

$$H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)-H\left(\frac{1}{n-1},\cdots\cdots,\frac{1}{n-1}\right)$$

$$=H\left(\frac{1}{n},\ \frac{n-1}{n}\right)-\frac{1}{n}H\left(\frac{1}{n-1},\cdots\cdots,\frac{1}{n-1}\right)\to 0$$

引理的第（1）个公式說明一个概率空間（各基本事件的出現概率相等），它的不肯定性并不与它包含的事件数目成正比例关系。例如一个包含 8 个等概率事件的空間，它的 H（ ）不是一个包含 4 个等概率事件的空間的 H（ ）的 2 倍，而是小于两倍，至少在事件数目足够大的时候是如此，否則当 $n\to\infty$ 时 $\frac{1}{n}H\left(\frac{1}{n},\cdots\cdots,\frac{1}{n}\right)$ 便无法趋于 0。下面一个例子可以更清楚地看出这一点。設有一个基本概率空間

$$X, x, P(x): \quad \overset{x_1,\ x_2,\ x_3,\ x_4}{P_1=P_2=P_3=P_4=\frac{1}{4}}。$$

要决定它那一个元素出現，我們可以分两步进行：第一步先决定是 $x_1 U x_2$ 出現呢？还是 $x_3 U x_4$ 出現。因为 $P(x_1 U x_2)=P(x_3 U x_4)=\frac{1}{2}$，所以我們須要收到 $H\left(\frac{1}{2}, \frac{1}{2}\right)$ 的信息量。第一步已經决定后，比方說 $x_3 U x_4$ 出現，再决定是 x_3 还是 x_4 出現。这个决定所需要得到的信息量也是 $H\left(\frac{1}{2}, \frac{1}{2}\right)$，因此总共需要的信息量为 $2H\left(\frac{1}{2}, \frac{1}{2}\right)$。现在把事件增加为 8 个。

$$X, x, P(x): \quad \overset{x_1,\ x_2,\ x_3,\ x_4,\ x_5,\ x_6,\ x_7,\ x_8}{P_1=P_2=P_3=P_4=P_5=P_6=P_7=P_8=\frac{1}{8}}。$$

我們首先决定出現的是 $x_1 U x_2 U x_3 U x_4$ 和 $x_5 U x_6 U x_7 U x_8$ 这两个事件中的那一个，然后按照决定 4 个元素时的两个步驟进行。其实我們一共分成三步，每决定一步应該收到 $H\left(\frac{1}{2}, \frac{1}{2}\right)$ 的信息量，总共要收到 $3H\left(\frac{1}{2}, \frac{1}{2}\right)$，而不是 $4H\left(\frac{1}{2}, \frac{1}{2}\right)$。

引理中第（2）个公式說明一个包含 n 个事件的概率空間，当 n 很大很大时，多一个事件或者少一个事件，其不肯定性几乎不变。这个概念看起来很簡单，可是它将成为引理（5）的重要依据。

引理（5）：

$$\frac{H\left(\frac{1}{p}, \cdots\cdots, \frac{1}{p}\right)}{\log p}=Cp=常数。$$

証：由（1.2-4）式，我們已經知道 p 是素数，它可以是 2 或者 2 以上的任何一个素数。今用 P 代表以 p 为其元素的集，它的左端点为 2，沒有右端点。如果有一个 p，比方說是 p_k，能使 Cp_k 大于或者等于所有其他的 Cp，（但至少大于其他的一个，即不是全部相等）我們称 Cp_k 为 Cp 在集 P 上的最大值；反之，如果有一个 p，比方說 p_n，能使 Cp_n 小于或者等于所有其他的 Cp，（但至少小于其他的一个）則 Cp_n 被称为 Cp 在集 P 上的最小值。

我們准备用以証明引理（5）的方法是首先假設 $Cp \neq$ 常数，即数列

$$Cp_1,\ Cp_2, \cdots\cdots, Cp_2, \cdots\cdots$$

中的各 p 不是全部相等，至少有一个不等。那么 Cp：要末，沒有最大（最小）值；要末，有最大（最小）值。然后証明：i，Cp 不能沒有最大（最小）值；ii，Cp 不能有最大（最小）值。从而否定了我們 $Cp \neq$ 常数的假設，反面地証明了 $Cp=$ 常数。我們用下图 2 来补充說明的不足。

这里簡单地插几句話。初看起来，好像我們已經假設至少有一个 Cp 不等在先，怎么又能假設沒有最大（最小）值于后？同时，既然能証明沒有最大（最小）值，不就証明 $Cp=$ 常数了嗎？其实，并不如此。一个无限数列可以一个小于一个，但并无最小值。例如