信息论基础

A.范恩斯坦

科学出版社

信 息 论 基 础

A. 范恩斯坦 著

江 泽 培 譯

科 学 出 版 社

1964

AMIEL FEINSTEIN
FOUNDATIONS OF INFORMATION THEORY
McGraw-Hill 1958

內容簡介

本书包括了从1948年C. E. Shannon的奠基性工作一直到1957年这一段时期內在信息論方面的重要的基本理論.

全书共分七章,前四章对于信息論的基本概念作了严格的叙述,对于一类重要的特殊通路(受干扰的无記忆的离散通路)圓滿地論証了信息传輸的基本定理. 第五章与第六章的內容是把前四章的結果相应地推广到一些通路(半連續离散通路,受干扰的有限記忆通路). 最末一章对于二元对称通路,詳細地討論了传輸誤差与传輸时間的关系,又討論了关于寻找最优編碼的問題.

此书可供数学工作者(特別是概率論方面的科学工作者),高等院校师生閱讀参考之用.

信息论基础

A. 范恩斯坦 著
江泽培 譯

*

科学出版社 出版 (北京朝阳門大街117号)
北京市书刊出版业营业許可証出字第061号
中国科学院印刷厂印刷 新华书店总經售

*

1964年2月第一版 书号:2806 字数:104,000
1964年2月第一次印刷 开本:850×1168 1/32
(京) 0001—6,000 印张:4 1/16

定价:(7) 0.70 元

序

作者写这本书的目的在于扼要而严格地叙述数学信息論的基本知識，自从 1948 年 C. E. Shannon 发表他的重要的工作之后，在这領域中已經作了許多工作，但是对有志于这方面的讀者来說，目前还沒有一本专門的书能够比較完备地介紹这些新結果. 希望这本书可以弥补这种缺陷.

信息論的发展，在目前阶段可以区分为两个不同的分支. 第一个是研究各种通路的一般性质，証明有关这些通路适合于传輸信息的各种要求的定理，第二个則是为了如何实际地运用这些一般性的定理. 現在看来，第一分支的工作比第二分支的工作要进展得快一些. 例如，虽然已經知道，利用容量不为零的受干扰的通路，可以以任意地接近于通路容量那么大小的速率来传輸信息，并且，还可以要求接收誤差的概率任意地小，但是，即使对于最簡单的不平凡通路，也还沒有什么具体的办法来实現这个理論上的結果，虽然如此，还是有不少有意义的、相当巧妙的文章，对于各种通路构造了一些特定的碼子. 如果要詳細一点介紹这方面的結果，就会大大增加这本书的原先拟定的篇幅. 可以在 7.3 节見到一些与它們有关的內容.

在第一章至第四章以及第七章里，主要的工具就是有限概率場的理論，例如，象 Feller [1] 书中所述的那些. 第五章与第六章所要求的数学上的預备知識比較深一些，所需要的是一般概率空間的概念，因此所要求的数学工具就是測度論与一般的 Lebesque 积分. 虽然用了不长的一节簡要地叙述了所需要的測度論的知識，但要想了解一切細节，讀者仍然需要参考一本这方面的专著. 第七章的內容与第五章及第六章完全独立，如果愿意的話，可以在 4.1 节之后，接着就閱讀第七章.

各章的注記包含了各种补充正文的材料，絕大部分参考文献是在注記中指出的，一些在正文之中被認为是"容易証明"的結果，其細节也是在注記中給出的。因此，在閱讀正文的时候，应当随时讀一讀这些注記。

A. 范恩斯坦

哥本哈根，1957 年 1 月

目　录

第一章　初等概念

在这本书里，我們将要介紹的信息論是从 1948 年 C. E. Shannon 的工作出发. 在他的奠基性的論文中，Shannon 提出了一个数学概型，对于信息的产生与传輸这些概念从量的方面給以定义. 他并且叙述和証明了一些非常一般的結果，表明了这些定义的重要与有效. 自从 1948 年以来，已有許多文章对于 Shannon 的原始工作加以簡化和推广. 同时，又有許多各种各样的工作試图把信息論应用到物理、化学以及生物学与心理学的各种不同的分支中去. 这里，我們不打算接触到这些应用；无論如何，已經有了一些专門論述这些应用的书，有兴趣的讀者可以去参考.

本书的主要內容多是演譯性的，也就是說，从少量的定义出发来推論其他的一切东西，这种途径正是我們将要采用的. 另一方面，信息論又引用了象信息容量、信息源、信息传輸率这样一些带有某种直观意义的名詞. 因此，在数学上每进展一步的时候，把我們的結果与直观所提示的东西加以比較，是很有意义的. 我們将会看到，信息論的基本概念很容易借助于直观来加以解释. 当然，我們不能够对于信息論的一些更高深的結果也做同样的要求，这里指的是那些复杂的根本性的极限定理，它們对于愿意进行深入鉆研的人們，能够起引导作用.

当我們試图建立一个定量的理論(对于象"信息的产生"与"信息的传輸"这些概念要給以明确的含义)时，卽刻就碰到两个困难. 首先，我們必須建立一个数学模型，从而可以談論信息的产生与传輸. 其次，我們必須对于所包含的信息給出量的刻划. 初看起来，好象第二个問題可以直接随着第一个問題的解决而得到解决. 当进一步考虑問題时，就会发現事情并不是这样.

直观地講，当我們知道了一个預先不能肯定其是否发生的事

件发生时，我們認为是获得了信息．而且，至少在一定的限度內可以合理地認为：事件愈是有可能发生，当我們知道它确实发生时，我們所得到的信息就愈少．暂且忽略这点，我們已能进行一些形式的討論．設 x 代表一个事件（就是这个事件的发生）而 x' 是它的对立事件（也就是这个事件不发生），又令 p_x 与 $p_{x'}$ 分别代表这两个事件的概率，于是 $p_x + p_{x'} = 1$．令 I_x 表示我們知道了事件 x 发生而获得的信息量．既然 x 仅由它的概率所决定，我們可以假定 I_x 是 p_x 的一个函数，就是說：$I_x = I(p_x)$，此处 $I(\ )$ 是在 p_x 的值域（即 $0 < p_x \leqslant 1$）上定义的非負函数，在目前的討論中 $p_x = 0$ 和 $p_x = 1$ 的場合是沒有什么意义的．类似地，我們令 $I_{x'} = I(p_{x'})$．因为获得信息量 I_x 的概率为 p_x，获得信息量 $I_{x'}$ 的概率为 $p_{x'}$，所以获得信息量的平均值（或期望值）就是 $p_xI(p_x) + p_{x'}I(p_{x'})$．类似地，如果我們有一組互不相容的事件 $x_1, \cdots, x_n$，而且 $p_{x_1} + \cdots + p_{x_n} = 1$，那就可以把 $p_{x_1}I(p_{x_1}) + \cdots + p_{x_n}I(p_{x_n})$ 作为我們知道了究竟是那一个 x_i 发生而获得的信息量的平均值．在計算这个信息量的平均值时，如果 $p_{x_i} = 0$，則与之相应的一項显然应当略而不計．現在还一点也看不出应当如何来选定函数 $I(\ )$．

但是，可以継續这种討論来得到对于 $I(\ )$ 的形状的較強的約制．考虑三个互不相容的事件 x, y, z，而 $p_x + p_y + p_z = 1$．于是 $H(x, y, z) = p_xI(p_x) + p_yI(p_y) + p_zI(p_z)$ 就表示我們知道了在 x, y, z 之中哪一个发生而获得的信息量的平均值．为了明确在 x, y, z 之中哪一个眞正发生，可以先明确 x 究竟发生或是不发生，而当 x 不发生时，再进一步明确 y, z 之中哪一个发生．明确第一件事所获得的信息量显然是 $p_xI(p_x) + (1 - p_x)I(1 - p_x)$，我們用 $H(x, x')$ 来代表它．如果 x 不发生，則 y 与 z 的（条件）概率就分别是 $\frac{p_y}{p_{x'}}$ 与 $\frac{p_z}{p_{x'}}$．因而明确第二件事所获得的信息量就是 $\frac{p_y}{p_{x'}} I\left(\frac{p_y}{p_{x'}}\right) + \frac{p_z}{p_{x'}} I\left(\frac{p_z}{p_{x'}}\right)$，我們用 $H(y, z|x')$ 来代表它．但是，只当事件 x' 发生时，我們才获得后一个信息量，所以我們可以取 $H(x, x') + p_{x'}H(y, z|x')$ 作为明确 x, y, z 之中哪一个眞正发生

所获得的总的平均信息量．于是，对于一切可能的正数 p_x, p_y, p_z, 要求函数 $I(\ )$ 满足关系式 $H(x, x') + p_{x'}H(y, z|x') = H(x, y, z)$, 就是很合理的了．

让我們用 p_x, p_y, p_z 表出函数 H; 这就是說，我們令 $H(x, x') = H(p_x, 1-p_x)$, $H(x, y, z) = H(p_x, p_y, p_z)$ 和 $H(y, z|x') = H\left(\frac{p_y}{1-p_x}, \frac{p_z}{1-p_x}\right)$. 于是，我們要求下面的关系成立：

$$H(p_x, p_y, p_z) = H(p_x, 1-p_x) + (1-p_x)H\left(\frac{p_y}{1-p_x}, \frac{p_z}{1-p_x}\right).$$

如果 x 是一个复合事件，可以运用同样的推理；这就是說，我們假設 x 由互不相容的事件 $u_1, \cdots, u_{n-1}$ 組成，它們的概率分別为 $p_1, \cdots, p_{n-1}$. 于是我們得到 $H(u_1, \cdots, u_{n-1}, y, z) = H(u_1, \cdots, u_{n-1}, x') + p_{x'}H(y, z|x')$, 或者，令 $q_1 = p_y$, $q_2 = p_z$, $p_n = p_{x'}$, 我們得到

$$H(p_1, \cdots, p_{n-1}, q_1, q_2) = H(p_1, \cdots, p_n) + p_nH\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

这是一个很強的要求；事实上，我們很快就会看到，它已經足够决定 $H(p_1, \cdots, p_n)$ 的形状，而不必考慮 H 是如何通过 $I(\ )$ 而定义的．由于在前面提过的一件事：当 $p_i = 0$ 时，$p_iI(p_i)$ 这一項应当略而不計，因此，卽使某些 p_i 是零的时候，仍然要求 $H(p_1, \cdots, p_n)$ 有定义，而且它在区域 $p_i \geqslant 0$, $i = 1, \cdots, n$, $p_1 + \cdots + p_n = 1$ 內是連續的．我們現在可以敍述如下的定理：

定理． 下面三个条件确定了函数 $H(p_1, \cdots, p_n)$（只剩一个常数因子沒有确定，这个因子刚好用来确定信息量单位的大小）．

1. $H(p, 1-p)$ 是綫段 $0 \leqslant p \leqslant 1$ 上 p 的連續函数．

2. $H(p_1, \cdots, p_n)$ 对它的自变量 $p_1, \cdots, p_n$ 来講，是对称的函数．

3. 如果 $p_n = q_1 + q_2 > 0$, 則

$$H(p_1, \cdots, p_{n-1}, q_1, q_2) = H(p_1, \cdots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{q_2}{p_n}\right).$$

以后，我們总是做这样的了解：函数 $H(p_1, \cdots, p_n)$ 仅仅对于整个一組概率有定义，也就是說，对于一組非負的数（其和数为 1）有定义.

把定理的証明分成下面一串引理来进行.

引理 1. 等式 $H(1, 0) = 0$ 成立.

証. 由条件 3，$H\left(\frac{1}{2}, \frac{1}{2}, 0\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2}H(1, 0)$. 但是，由条件 2 与 3，我們有 $H\left(\frac{1}{2}, \frac{1}{2}, 0\right) = H\left(0, \frac{1}{2}, \frac{1}{2}\right) = H(0, 1) + H\left(\frac{1}{2}, \frac{1}{2}\right)$，由此推得 $H(1, 0) = 0$，这是因为 $H(1, 0) = H(0, 1)$.

引理 2. 等式 $H(p_1, \cdots, p_n, 0) = H(p_1, \cdots, p_n)$ 成立.

証. 由于条件 2 我們可以假定 $p_n > 0$，于是，利用条件 3，再应用引理 1，就得到我們的結果.

引理 3. 我們有

$$H(p_1, \cdots, p_{n-1}, q_1, \cdots, q_m) = H(p_1, \cdots, p_n) + p_n H\left(\frac{q_1}{p_n}, \cdots, \frac{q_m}{p_n}\right),$$

其中 $p_n = q_1 + \cdots + q_m > 0$.

証. 当 $m = 2$ 时，这个等式恰好就是条件 3. 我們按 m 来进行数学归納法；由引理 2，显然只需考虑 $q_i > 0 (i = 1, 2, \cdots, m)$ 的場合. 假定对于某个 m 我們所要推証的論断对于一切 n 皆成立. 于是，利用条件 3，我們就有

$$\begin{aligned} H(p_1, \cdots, p_{n-1}, q_1, \cdots, q_{m+1}) &= H(p_1, \cdots, p_{n-1}, q_1, p') + p' H\left(\frac{q_2}{p'}, \cdots, \frac{q_{m+1}}{p'}\right) \\ &= H(p_1, \cdots, p_n) + p_n H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) \\ &\quad + p' H\left(\frac{q_2}{p'}, \cdots, \frac{q_{m+1}}{p'}\right), \end{aligned}$$

其中 $p' = q_2 + \cdots + q_{m+1}$.

但是,又有

$$H\left(\frac{q_1}{p_n}, \cdots, \frac{q_{m+1}}{p_n}\right) = H\left(\frac{q_1}{p_n}, \frac{p'}{p_n}\right) + \frac{p'}{p_n} H\left(\frac{q_2}{p'}, \cdots, \frac{q_{m+1}}{p'}\right).$$

把这个关系代入上面的等式,我們就得到引理对于 $m+1$ 的論断.

引理 4. 我們有

$$H(q_{11}, \cdots, q_{1m_1}; \cdots; q_{n1}, \cdots, q_{nm_n})$$
$$= H(p_1, \cdots, p_n) + \sum_{i=1}^{n} p_i H\left(\frac{q_{i1}}{p_i}, \cdots, \frac{q_{im_i}}{p_i}\right),$$

其中 $p_i = q_{i1} + \cdots + q_{im_i} > 0$.

証. 利用引理 3, 我們有

$$H(q_{11}, \cdots, q_{1m_1}; \cdots; q_{n1}, \cdots, q_{nm_n})$$
$$= p_n H\left(\frac{q_{n1}}{p_n}, \cdots, \frac{q_{nm_n}}{p_n}\right)$$
$$+ H(q_{11}, \cdots, q_{1m_1}; \cdots; q_{n-1,1}, \cdots, q_{n-1,m_{n-1}}; p_n).$$

把 p_n 移到最左边, 継續簡化下去,經过 n 个步驟我們最后就得到所期望的結果.

現在, 对于 $n \geqslant 2$, 設 $F(n) = H\left(\frac{1}{n}, \cdots, \frac{1}{n}\right)$, 而且 $F(1) = 0$. 对于 $m_1 = \cdots = m_n = m$, $q_{ij} = \frac{1}{mn}$, 的場合, 利用前面一条引理,我們得到 $F(mn) = F(m) + F(n)$. 当 $m = 1$ 或者 $n=1$ 时, 这个关系式显然成立. 又对于 $H\left(\frac{1}{n}, \cdots, \frac{1}{n}\right)$ 运用引理 3, 我們得到

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

从而有

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

我們現在証明下面的

引理 5. 当 $n \to \infty$ 时, $\mu_n = \dfrac{F(n)}{n} \to 0$,

并且 $\lambda_n \equiv F(n) - F(n-1) \to 0$.

証. 由 $H(p, 1-p)$ 的連續性可以推出 $\eta_n \to H(0, 1) = 0$ (当 $n \to \infty$). 又有

$$n\eta_n = nF(n) - (n-1)F(n-1),$$

于是 $nF(n) = \sum_{k=1}^{n} k\eta_k$, 或者

$$\frac{F(n)}{n} = \frac{1}{n^2}\sum_{k=1}^{n} k\eta_k = \frac{n+1}{2n}\,\frac{2}{n(n+1)}\sum_{k=1}^{n} k\eta_k.$$

$\dfrac{2}{n(n+1)}\sum_{k=1}^{n} k\eta_k$ 就是序列 $\eta_1, \eta_2, \eta_2, \eta_3, \eta_3, \eta_3, \cdots$ 的前 $\dfrac{n(n+1)}{2}$ 項的算术平均数, 而这个序列的极限是零. 于是, 当 $n \to \infty$ 时, $\dfrac{2}{n(n+1)}\sum_{k=1}^{n} k\eta_k \to 0$, 从而有 $\lim\limits_{n\to\infty} \dfrac{F(n)}{n} = 0$. 最后, 当 $n \to \infty$ 时我們有 $\lambda_n \equiv F(n) - F(n-1) = \eta_n - \dfrac{1}{n}F(n-1) \to 0$.

我們現在来确定 $F(n)$的形狀. 由于 $F(nm)=F(m)+F(n)$, 我們只需知道当 n 为質数时 $F(n)$ 的值. 事实上, 对于任何 n, 令 $n = p_1^{\alpha_1}\cdots p_s^{\alpha_s}$ 是 n 的質因数分解式; 我們就有 $F(n) = \alpha_1 F(p_1) + \cdots + \alpha_s F(p_s)$. 对于一切的質数 p, 我們令 $F(p) = c_p \ln p$, 这里 ln 表示通常的自然对数. 于是, $F(n) = \alpha_1 c_{p_1} \ln p_1 + \cdots + \alpha_s c_{p_s} \ln p_s$.

引理 6. 序列 c_p, $p = 2, 3, 5, 7, 11, \cdots$, 具有最大項.

証. 假定引理不成立; 那就可以找到一串質数 $p_1 < p_2 < p_3 < \cdots$ 使得 $p_1 = 2$ 而 p_{i+1} 是一切大于 p_i 的質数之中使得不等式 $c_{p_{i+1}} > c_{p_i}$ 成立的最小者. 由于这种构造, 如果 q 是一个小于 p_i 的質数, 則 $c_q < c_{p_i}$. 对于 $i > 1$, 令 $p_i - 1 = q_1^{\alpha_1}\cdots q_s^{\alpha_s}$ 是 $p_i - 1$

的質因数分解式. 現在

$$
\begin{aligned}
\lambda_{p_i} &= F(p_i) - F(p_i - 1) \\
&= F(p_i) - \frac{F(p_i)}{\ln p_i} \ln (p_i - 1) \\
&\quad + c_{p_i} \ln (p_i - 1) - F(p_i - 1) \\
&= \frac{F(p_i)}{p_i} \frac{p_i}{\ln p_i} \ln \frac{p_i}{p_i - 1} + \sum_{j=1}^{s} \alpha_j (c_{p_i} - c_{q_j}) \ln q_i.
\end{aligned}
$$

因为 $p_i - 1$ 必須是偶数, q_i 之中一定有一个是 2. 又因为 $c_{p_i} > c_{q_j}$ (对于 $j = 1, 2, \cdots, s$), 我們有

$$\sum_{j=1}^{s} \alpha_j (c_{p_i} - c_{q_j}) \ln q_j \geqslant (c_{p_i} - c_2) \ln 2 \geqslant (c_{p_2} - c_2) \ln 2.$$

但是, 当 $i \to \infty$, $p_i \to \infty$; 根据引理 5, $\lambda_{p_i} \to 0$ 并且 $\frac{F(p_i)}{p_i} \to 0$, 同时又容易証明 $\frac{p_i}{\ln p_i} \ln \frac{p_i}{p_i - 1} \to 0$. 因此,我們必須有 $(c_{p_2} - c_2) \ln 2 \leqslant 0$ 或 $c_{p_2} \leqslant c_2$, 这与 p_2 的定义相矛盾.

用同样方法可以証明序列 $c_p (p = 2, 3, 5, \cdots)$ 具有最小項.

引理 7. $F(n) = c \ln n$, 其中 c 是常数.

証. 只須証明所有的 c_p 相等. 假定有質数 p' 使得 $c_{p'} > c_2$. 設 p 是使得 c_p 成为最大項的那个質数; 于是 $c_p > c_2$. 設 m 是正整数, $q_1^{\alpha_1} \cdots q_s^{\alpha_s}$ 是 $p^m - 1$ 的質因数分解式. 由 $F(mn) = F(m) + F(n)$ 可以得出 $\frac{F(p^m)}{\ln p^m} = c_p$; 于是,正如引理 6 的証明中一样, 我們可得到 $\lambda_{p^m} \geqslant \frac{F(p^m)}{p^m} \frac{p^m}{\ln p^m} \ln \frac{p^m}{p^m - 1} + (c_p - c_2) \ln 2$. 令 $m \to \infty$, 我們有 $(c_p - c_2) \ln 2 \leqslant 0$, 这就与 $c_p > c_2$ 相矛盾. 用完全同样的办法,可以証明, 不可能存在任何質数 q 使得 $c_q < c_2$; 这样一来,一切 c_p 皆相等. 我們現在能够完成定理的証明了.

令 $p = \frac{r}{s}$ (r, s 皆整数). 根据引理 4,我們有

$$
\begin{aligned}
&H\left(\frac{1}{s}, \cdots, \frac{1}{s}\right) \\
&\quad = H\left(\frac{r}{s}, \frac{s-r}{s}\right) + \frac{r}{s} H\left(\frac{1}{r}, \cdots, \frac{1}{r}\right) \\
&\qquad + \frac{s-r}{s} H\left(\frac{1}{s-r}, \cdots, \frac{1}{s-r}\right),
\end{aligned}
$$

从而有

$$
\begin{aligned}
H(p, 1-p) &= F(s) - pF(r) - (1-p)F(s-r) \\
&= c \ln s - pc \ln r - (1-p)c \ln (s-r) \\
&= c\left(p \ln \frac{s}{r} + (1-p) \ln \frac{s}{s-r}\right) \\
&= c\left(p \ln \frac{1}{p} + (1-p) \ln \frac{1}{1-p}\right).
\end{aligned}
$$

由于H的連續性，这个結果即刻推广到一切的无理数 p. 利用条件 3，按 n 进行数学归納法就得到 $H(p_1, \cdots, p_n) = c \sum_{i=1}^{n} p_i \ln \frac{1}{p_i}$.

我們注意到，$H(p_1, \cdots, p_n)$ 具有形状 $\sum_{i=1}^{n} p_i I(p_i)$，并且，如果我們取 $c > 0$，則 $I(p)$ 是 $1-p$ 的一个增函数.

注 記

i. C. E. Shannon 的原始工作（Shannon [1]）曾經轉載于 Shannon 与 Weaver [1]. 在 Brillouin [1] 中詳細地討論了信息論在物理学中的应用；Quastler [1，2] 汇編了两个論文集，这些論文討論了信息論在化学、生物学与心理学中的各种可能的应用.

早在 Shannon 的工作之前，人們已經認識到大多数的通訊系統都具有統計的性质. 在 Wiener [1] 里特别強調了这一点，而且，他是第一个利用通訊系統的統計性质来研究預測与濾过問題的人. 但是，公平地說，象通路和信息量这样一些概念，以及基本的翻碼定理的陈述，却完全是属于 Shannon 的.

ii. 除了我們加于$H(p_1, \cdots, p_n)$的三个条件之外，如果还要求

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

是n的增函数，则对$H(p_1, \cdots, p_n)$的形状的推导就会大大簡化(参考 Shannon 与 Weaver [1] 的附录 2). 在 Khintchine [1] 的工作中，对于$H(p_1, \cdots, p_n)$除了要求条件 1 与 2，又假定引理 2 与 4 成立，并且要求$H(p_1, \cdots, p_n) \leqslant F(n)$，从而推导出$H(p_1, \cdots, p_n)$的形状。 我們在这里所采用的簡洁的証明属于 Fadiev [1]. 至于用到的一个事实：从$\lim\limits_{i \to \infty} a_i = a$可以推出$\lim\limits_{n \to \infty} \frac{1}{n} \sum\limits_{i=1}^{n} a_i = a$，这可以利用通常的 ε-δ 方法很容易地推証出来(参考 Halmos [1] 的书 p. 202，定理 B).

第二章　$H(X)$ 的基本性質

2.1. 引　言

在前一章里，我們証明了对于 $H(p_1, \cdots, p_n)$ 的几条簡单而合理的要求就足以完全地确定它（只剩一个常数因子沒有确定）．这个結果虽然有趣，但是并不能因而推論 $H(p_1, \cdots, p_n)$ 这个量具有什么有益的或是有意义的应用．我們在这里希望強調的是：$H(p_1, \cdots, p_n)$ 的用处不在于它能够满足第一章的条件 1—3，而在于它确实作为一个重要的角色出現在許多翻碼和通訊的基本問題中．此后我們只假定 $H(p_1, \cdots, p_n) = -c\sum_{i=1}^{n} p_i \ln p_i$ 这种形状而不加其它評論，让以后的各种定理来証实它．証明这些定理正是我們的主要目的．

2.2. 基本不等式

設 X 是包含有限个元素 x 的抽象集合．又設 $p()$ 是定义在 X 上的概率分布，这就是說，对于 X 的每一个子集合 Q，我們定义 $p(Q)$ 为一个非負的数，具有性质 $p(X) = 1$ 以及 $p(Q_1 \cup Q_2) = p(Q_1) + p(Q_2)$，只要 Q_1 与 Q_2 不相交．我們把事物 (X, x) 的全体与 $p(\ )$ 称为一个有限的概率空間．

由第一章所做的討論，我們可以把任何一个有限概率空間看成一个信息源．我們把非負的数 $-\sum_{X} p(x) \log p(x)$ 定义为这个信息源的信息量 $H(X)$，今后我們把对数一律了解为以 2 为底．对数的底的选择显然是对单位大小的一种規定；在我們所做的这种規定之下，信息量单位称为二进制的信息量单位．为了不引起