信息論的數學理論

王壽仁 著

科學出版社

信息論的數學理論

王壽仁 著

科學出版社
1957

內 容 提 要

本書首先簡單地敍述怎樣把傳送信息過程表述爲數學問題，然後詳盡地綜述了所謂信息論的數學理論中的現代結果。本書只對離散信息加以討論，而且給了 Shannon 的基本定理以證明。

信息論的數學理論

王壽仁著

*

科學出版社出版（北京朝陽門大街 117 號）

北京市書刊出版業營業許可證出字第 061 號

大众文化印刷廠印刷　新華書店總經售

*

1957 年 8 月第一版	書號：0842　印張：2 1/16
1957 年 8 月第一次印刷	開本：850×1168 1/32
(滬) 0001—2,061	字數：49,000

定價：(11)0.46 元

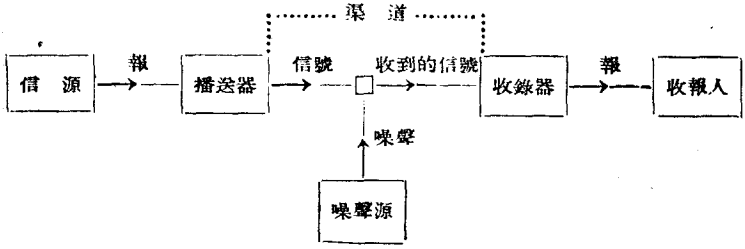
目　錄

引言……(1)
§ 1. 數學定義……(6)
§ 2. 熵的概念……(11)
§ 3. 熵的唯一性……(19)
§ 4. 信源的熵率及渠道的傳信率……(22)
§ 5. 遍歷性……(26)
§ 6. 漸近等同分割性，McMillan 定理……(32)
§ 7. 應用……(38)
§ 8. Feinstein 的三個不等式……(45)
§ 9. Feinstein 的基本引理……(49)
§10. Shannon 基本定理……(55)
結尾語……(62)
參考文獻……(63)

引　言

信息論是概率論及數理統計應用的一門新學問．它是由於要想了解和改進傳信技巧而在最近十年內才發展起來的．信息論可以說是由 C. E. Shannon 的論文發表[1](1948) 而誕生出來．自從它誕生出來以後，這一門學問就獲得了科學理論的成熟性與分支性，這種現象在數學發展中是少見的，這種情況恰恰可以與積分方程理論中 Fredholm 的文章發表時的情況相媲美．

信息論是討論如何得到最有效的傳送信息．傳送信息的手段叫作傳信系統．傳信系統由五個部分所構成，這五個部分之間的關係可以由下面的示意圖來表明：



1．信源．產生"報"的主體叫作信源．"報"可以是各種各樣的．例如一串字母(電報)，一個時間 t 的函數 $f(t)$ (無線電)，一個時間 t 及其它變數 x, y 的函數 $f(x, y, t)$ (在無色電視中 $f(x, y, t)$ 是在放影板上 (x, y) 點及時間 t 時的光強)．

2．播送器．它的作用在於把"報"變成適於渠道能傳送的"信號"．在電話中播送器把聲音的不同的壓力變成相應的電流的強度；在電報中播送器把字母拍成一串符號：點、橫和間隔等．

3．渠道．渠道卽是把信號由播送器傳到收錄器時所用到的

手段或媒介．它可以是一條電線，一段無線電的頻率等．在傳送時，信號可以被噪聲干擾，在上圖中，我們表示出來被噪聲源發出的噪聲干擾了之後的信號．

4．收錄器．恰是反播送器，它再把信號變成報．

5．收報人．即是受報的對象．

傳信問題最簡單的數學描述粗略的說是和統計中的參數估計問題相像．有一個參數 x 代表發出的報，x 在一個相當抽象的域或者至少是多維的域內變動．還有一個變數 y，通常也是相當抽象的，y 代表收到的報．在實際傳信中，所收到的報很少是原來發出的報的一個完整無誤的翻版，或者是原來所發出的報的完全可以斷定的改形．所以 y 就成了一個依賴於參數 x 的隨機變數．所謂傳信問題就是：給了一個 y 的抽樣來估計未知參數 x．

上述的模型有一個缺點，這一缺點導致我們必需對上述模型加以改進．當然在許多的技術領域中，有這樣的傳信問題，在其中確實是可以假設所收到的 y 不具有關於參數 x 的任何先驗知識；但是經驗告訴了我們，收報人對於參數 x 的可能值是會有很多的先驗知識的；例如給人寫信時往往寫錯了或拼錯了字，但收信人會利用他自己關於文字的先驗知識來判斷原意是什麼．所以作爲傳信問題的模型應該把這種可能性也考慮在內．

例如在英文裏，字母 E 出現的頻率就比 Q 出現的頻率大，TH 出現的頻率就比 XP 出現的頻率大等等．所以在研究傳信問題時，我們不應該再把發出的報本身看作是毫無拘束的，而應該把它看作是一個隨機變數，而此隨機變數的分佈恰恰反映了人們對於它的先驗知識．今後我們把傳信模型作如下的改進：我們有兩個隨機變數，一個爲 x，代表發出的報，一個爲 y，代表收到的報．對於新的模型來說，傳信問題仍然還可以看作是一個估計問題：在一個聯合抽樣 (x, y) 中，已知 y 值，試估計 x 值．Wiener 的平穩時間敍列的外推內插及光滑化一書就是採用了這種觀點，在 Wie-

ner 的工作中，x, y 都是數值的時間敍列，而且有一種很自然的數值方法來衡量 x 的眞值與估計值之偏差.

統計學家也可以把傳信問題看作是一個假設檢驗問題：所觀察的 y 具有一個分佈，此分佈依賴於假設 "x"，問題就是要決定在觀察時哪一個 x 是被收到了. Shannon 的工作就是採取這種觀點的. 在 Shannon 的工作裏時間敍列值是抽象值，但是他的結果對於一個錯的假設檢驗的"錯誤性"沒有一個自然的衡量法. 我們只能够得到一個取決的準則，也就是說在 (x, y) 空間中得到關於所採假設爲眞確時的全概率.

"假設之檢驗"與"估計"的差異並不是那麼突出的，對於 Shannon 的模型我們還用 "估計" 這個不嚴格的字眼. 下面我們將要提一提信息論與數理統計中通常研究 (x, y) 模型的不同的地方. 第一，在信息論中隨機變數 x, y 都是時間敍列，而且過去的事件(已經知道的)與未來事件 (不知道的) 之間的差異是很細心的被考慮到. 第二，信息論反映傳信工程的特殊興趣. 一個統計學家遇到上面的模型時，首先在他的面前有一個已知的二維分佈或者至少要選擇一個他所認爲合適的二維分佈，然後他的注意力集中於下列的問題：

(a)　用什麼準則來比較 x 的各種估計?

(b)　如果準則選定，那末什麼是 x 的最佳估計，其最佳程度如何?

(c)　各種估計 x 的競爭方法與最佳方法如何比較?

這些問題當然也在傳信問題中發生，而且 Wiener 的工作就是集中於這樣的普遍問題. 但在傳信工程裏，由於我們可以對二維分佈加以某種條件(下面就要談到)，於是對於上面的問題(b)更可深入地追問在這種條件下怎樣能够體現最佳估計.

在傳信工程裏，我們可以對二維分佈加以某種條件，在這裏所能够有的這種加條件的自由恰像一個試驗設計者所具有的自由一

樣：通常技術方面供給工程師一種通信手段；一個隨機變數 y，其活動範圍爲 Y，代表在收錄器所發生的事件，同時還有 y 的分佈，此分佈依賴於一個參數 x，x 的活動範圍爲 X，代表播送器產生出來的可能事件；此外還有另一隨機變數 θ，它的活動範圍爲 Θ，Θ 代表可能的"報"的空間，在此空間上有一個相應的概率測度.

我們還沒有談到"報" θ 與放在渠道入口的"刺戟" x 之間的關係. 此處恰恰就是傳信工程上可以加條件的地方，因爲必需考慮到過去事件與未來事件的差異，所以我們就可以有某種自由來選擇一個由 Θ 到 X 中的映像函數 $x = x(\theta)$ [參看後面 Shannon 第一定理的證明]. 這裏選擇映像函數相當於選擇一種拍報方法，使得原來的報轉換成適於渠道傳送的形狀. 爲了說明所加的條件的工作，讓我們設 y 的概率是 y 空間上的某一個測度 ν，此測度依賴於 x，設其密度爲 $\rho_x(y)$，θ 的概率是 θ 空間上的某一個測度 μ，設其密度爲 $\sigma(\theta)$，如果加上上述 x 與 θ 之間的關係的條件，則在插積空間 $\Theta \times Y$ 上的函數 $\sigma(\theta)\,\rho_{x(\theta)}(y)$ 就是對於測度乘積 $\mu \times \nu$ 而言的 θ, y 的聯合分佈的密度. 在信息論裏就是處理這個聯合分佈. 信息論的實際工作者對於 Shannon 工作所發生興趣的地方就是關於函數 $x(\theta)$ 的那些定理，亦卽如何選擇拍報方法的定理.

信息論在一開始的時候就對數學家提出來一系列的新問題，其中也有困難的問題. 信息論的早期工作者們懷着想得到最終結性的能行結果這一根本的目的，所以在一開始時，他們不能充分注意到很多數學上的困難. 他們不得不在他們的工作中的很多地方，或則滿意於不太嚴格的推理，或則爲了使證明簡易而勉強限制研究對象(信源，渠道，拍報等等). 因此信息論的早期工作都帶着數學上的不完全性，以致使得特別是要想學習這一新分支的數學家們感到很大的困難. 最近出版的 Goldman 的信息論教程還是如此.

1953 年 B. McMillan 發表了一篇文章，McMillan 的研究目的是打算使信息論獲得堅固的數學基礎，在他的工作中，離散源理論中的概念第一次得到了清晰的數學定義，此外他還證明了一個遍歷的零散信源具有漸近等同分割性，這一特性是信息論中一切漸近計算的基礎；而 Shannon 只能證明馬爾可夫型的信源具有這種特性．由於 McMillan 的定理，關於離散信源的一切理論可以不必局限於馬爾可夫型的信源，從而擴大了應用範圍．McMillan 還想利用 Shannon 的方法把 Shannon 的有關帶干擾渠道的基本定理的證明嚴格化，這時就發現了 Shannon 所給的證明草案中含有缺陷，這些缺陷甚至對於馬爾可夫型的信源而言，仍然是存在的，McMillan 對於這些缺陷的澄清只是註釋一下而未能徹底解決．

1954 年出現了 A. Feinstein 的論文，這篇論文仍以離散源爲對象．他充分地採用了 McMillan 的數學工具同時不追隨 Shannon 的證明路線而用了完全新的"字母行的有分別的組"的這一概念，終於找到了 Shannon 有關帶干擾渠道的基本定理的嚴格證明．但 Feinstein 只處理了簡單情形：信源發出的信號串中的信號是相互獨立的，渠道記憶力等於零．對於複雜的情形，Feinstein 只是簡略地提了一下，詳細表達出來時仍有很多的困難．

最近 А. Я. Хинчин 總結了這方面的工作，而且對 Shannon 的兩個定理分別加以證明．如所周知，Shannon 關於帶干擾渠道的定理給了兩個不同的陳述：第一個是以估計錯誤推測的概率來陳述的，第二個以傳送的可疑性（equivocation）這一量的估計來陳述的．McMillan 的分析結論說明這兩種陳述不是等價的，第一個比第二個結果更精確．Feinstein 的更細緻的研究工作指出來雖然第二個是第一個的推論，但是要想從第一個推出第二個來是有相當的困難的．同時又因這兩種陳述都有實際的意義，所以 Хинчин 還是分別陳述爲兩個定理而分別加以證明．

因爲離散情形是其它如連續情形或混合情形的基礎，離散情

形不只對信息論而且對計算機理論以及電話交換的設計都有用處，所以這裏仍以離散情形爲研究的對象，我們綜合了以往的結果而對離散情形給以全面的完整的介紹。

§1. 數學定義

令 A 爲一有窮集合，我們稱此集合爲字母集 (alphabet)．其中的元素叫作字母 (letter)．令 I 爲正負整數集合：$I=(\cdots -1, 0, 1, 2, \cdots)$．設字母集 A 爲已知，令 A^I 代表如下形狀的字母敍列所成的集合：

$$x=(\cdots x_{-1}, x_0, x_1, x_2, \cdots),$$

此處 $x_t \in A$，$t \in I$．x_t 稱爲在時候 t 時 x 的字母．

對於一個整數 t $(-\infty < t < \infty)$，一個正整數 n 及 A 中的字母 $a_0, a_1, \cdots, a_{n-1}$，我們考慮 A^I 中的那些 x，我們滿足下列條件：

$$x_{t+k}=a_k, \quad 0 \leqslant k \leqslant n-1;$$

所有的這些 x 組成了 A^I 中的子集合，此子集合稱爲基本集合，也就是 A^I 中的圓柱體 (Cylinder)．A^I 中的基本集合依賴於 t，n 及一串字母 $a_0, \cdots, a_{n-1}$．上面的條件所規定的基本集合記作 $[x_t=a_0, x_{t+1}=a_1, \cdots, x_{t+n-1}=a_{n-1}]$．

A^I 中一切基本集合(取不同的 t，n 及字母串 $a_0, \cdots, a_{n-1}$)產生一個 Borel 域，此 Borel 域記作 F_A．

有了以上的概念我們就可以數學地定義一個信源了．把字母集 A 中的字母看作是信源所能產生的字母，而把 x 看作是信源可能產生的一串字母而這一串字母卽組成一個報．如果令 μ 爲 Borel 域 F_A 上的一個概率測度，於是就得一個隨機過程 $[A^I, F_A, \mu]$．這一隨機過程卽是一個信源．基本集合就是一個事件，此事件就是說在 n 個固定的時候信源發出來 n 個固定的字母，此 n 個在相連 n 個時候所發出來的字母稱爲長爲 n 的字母行．大家熟

知在定義這一隨機過程時只需知道每一基本集合 Z 的概率 $\mu(Z)$ 就够了. 信源的統計特性完全由字母集 A 及 F_A 上的概率測度 μ 定出,所以信源記爲 $[A, \mu]$.

令 $[A, \mu]$ 爲一信源,而 T 爲坐標的推移變換 (shift), 即若 $x = (\cdots x_{-1}, x_0, x_1, \cdots)$ 則 $x' = Tx = (\cdots x'_{-1}, x'_0, x'_1 \cdots)$, 此處 $x'_t = x_{t+1}$, $t \in I$. T 把 F_A 中的元素 S 仍變爲 F_A 中的元素 TS (保持可測性). 我們說信源 $[A, \mu]$ 是平穩的 (stationary), 如果下列條件 (i) 被滿足;我們說信源是遍歷的 (ergodic), 如果下列條件 (i) 及 (ii) 同時被滿足:

(i) 若 $S \in F_A$, 則 $\mu(S) = \mu(TS)$,

(ii) 若 S 爲對 T 不變集合,亦即若 $S = TS$, 則 $\mu(S)$ 或爲零或爲 1.

平穩性就是說信源的輸出的概率狀況不隨時候而變化.

從實際的眼光來看,產生報或字母的機構就是信源,而傳送信號的機構就是渠道,所謂信號事實上就是某一集合中的元素,所謂信號就是字母. 所以欲對渠道下一數學定義, 必先知道這一渠道能够傳送的字母所組成的字母集,此字母集稱爲輸入字母集,我們可以設輸入字母集是有窮集合而記爲 A. A 中的字母叫作輸入字母. 通常,渠道所輸出的信號或字母就是收到的信號,輸出的信號可以與輸入者具有完全不同的性質. 所以我們也應有一個輸出字母集,此字母集也假定爲有窮集合而記爲 B.

如果輸入於渠道的字母爲 A 中的字母 a, 而在渠道的出口處就能得到一個 B 中的相應的字母 $b = b(a) \in B$, 那末我們就說此渠道無干擾. 通常由於有噪聲存在,當我們在不同的時候在渠道的入口處輸入同一字母 $a \in A$ 時, 我們在渠道出口處所得到的 B 中的字母是不同的, 這時我們就需要不同字母出現的概率, 它們的概率可以與輸入字母 a 及 a 前輸入的字母有關. 現在我們用最廣泛的形式來陳述這種依賴關係. 考慮 A^I 中的一切敍列

$$x = (\cdots x_{-1}, x_0, x_1, \cdots) \quad (x_i \in A, i \in I)$$

對於每一個這樣的敍列輸入於渠道後就在渠道的出口處收到一個敍列

$$y = (\cdots y_{-1}, y_0, y_1, \cdots) \quad (y_i \in B, i \in I).$$

任意一個 y_n 等於 B 中的固定字母 b 的概率在最廣泛的情形下可以認爲與所有輸入的字母 x_i 有關，亦即 $y_n = b$ 的概率是 $x \in A^I$ 的函數，把此概率記爲 $\nu_x(y_n = b)$. 這一概率就是收到的字母敍列 $y(\in B^I)$ 屬於圓柱體或基本集合 $y_n = b$ 的條件概率，而條件就是已知輸入的字母敍列爲 $x(\in A^I)$. 顯然，欲完全描寫一個渠道，不只應該知道這樣的基本集合：$y_n = b$ 的條件概率，而且也應該知道 B^I 中的任意基本集合 S 的條件概率 $\nu_x(S)$，由此自然對 Borel 域 F_B 中的任意集合的條件概率就知道了. 所以一個渠道決定於三個因素：1) 輸入字母集 A，2) 輸出字母集 B，3) 當輸入於渠道的是字母敍列 x 時，輸出的字母敍列 y 落於 F_B 中的任意集合 S 中的條件概率 $\nu_x(S)$. 因此渠道記爲 $[A, \nu_x, B]$.

渠道 $[A, \nu_x, B]$ 稱爲平穩的，如果對於任意 $x \in A^I$ 及 $S \in F_B$ 我們有 $\nu_{Tx}(TS) = \nu_x(S)$，此處 T 爲推移變換.

在應用問題中，$\nu_x(y_n = b)$，不依賴於 x_l，$l > n$. 這就是說對於一切 x 只要是它們的 $\cdots x_{n-1}$，x_n 都相同時，則 $\nu_x(y_n = b)$ 即具有同一數值. 這種情形稱之爲無預期性. 在很多的應用裏，$\nu_x(y_n = b)$ 不但只依賴於 x_n 及其以前的 x_k，$k < n$，而且只依賴於 x_n 以前的有限個 x_k，例如說只依賴於 $x_n, x_{n-1}, \cdots, x_{n-m}$. 這也就是說對於一切 x 當它們的 $x_{n-m}, \cdots, x_{n-1}, x_n$ 都相同時，則 $\nu_x(y_n = b)$ 即具有同一數值. 這時我們說渠道具有有限記憶. 使上述情形能够成立的最小整數 m 叫作渠道的記憶力. 當 $m = 0$ 時稱渠道無記憶力，這時 y_n 的條件概率只依賴於 x_n.

如果我們有一個信源 $[A, \mu]$ 和一個渠道 $[A, \nu_x, B]$，這時我

們即可利用此渠道以傳送由信源 $[A,\mu]$ 所發出的報. 信源 $[A,\mu]$ 所發出的每一個字母 x_n 可以作爲渠道 $[A, \nu_x, B]$ 的輸入字母, 經過傳送渠道即輸出一個字母 $y_n\in B$. 若由 $[A,\mu]$ 發出一個報

$$x = (\cdots, x_{-1}, x_0, x_1, \cdots),$$

則在渠道的出口處收到 B 中的一串字母

$$y = (\cdots, y_{-1}, y_0, y_1, \cdots).$$

這種安排我們稱之爲信源 $[A, \mu]$ 與渠道 $[A, \nu_x, B]$ 相結合. 用概率論的眼光來看這種結合乃是下屬兩種事件的結合: 1) 從 A^I 中取出一個隨機的報 x 及 2) 對於這一個所取定的 $x\in A^I$ 在渠道的出口處收到隨機的 $y\in B^I$.

今考慮 A 及 B 的插乘積 $C = A \times B$. C 即是由所有的字母偶 (a, b), $a\in A$, $b\in B$ 所構成的集合. 下列形狀的敍列

$$(x, y) = (\cdots (x_{-1}, y_{-1}), (x_0, y_0), (x_1, y_1), \cdots).$$

此處 $x_t\in A$, $y_t\in B$, $t\in I$, 組成一個集合記爲 C^I. C^I 也可以看作等於 $A^I \times B^I$. 若 $M\in F_A$, $N\in F_B$, 於是 $S = M \times N$ 是 C^I 中的一個集合. 對於這種 S 我們可以依條件概率定義算出 S 的概率 $\omega(S)$:

$$\omega(S) = \omega(M\times N) = \int_M \nu_x(N)\, d\mu(x). \tag{1.1}$$

特別言之, 當 S 是 C^I 中的基本集合時, 則 S 可以看作是 A^I 中某一基本集合 S_1 及 B^I 中某一基本集合 S_2 的插乘積. 依(1.1), C^I 中的基本集合的概率可以定出, 因之對 C^I 中基本集合產生的 Borel 域 F_C 中的任意集合的概率也唯一確定.

由此我們得到一個新的隨機過程 $[C, F_C, \omega]$, 這一過程也可以看作是一個信源, 此信源稱爲複式信源, 記爲 $[C, \omega]$. 容易證明下述事實:

若信源 $[A, \mu]$ 及渠道 $[A, \nu_x, B]$ 都是平穩的, 則由它們的結

合而產生的複式信源 $[C,\omega]$ 也是平穩的.

$[C,\omega]$ 相當於渠道輸入 x 及輸出 y 的聯合分佈，$[A,\mu]$恰恰相當於輸入 x 的邊際分佈. 輸出 y 的邊際分佈也是一個隨機過程，它等於 $[B^I, F_B, \eta]$，此處 η 由 (1.1) 直接導出等於

$$\eta(N)=\omega(A^I\times N)=\int_{A^I}\nu_x(N)\,d\mu(x)\,. \tag{1.2}$$

所以輸出 y 的邊際分佈相當於一個信源，記作$[B,\eta]$，容易證明：若 $[A,\mu]$ 及 $[A,\nu_x,B]$ 爲平穩的，則 $[B,\eta]$ 也是平穩的.

現在我們對於播送器及收錄器給以數學定義. 這兩個機器統稱之爲轉換器. 通常信源 $[A_0,\mu]$ 的字母集 A_0 與渠道 $[A,\nu_x,B]$ 的輸入字母集 A 不見得是同一字母集，所以欲用此渠道傳送由信源 $[A_0,\mu]$ 所發出的報 θ:

$$\theta=(\cdots,\theta_{-1},\theta_0,\theta_1,\cdots)\,.$$

此處 $\theta_t\in A_0$, $t\in I$, 必需經過一個轉換器使得 θ 中的字母轉換爲渠道所能接受的字母. 轉換器事實上就是一個把 θ 變到 x 的變換:

$$x=x(\theta)\,. \tag{1.3}$$

此處 $\theta\in A_0^I$, $x\in A^I$. 顯見，轉換器可以視爲無干擾的渠道，因爲這恰恰相當於輸入爲 θ，而輸出爲 θ 的一定的函數 $x(\theta)$. 所以我們可以把轉換器用渠道的記號記爲 $[A_0,\rho_\theta,A]$ 其中的 $\rho_\theta(M)$, $M\in A^I$ 等於

$$\rho_\theta(M)=\begin{cases}1, & \text{若 } x(\theta)\in M,\\ 0, & \text{若 } x(\theta)\notin M.\end{cases}$$

但是上述這樣廣泛的轉換器在實際上並不是有用的，例如假若對於能爲渠道傳送的報中的任意一個字母 x_k 我們必需知道整個的 θ 敍列，亦卽必需知道信源 $[A_0,\mu]$ 發出的字母無窮敍列，這是不切實用的；所以在實際上爲了能够定出要想傳送的字母 $x_k\in A$

(或者一個有窮長的字母行)時，只需知道某一個有窮串的 $\theta_k \in A_0$ 的轉換器才是眞正有用的轉換器．實際上今後我們可以看到，最方便的轉換器也可以說是最方便的拍報方法是把 θ 及 x 都分成某種長的字母行，而且把它們自左至右的編好號，然後依照一個規則使 θ 中的第 k 個字母行對應於 x 中的第 k 個字母行．

設我們欲用渠道 $[A, \nu_x, B]$ 傳送信源 $[A_0, \mu]$ 所發出的報，此處字母集 A_0 與 A 不相同，這時必需在信源 $[A_0, \mu]$ 及渠道 $[A, \nu_x, B]$ 之間裝置一個轉換器，使此轉換器把 $[A_0, \mu]$ 所發出的報 θ 拍發成爲能爲渠道 $[A, \nu_x, B]$ 傳送的報 x，然後把 x 輸入於渠道，於是在渠道的出口處收到一個報 $y \in B'$．這樣一來把信源 $[A_0, \mu]$ 經過轉換器而與渠道 $[A, \nu_x, B]$ 結合起來．所以轉換器與渠道聯接可以看作是一個新的渠道 $[A_0, \lambda_\theta, B]$，而 λ_θ 很容易計算．令 $S \in F_B$，因爲轉換器把 $\theta (\in A_0')$ 轉換爲 $x(\theta)$ $(\in A')$，所以 $y \in S$ 的條件概率 $\lambda_\theta(S)$ (條件爲 θ 已知) 等於對於渠道 $[A, \nu_x, B]$ 而言當 $x = x(\theta)$ 爲已知時 $y \in S$ 的條件概率，這就是說

$$\lambda_\theta(S) = \nu_{x(\theta)}(S).$$

由此可知渠道 $[A_0, \lambda_\theta, B] = [A_0, \nu_{x(\theta)}, B]$.

特別言之，信源 $[A_0, \mu]$ 及渠道 $[A, \nu_x, B]$ 經過轉換器而結合起來就產生出一個複式信源 $[C, \omega]$，此處 $C = A_0 \times B$，而概率 $\omega(M \times N)$，$M \in F_{A_0}$，$N \in F_B$ 等於

$$\omega(M \times N) = \int_M \lambda_\theta(N)\, d\mu(\theta) = \int_M \nu_{x(\theta)}(N)\, d\mu(\theta).$$

§2. 熵 的 概 念

上面已經把傳送系統的各個組成部分都給了數學定義．今後我們要研究如何利用這個系統得到效果最好的傳送．談到效果時就應該以某些量來衡量，這些量的基礎就是熵．設有一個有窮概

率場

$$A=\begin{pmatrix} A_1\cdots A_n \\ p_1\cdots p_n \end{pmatrix},$$

此處 $A_1, \cdots, A_n$ 爲所有可能的情況，而 $p_1, \cdots, p_n$ 爲這些情況的相應的概率．我們對這個概率場作試驗時，每作一次試驗 $A_i(i=1, \cdots, n)$ 中之一必定出現，但是事前我們不能肯定哪一個出現，所以說每一個概率場的試驗結果必具有一定的不肯定性．例如，我們就下列兩個概率場作試驗

$$\begin{pmatrix} A_1 & A_2 \\ 0.5 & 0.5 \end{pmatrix} \quad \begin{pmatrix} A_1 & A_2 \\ 0.99 & 0.01 \end{pmatrix}$$

很容易看出這兩個試驗的結果的不肯定性是不同的，第一個概率場的不肯定性要大．就第二個概率場作試驗時，其結果可以說"幾乎肯定"爲 A_1，但就第二個概率場作試驗時，我們絲毫不能給以預言．再看第三個概率場

$$\begin{pmatrix} A_1 & A_2 \\ 0.3 & 0.7 \end{pmatrix}$$

顯見這個概率場的不肯定性是介乎上述二者之間的．

我們希望找到一個數值以衡量概率場 A 的不肯定性，這種不肯定性表示在一次試驗裏它的結果是哪一個情況的先驗的不肯定性．當然這個數值要依賴於 $p_1, \cdots, p_n$．令此值記作 $H(p_1, \cdots, p_n)$．我們說它可以取成下列形狀：

$$H(p_1, \cdots, p_n)=-\sum_{i=1}^{n} p_i \log p_i, \tag{2.1}$$

此處對數基當然可以取爲任意一個不等於 1 的正數，但在信息論中通常取作 2，上式中若有一個 $p_i=0$，則其相應的項 $p_i \log p_i$ 了解爲零．由於 $H(p_1, \cdots, p_n)$和熱力學的熵有某些相似之處，我們

把 $H(p_1, \cdots, p_n)$ 叫作概率場 A 的熵. 我們由下討論可以看出這樣取的熵 $H(p_1, \cdots, p_n)$ 的確反映了我們對於先驗不肯定性的感性知識.

首先我們知道,當而且只當 $p_1, \cdots, p_n$ 之中的一個等於 1 時,熵 $H(p_1, \cdots, p_n) = 0$. 這是合乎道理的,因爲當未作試驗之前,我們完全可以肯定對應於 $p_i = 1$ 的那個情況 A_i 必然出現,因此就沒有什麼不肯定性了,所以熵應該等於 0. 對於其它的一切情形熵值爲正.

此外顯然可知,當 $A_1, \cdots, A_n$ 等同概然時,即 $p_i = \frac{1}{n}$ $(i = 1, \cdots, n)$ 時,這個概率場的不肯定性最大. 我們這樣取的熵 $H(p_1, \cdots, p_n)$ 恰有這種性質,當 $p_i = \frac{1}{n}$ $(i = 1, \cdots, n)$ 時, $H(p_1, \cdots, p_n)$ 之值爲最大. 欲證此,我們知道對於任意凹函數 $f(x)$ 有下列不等式:

$$f\left(\frac{1}{n}\sum_{i=1}^{n} b_i\right) \leqslant \frac{1}{n}\sum_{i=1}^{n} f(b_i).$$

此處 b_i 爲任意正數. 令 $f(x) = x\log x$ 並於上列不等式中取 $b_i = p_i$, 則得

$$f\left(\frac{1}{n}\right) = \frac{1}{n}\log\frac{1}{n} \leqslant \frac{1}{n}\sum_{i=1}^{n} p_i \log p_i = -\frac{1}{n}H(p_1, \cdots, p_n),$$

故

$$H(p_1, \cdots, p_n) \leqslant \log n = H\left(\frac{1}{n}, \cdots, \frac{1}{n}\right).$$

設有兩個概率場:

$$A = \begin{pmatrix} A_1 \cdots A_a \\ p_1 \cdots p_a \end{pmatrix}; \quad B = \begin{pmatrix} B_1 \cdots B_b \\ q_1 \cdots q_b \end{pmatrix}.$$

設此二概率場相互獨立. 由此可知 A_k, B_l 同時出現的概率 π_{kl} 等