MỤC LỤC

LỜI NÓI ĐẦU

Học phần lý thuyết thông tin là môn học thuộc nhóm các môn cơ sở ngành của ngành An toàn thông tin, đóng vai trò là nền tảng của những môn học khác thuộc nhóm chuyên ngành. Tuy vậy, hiện nay ở Học viện An ninh nhân dân chưa có giáo trình giành cho môn học này. Ngoài ra, hiện nay trong nước chưa có tài liệu nào về môn học này giành cho ngành Công nghệ thông tin và An toàn thông tin, phần lớn các giáo trình gần với môn học đều ở các ngành học khác như: điện, điện tử viễn thông vv..

Giáo trình được sử dụng cho môn học Lý thuyết thông tin giành cho ngành Công nghệ thông tin ở các trường đại học lớn trong nước thường viết bằng tiếng anh, trong đó có cuốn *Elements of Information Theory* của nhà xất bản *Wiley & Sons* thường được sử dụng. Đối với các lớp chuyên ngành Khoa Công nghệ và An ninh thông tin vẫn sử dụng tài liệu này làm tài liệu tham khảo chính, ngoài ra còn cuốn tập bài giảng môn lý thuyết thông tin được nhóm tác giả Phạm Văn Cảnh, Phạm Thị Hằng biên soạn đã được nghiệm thu cấp Khoa.

Với thực trạng trên, tập bài giảng cấp Học viện môn Lý thuyết thông tin được biên soạn nhằm mục đích: nâng cấp cuốn tập bài giảng đã có; bổ sung, tham khảo các tài liệu có giá trị trong và ngoài nước; bám sát chương trình khung của môn học. Cuốn tập bài giảng này sẽ cung cấp các Các kiến thức cơ bản trong lý thuyết thông tin, bao gồm: Entropy, Entropy hợp, có điều kiện, thông tin tương hỗ, mối liên hệ giữa các đại lượng đo lượng tin; nén dữ liệu; kênh truyền.

Cuốn tập bài giảng mới biên soạn sẽ là nguồn tài liệu thống nhất phục

vụ giảng dạy và học tập học phần Lý thuyết thông tin và là tiền để để nhóm tác giả tiếp tục biên soạn cuốn Giáo trình Lý thuyết thông tin sau này.

CHƯƠNG 1
MỞ ĐẦU

I Giới thiệu

Lý thuyết thông tin là một nhánh của toán học ứng dụng và kĩ thuật điện nghiên cứu về đo đạc lượng thông tin. Lý thuyết thông tin được xây dựng bởi *Claude E. Shannon*. Claude Elwood Shannon sinh ngày 30 tháng 4 năm 1916 - 24 tháng 2 năm 2001) là nhà toán học, kĩ sư điện tử, và mật mã học người Mỹ, được biết đến là "cha đẻ của lý thuyết thông tin".

Shannon nổi tiếng nhất vì đã xây dựng nên lý thuyết thông tin với bài báo năm 1948. Tuy nhiên ông cũng là người đưa ra lý thuyết thiết kế máy tính số và mạch số năm 1937, khi đang là một sinh viên cao học 21 tuổi tại Viện công nghệ Massachusetts, Hoa Kỳ (MIT), ông đã viết luận án chứng minh rằng ứng dụng điện tử của đại số Boole có thể xây dựng và giải quyết bất kì quan hệ số hay logic nào. Có người coi nó là luận án cao học quan trọng nhất của mọi thời đại. Shannon cũng đóng góp cho ngành phân tích mật mã trong chiến tranh thế giới thứ hai và sau đó, với những đóng góp cơ bản cho việc phá mật mã. để xác định giới hạn cơ bản trong các hoạt động xử lý tín hiệu chẳng hạn như nén dữ liệu hay lưu trữ và truyền dẫn dữ liệu.

Lý thuyết thông tin trả lời hai câu hỏi cơ bản trong lý thuyết truyền thông: Đâu là giới hạn nén dữ liệu và đâu là giới hạn truyền dữ liệu. Vì lý do này mà một số người coi lý thuyết thông tin là một phần của lý thuyết truyền thông. Theo chúng tôi thì nó rộng hơn thế nhiều. Thực

sự, lý thuyết thông tin có những đóng góp hết sức cơ bản cho vật lý học thống kê, khoa học máy tính, suy luận thống kê, và xác suất thống kê.

II Lịch sử phát triển

Sự kiện quan trọng nhất đánh dấu sự khởi đầu của lý thuyết thông tin là bài báo của Claude E. Shannon 'A Mathematical Theory of Communication' ở Bell System *Technical Journal* vào tháng 7 và tháng 10 năm 1948.

Trước bài báo này, một số ý tưởng về lý thuyết thông tin đã được phát triển tại Bell Labs, trong trường hợp đặc biệt khi tất cả các sự kiện đều có cùng xác suất. Bài báo năm 1924 của Harry Nyquist, 'Certain Factors Affecting Telegraph Speed', chứa một phần lý thuyết định lượng "tri thức" (intelligence) và "tốc độ đường truyền" (line speed), đưa ra mối liên hệ $W = K \log m$, trong đó W là tốc độ dẫn truyền tri thức, m là số cấp điện áp có thể sử dụng tại mỗi bước và K là một hằng số.

Bài báo năm 1928 của Ralph Hartley, 'Transmission of Information', sử dụng thuật ngữ 'thông tin' (information) như một đại lượng đo được, thể hiện khả năng phân biệt giữa các dãy kí hiệu của người nhận, do đó lượng hóa thông tin bởi $H = \log S^n = n \log S$, trong đó S là số kí hiệu có thể sử dụng, và n là số kí hiệu được truyền đi. Đơn vị tự nhiên của thông tin do đó là một chữ số thập phân, sau này được đổi tên là hartley để ghi danh đóng góp của ông, là một đơn vị đo thông tin.

Năm 1940, Alan Turing đã sử dụng những ý tưởng tương tự cho phân tích thông kê để phá bộ mã Enigma của Đức trong chiến tranh thế giới thứ hai.

Phần lớn lý thuyết toán học đằng sau lý thuyết thông tin với các sự

kiện có xác suất khác nhau được xây dựng trong ngành nhiệt động học bởi Ludwig Boltzmann và J. Willard Gibbs. Mối liên hệ giữa entropy thông tin và entropy nhiệt động học, bao gồm đóng góp quan trọng của Rolf Landauer trong thập kỉ 1960, được mô tả trong trang Entropy trong nhiệt động học và lý thuyết thông tin.

III Vai trò của lý thuyết thông tin

Ngay từ những ngày đầu, Lý thuyết thông tin đã mở rộng phạm vi ứng dụng ra nhiều lĩnh vực khác, bao gồm suy luận thống kê, xử lý ngôn ngữ tự nhiên, mật mã học, các mạng lưới bên cạnh mạng lưới viễn thông - chẳng hạn như trong thần kinh, sự tiến hóa và chức năngcủa các mã phân tử, lựa chọn mô hình trong sinh thái học, vật lý nhiệt, máy tính lượng tử, phát hiện sao chép và các hình thức phân tích dữ liệu khác.

Một độ đo cơ bản của thông tin là entropy, thường được diễn đạt dưới dạng số lượng bit cần thiết trung bình để lưu trữ hoặc dẫn truyền. Entropy lượng hóa sự không chắc chắn trong việc dự đoán giá trị của một biến ngẫu nhiên. Ví dụ như, xác định kết quả của một lần tung đồng xu công bằng (hai kết quả có khả năng như nhau) cho ít thông tin hơn (entropy nhỏ hơn) là xác định kết quả của một lần tung xúc sắc (sáu kết quả có khả năng như nhau).

Đối với kỹ thuật điện tử và khoa học máy tính, vào đầu thập niên 40 của thế kỷ trước người ta cho rằng không thể truyền tin với tốc độ dương với xác suất lỗi không đáng kể. Shannon đã làm giới nghiên cứu lý thuyết truyền thông phải kinh ngạc khi chứng minh rằng xác suất lỗi có thể nhỏ tùy ý cho tất cả các tốc độ truyền nhỏ hơn dung lượng kênh. Dung lượng này có thể được tính đơn giản dựa vào đặc điểm nhiễu của kênh truyền.

Shannon sau đó biện luận xa hơn rằng các quá trình ngẫu nhiên như âm nhạc hay tiếng nói có một độ phức tạp mà tín hiệu không thể được nén nhỏ hơn. Ông ta đặt tên đại lượng này là entropy, chiều theo sự sử dụng song song từ này trong nhiệt động lực học, và chỉ ra rằng nếu entropy của nguồn nhỏ hơn dung lượng kênh thì có thể gần như đạt được truyền thông không lỗi.

Nghiên cứu gần đây về các khía cạnh của lý thuyết thông tin tập trung vào lý thuyết thông tin mạng: lý thuyết về tốc độ truyền thông đồng thời từ nhiều người gửi tới nhiều người nhận trong điều kiện có xung đột và nhiễu.

Trong kỹ thuật nén dữ liệu, các ứng dụng cơ bản của lý thuyết thông tin bao gồm nén không mất dữ liệu (chẳng hạn như ZIP), nén mất dữ liệu (chẳng hạn MP3, JPG), mã hóa kênh (chẳng hạn như trong DSL). Lý thuyết thông tin nằm ở phần giao nhau giữa toán học, thống kê, khoa học máy tính, vật lý, thần kinh, và kĩ thuật điện. Các ngành hẹp quan trọng của lý thuyết thông tin bao gồm mã hóa nguồn, mã hóa kênh, lý thuyết thông tin thuật toán, bảo mật theo lý thuyết thông tin.

Đối với toán học, các đại lượng cơ bản của Lý thuyết thông tin – entropy, entropy tương đối, và thông tin tương hỗ – được định nghĩa như các hàm của các phân phối xác suất. Nó cũng giúp mô tả đặc điểm của các chuỗi biến ngẫu nhiên dài và cho phép chúng ta ước lượng xác suất của các sự kiện hiếm (lý thuyết độ lệch lớn) và tìm các số mũ lỗi trong kiểm định giả thuyết thống kê.

IV Cơ sở toán học

Trong mục này, ta sẽ nhắc lại một số kiến thức về lý thuyết xác suất cơ bản phục vụ cho các chương sau.

1. Phép thử ngẫu nhiên, Không gian mẫu và Biến cố

Ta sẽ gọi một thí nghiệm (thực hiện một số công việc nào đó) mà kết quả của nó không thể nào biết trước được là một *Phép thử ngẫu nhiên.*

Tập hợp tất cả các kết cục (outcome) có thể xảy ra khi thực hiện một phép thử ngẫu nhiên được gọi là *Không gian mẫu* và ký hiệu là Ω. Còn mỗi kết cục $\omega \in \Omega$ được gọi là một *biến cố sơ cấp* (hay là một *điểm mẫu*). Như thế không gian mẫu còn được gọi là *Không gian các biến cố sơ cấp.*

Ví dụ. Gieo cùng một lúc hai con xúc xắc thì không gian mẫu bao gồm 36 điểm mẫu

$$\Omega = \begin{bmatrix} (1,1) & (1,2) & (1,3) & (1,4) & (1,5) & (1,6) \\ (2,1) & (2,2) & (2,3) & (2,4) & (2,5) & (2,6) \\ (3,1) & (3,2) & (3,3) & (3,4) & (3,5) & (3,6) \\ (4,1) & (4,2) & (4,3) & (4,4) & (4,5) & (4,6) \\ (5,1) & (5,2) & (5,3) & (5,4) & (5,5) & (5,6) \\ (6,1) & (6,2) & (6,3) & (6,4) & (6,5) & (6,6) \end{bmatrix}, \tag{1.1}$$

trong đó kết cục $\omega = (i, j)$ có nghĩa là ở con xúc xắc thứ nhất xuất hiện i chấm và ở con xúc xắc thứ hai xuất hiện j chấm.

Mỗi tập hợp con A của không gian mẫu Ω được gọi là một *Biến cố.* Nói cách khác, ta xác định các *biến cố* là tất cả các tập hợp con $A \subset \Omega$. Khi thực hiện phép thử ngẫu nhiên, thì hoặc là kết cục xảy ra $\omega \in A$

hoặc là kết cục xảy ra $\omega \notin A$. Nếu $\omega \in A$, ta nói rằng ω là *thuận lợi* cho biến cố A .

Giả sử A và B là hai biến cố của không gian mẫu Ω . Nếu biến cố A xảy ra mà biến cố B cũng xảy ra, thì ta nói rằng biến cố A *kéo theo* biến cố B và ký hiệu là $A \subset B$, hoặc $B \supset A$ với nghĩa là "B được kéo theo bởi A".

Nếu biến cố A kéo theo biến cố B đồng thời biến cố B cũng kéo theo biến cố A , nghĩa là hai biến cố A và B kéo theo lẫn nhau, thì ta nói rằng hai biến cố A và B là *tương đương* và ký hiệu là $A = B$.

Với hai biến cố A và B của không gian mẫu Ω ta định nghĩa biến cố $A \cup B$ bao gồm tất cả các điểm mẫu thuộc vào biến cố A hoặc biến cố B. Tức là biến cố $A \cup B$ xảy ra nếu ít nhất một trong hai biến cố A hoặc B xảy ra. Biến cố $A \cup B$ thường được gọi là *hợp* (hoặc là *tổng*) của biến cố A và biến cố B .

Một biến cố được gọi là *chắc chắn* nếu nó luôn luôn xảy ra. Biến cố chắc chắn được ký hiệu là Ω, bởi vì khi thực hiện phép thử ngẫu nhiên thì chắc chắn thu được một điểm mẫu của không gian mẫu.

Với hai biến cố A và B , ta xác định một biến cố mới $A \cap B$ (cũng còn ký hiệu là AB), được gọi là *giao* (hoặc là *tích*) của A và B . Như thế AB bao gồm các điểm mẫu thuộc vào cả hai biến cố A và B . Tức là biến cố AB xảy ra nếu cả hai biến cố A và B đều xảy ra.

Một biến cố bao gồm các điểm mẫu thuộc vào biến cố A nhưng không thuộc vào biến cố B được gọi là *hiệu* của biến cố A với biến cố B và được ký hiệu là $A \setminus B$.

Định nghĩa IV.1 *Ta định nghĩa hiệu đối xứng của hai biến cố A và B*

bởi công thức

$$A\Delta B = (A \setminus B) + (B \setminus A) \ . \tag{1.2}$$

Hai biến cố A và $\overline{A}$ gọi là *đối lập* nếu thoả mãn các điều kiện sau:

$$A\overline{A} = \emptyset \quad , \ \texttt{và} \quad A + \overline{A} = \Omega \tag{1.3}$$

Rõ ràng với mỗi biến cố A ta có $\overline{A} = \Omega \setminus A$. Biến cố $\overline{A}$. Nói cách khác, với mỗi biến cố A ta xác định biến cố $\overline{A}$, gọi là biến cố *đối lập* của A , bao gồm những điểm của không gian mẫu Ω mà không thuộc vào A.

Nếu các biến cố $B_1, B_2, ..., B_n$ là xung khắc từng đôi một,

tức là $B_iB_j = \emptyset$ với $i \neq j$ và $A = B_1 + B_2 + \cdots + B_n$, thì ta nói rằng *A được phân chia thành các biến cố* $B_1, B_2, ..., B_n$.

Các biến cố $B_1, B_2, ..., B_n$ lập thành *nhóm đầy đủ các biến cố* nếu chúng thoả mãn các điều kiện sau:

$$B_iB_j = \emptyset \quad \texttt{với } i \neq j \quad , \ \texttt{và} \quad B_1 + B_2 + \cdots + B_n = \Omega \ . \tag{1.4}$$

Một số tính chất trong quan hệ giữa các biến cố:

Tính giao hoán: $A \cup B = B \cup A \quad , \quad AB = BA$.

Tính kết hợp: $A \cup (B \cup C) = (A \cup B) \cup C$,

$$A(BC) = (AB)C$$

Tính phân phối: $A(B \cup C) = AB \cup AC$,

$$A \cup (BC) = (A \cup B)(A \cup C) \ .$$

Tính luỹ đẳng: $A \cup A = A \quad , \quad AA = A$.

Tính đối ngẫu: $\overline{A \cup B} = \overline{A} \cap \overline{B}$,

$$\overline{A \cap B} = \overline{A} \cup \overline{B} \ .$$

2. Xác suất của các biến cố

Định nghĩa cổ điển

Giả sử rằng không gian mẫu Ω bao gồm một số hữu hạn các kết cục và khả năng xảy ra của chúng là như nhau. Chẳng hạn, gieo một con xúc xắc cân đối và đồng chất.

Xác suất $P(A)$ của biến cố A bằng số lượng các kết cục thuận lợi cho A chia cho số lượng các kết cục có thể xảy ra, tức là

$$P(A) = \frac{|A|}{|\Omega|} = \frac{m}{n} \tag{1.5}$$

trong đó $|A| = m$ chỉ số lượng điểm mẫu của A .

Ví dụ. Gieo hai con xúc xắc cân đối và đồng chất. Không gian mẫu bao gồm 36 kết cục có khả năng như nhau là bằng $\frac{1}{36}$.

Gọi A là biến cố tổng số chấm thu được bằng 12 thì A chỉ có một kết cục thuận lợi cho nó là $(6, 6)$, do đó $P(A) = \frac{1}{36}$.

Gọi B là biến cố tổng số chấm thu được bằng 11 thì các kết cục thuận lợi cho B là $(5, 6)$ và $(6, 5)$, do đó $P(B) = \frac{2}{36} = \frac{1}{18}$.

Các tính chất của xác suất theo định nghĩa cổ điển:

1) $0 \le P(A) \le 1$, với mọi biến cố A và $P(\emptyset) = 0$, $P(\Omega) = 1$.
2) Nếu $A \subset B$, thì $P(A) \le P(B)$.
3) Nếu $AB = \emptyset$, thì $P(A + B) = P(A) + P(B)$.
4) $P(\overline{A}) = 1 - P(A)$, với mọi biến cố A .
5) $P(A \cup B) = P(A) + P(B) - P(AB)$, với mọi cặp biến cố A, B.

3. Xác suất có điều kiện và các công thức cơ bản

Nhiều khi ta cần phải tính xác suất của biến cố A khi một biến cố B đã xảy ra với xác suất dương. Đó gọi là xác suất *có điều kiện* của biến cố A khi biến cố B đã xảy ra và ký hiệu là $P(A|B)$. Xác suất có điều kiện của biến cố A khi biến cố B đã xảy ra có thể tính theo công thức sau

$$P(A|B) = \frac{P(AB)}{P(B)} \quad , \textit{ trong đó } \quad P(B) > 0 \tag{1.6}$$

Tương tự, ta có thể định nghĩa

$$P(B|A) = \frac{P(AB)}{P(A)} \quad , \textit{ trong đó } \quad P(A) > 0 \tag{1.7}$$

Từ các công thức (1.6) và (1.7) ta suy ra công thức *nhân xác suất* sau đây:

$$P(AB) = P(A)P(B|A) = P(B)P(A|B) \tag{1.8}$$

Công thức (1.8) còn đúng ngay cả khi A hoặc B là các biến cố không. Chẳng hạn với $P(A) = 0$, ta có $P(A|B) = 0$ và $P(AB) = 0$.

Công thức nhân xác suất mở rộng cho n biến cố $A_1, A_2, ..., A_n$ như sau:

$$P(A_1A_2...A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2...A_{n-1}) \tag{1.9}$$

Xác suất có điều kiện cũng có những tính chất tương tự như xác suất

bình thường. Ngoài ra, ta còn có

$$P(B|B) = \frac{P(BB)}{P(B)} = \frac{P(B)}{P(B)} = 1 \ , \tag{1.10}$$

$$P(\emptyset|B) = 0 \quad , \tag{1.11}$$

$$P(\Omega|B) = 1 \quad , \tag{1.12}$$

$$P(A|B) = 1 \ , \ \textit{nếu } B \subset A \quad , \tag{1.13}$$

$$P(A|B) + P(\overline{A}|B) = 1 \quad , \tag{1.14}$$

tuy nhiên, nói chung

$$P(A|B) + P(A|\overline{B}) \neq 1 \quad , \tag{1.15}$$

$$P(A|B) + P(\overline{A}|\overline{B}) \neq 1 \quad . \tag{1.16}$$

Sự độc lập của các biến cố.

Ta nói rằng biến cố A *độc lập* với biến cố B nếu thoả mãn $P(A|B) = P(A)$, tức là sự xuất hiện của biến cố B không ảnh hưởng tới việc xuất hiện của biến cố A .

Nếu biến cố A độc lập với biến cố B , thì ta có

$$P(A)P(B|A) = P(B)P(A) \ , \tag{1.17}$$

suy ra $P(B|A) = P(B)$, tức là biến cố B cũng độc lập với biến cố A .

Từ đó ta nói rằng hai biến cố A và B là *độc lập với nhau*, hay nói đơn giản là *độc lập*.

Ta cũng thấy rằng nếu hai biến cố A và B là độc lập thì mỗi cặp biến cố $(\overline{A}, B), (A, \overline{B}), (\overline{A}, \overline{B})$ cũng độc lập. Chứng minh khẳng định này dành cho bạn đọc xem như là bài tập.

Khi hai biến cố A và B là độc lập thì công thức (1.3) trở thành

$$P(AB) = P(A)P(B) \tag{1.18}$$

vì thế người ta còn định nghĩa sự độc lập của hai biến cố A và B bằng công thức trên đây.

Các biến cố $A_1, A_2, ..., A_n$ được gọi là *độc lập*, hay *độc lập toàn bộ*, nếu với mỗi m $(1 \leq m \leq n)$ và với mọi $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ ta có

$$P(A_{i_1} A_{i_2} \cdots A_{i_m}) = P(A_{i_1}) P(A_{i_2}) \cdots P(A_{i_m}) \ . \tag{1.19}$$

Khi các biến cố $A_1, A_2, ..., A_n$ là độc lập thì công thức (1.9) trở thành

$$P(A_1 A_2 \cdots A_n) = P(A_1) P(A_2) \cdots P(A_n) \tag{1.20}$$

CHƯƠNG 2
ENTROPY

I Độ đo lượng tin

Mỗi một thông tin con người nhận được đều có khả năng xảy ra khác nhau. Vậy làm thế nào để có thể so sánh khả năng của các thông tin với nhau. Với mục đích như vậy *lượng tin* dùng để so sánh định lượng các tin với nhau, sự định lượng này nằm ở khả năng xảy ra của thông tin.

Xét một tin x có xác suất xuất hiện là $p(x)$, thì chúng ta có thể xem tin này là một tin trong một tập có $1/p(x)$ tin với các tin có xác suất xuất hiện như nhau. Nếu $p(x)$ càng nhỏ thì $1/p(x)$ càng lớn có nghĩa là "lượng tin" khi nhận được tin này cũng sẽ càng lớn.

Vậy lượng tin của một tin tỷ lệ thuận với số khả năng của một tin và tỷ lệ nghịch với xác suất xuất hiện của tin đó. Xác suất xuất hiện của một tin tỷ lệ nghịch với độ bất ngờ khi nhận được một tin.

Định nghĩa I.1 *Xét một nguồn $A = \{a_1, a_2, ..., a_m\}$ với các xác suất xuất hiện là $P(a_i), i = 1, .., m$. Lượng tin trong mỗi tin là a_i là $I(a_i)$ được định nghĩa bởi:*

$$I(a_i) = \log \frac{1}{p(a_i)} = -\log p(a_i) \tag{2.1}$$

Nếu cơ số logarit cơ số 2 thì đơn vị bits, nếu cơ số e thì đơn vị là nats, nếu cơ số là 10 thì đơn vị là Hartley.

Ví dụ. Cho một nguồn tin gồm 8 tin $U = \{u_0, u_1, u_2, u_3, u_4, u_5, u_6, u_7\}$, với các xác suất tương ứng là: $1/4, 1/4, 1/8, 1/8, 1/16, 1/16, 1/16, 1/16$.

Lượng tin cho mỗi tin là: $I(u_0) = 2$, $I(u_1) = 2$, $I(u_2) = 3$, $I(u_3) = 3$, $I(u_4) = 4$, $I(u_5) = 4$, $I(u_6) = 4$, $I(u_7) = 4$.

Ngoài ra chúng ta còn có thể định nghĩa lượng tin chứa trong một dãy $x = a_1a_2a_3...a_n$ với $a_i \in A$ độc lập với nhau là:

$$I(x) = \log \frac{1}{p(x)} = -\sum_{i=1}^{n} \log p(a_i) \tag{2.2}$$

Lượng tin trung bình của A được tính bằng công thức sau:

$$I(A) = \sum_{a_i \in A} p(a_i).I(a_i) = -\sum_{a_i \in A} p(a_i) \log p(a_i) \tag{2.3}$$

II Entropy

Phần này sẽ giới thiệu định nghĩa cơ bản của lý thuyết thông tin là Entropy. Với một nguồn tin cho trước, để định lượng mức độ không chắc chắn hay tính bất định của nguồn tin tức đó.

Định nghĩa II.1 *Entropy là độ đo mức độ không chắc chắn của một biến ngẫu nhiên cho trước. Giả sử X là một biến ngẫu nhiên rời rạc có tập giá trị là χ và hàm phân phối xác suất $p(x) = Pr\{X = x\}, x \in \mathcal{X}$. Khi đó Entropy của biến ngẫu nhiên X được định nghĩa bởi cộng thức:*

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \tag{2.4}$$

Với định nghĩa trên, ta thấy rằng một sự kiện xảy ra với xác suất càng nhỏ thì sự kiện đó càng ít xảy ra trong một tập sự kiện cho trước, cũng có nghĩa là tính không chắc chắn càng lớn. Một biến ngẫu nhiên càng biến cố có phân phối càng đều thì tính không chắc chắn càng lớn. Một phân phối càng lệnh nhau (có nhiều sự kiện có xác suất nhỏ và lớn) thì

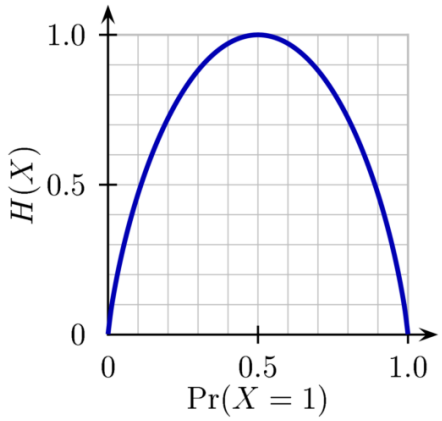
tính không chắc chắn càng ít do đó thì Entropy nhỏ hơn so với phân phối đều.

Nếu X là mộ biến ngẫu nhiên liên tục với hàm mật độ là $f(x)$ thì định nghĩa Entropy có thể được biểu diễn là:

$$h[f] = -\int_{-\infty}^{+\infty} f(x)\log_2 f(x) \tag{2.5}$$

Định nghĩa này thường được gọi là entropy Boltzmann hay entropy liên tục, hay entropy vi phân.

Ví dụ. Cho biến ngẫu nhiên $X = \{0, 1\}$ với xác suất $P(X = 1) = p, P(X = 0) = 1 - p$.
Khi đó Entropy của biến ngẫu nhiên X là: $H(X) = H(p) = -p\log p - (1-p)\log(1-p)$.



Hình 2.1 Quan hệ giữa p và H(p)

Ví dụ. Cho biến ngẫu nhiên $X = \{x_1, x_2, x_3\}$ và các xác suất: $p(x_1) = \frac{1}{2}, p(x_1) = \frac{1}{4}, p(x_1) = \frac{1}{2}$. Entropy của biến ngẫu nhiên đó là:

$$H(X) = H(\frac{1}{2}; \frac{1}{4}; \frac{1}{2}) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{4}\log\frac{1}{4} - \frac{1}{4}\log\frac{1}{4} = \frac{3}{2} \ \ (bit)$$

Ta cũng có thể viết $H(p)$ để thể hiện lượng tin trên. Từ giờ trở đi ta sẽ dùng ký hiệu log để thay cho logarit cơ số 2.

Nếu ta dùng khái niệm kỳ vọng thì Entropy được biểu diễn lại nhu sau:

$$H(X) = -\sum_{x\in\chi} p(x)\log_2 p(x) = \sum_{x\in\chi} p(x)\log_2 \frac{1}{p(x)} = E_p \log \frac{1}{p(x)} \quad (2.6)$$

Công thức trên ngụ ý rằng, Entropy của một biến ngẫu nhiên là kỳ vọng của biến ngẫu nhiên $1/\log(x)$ theo phân phối đầu vào của X.

III Entropy hợp, entropy có điều kiện và thông tin tương hỗ

Mục này sẽ định nghĩa Entropy của biến ngẫu nhiên hợp (X, Y). Ý tưởng chính của định nghĩa này là có thể xem cặp biến ngẫu nhiên trên nhu một vector trên không gian mẫu của chúng.

Định nghĩa III.1 *Entropy của hai biến ngẫu nhiên rời rạc X và Y với không gia mẫu là $\mathcal{X}, \mathcal{Y}$ có phân phối hợp là $p(x, y)$ được đỉnh nghĩa bởi:*

$$H(X, Y) = -\sum_{x\in X}\sum_{y\in Y} p(x, y)\log p(x, y) \quad (2.7)$$

Định nghĩa trên có thể biểu diễn dưới dạng kỳ vọng như sau: Hay:

$$H(X, Y) = -E\log p(X, Y) \quad (2.8)$$

Chúng ta cũng định nghĩa Entropy có điều kiện của một biến ngẫu nhiên khi đã biết một biến ngẫu nhiên khác như sau:

Định nghĩa III.2 *Entropy của biến ngẫu nhiên Y với điều kiện X với*

hàm phân phối xác suất hợp là $p(x, y)$ được định nghĩa như sau:

$$H(Y|X) = \sum_{x\in\mathcal{X}} p(x)H(Y|X = x) \tag{2.9}$$

$$= -\sum_{X\in\mathcal{X}} p(x) \sum_{Y\in\mathcal{Y}} p(y|x) \log p(x|y) \tag{2.10}$$

$$= -\sum_{X\in\mathcal{X}} \sum_{Y\in\mathcal{Y}} p(x, y) \log p(x|y) \tag{2.11}$$

$$= -E \log P(Y|X) \tag{2.12}$$

Trong định nghĩa trên, ta nhận thấy rằng Entropy có điều kiện, Entropy hợp có mối liên hệ với nhau và được khẳng định qua định lý sau,

Định lý III.1 (Luật xích)

$$H(X, Y) = H(X) + H(Y|X) \tag{2.13}$$

Chứng minh.

$$\begin{aligned} H(X, Y) &= -\sum_{x\in X} \sum_{y\in Y} p(x, y) \log p(x, y) \\ &= -\sum_{x\in X} \sum_{y\in Y} p(x, y) \log p(x)p(y|x) \\ &= -\sum_{x\in X} \sum_{y\in Y} p(x, y) \log p(x) - \sum_{x\in X} \sum_{y\in Y} p(x, y) \log p(y|x) \\ &= H(X) + H(Y|X) \end{aligned}$$

Hệ quả III.1 $H(X, Y|Z) = H(X|Z) + H(Y|X, Z)$

Ví dụ. Cho hai biến ngẫu nhiên X, Y có phân phối hợp như sau: Tính $H(X), H(Y), H(X|Y)$?

Y \ X	1	2	3	4
1	$\frac{1}{8}$	$\frac{1}{16}$	$\frac{1}{32}$	$\frac{1}{32}$
2	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{32}$	$\frac{1}{32}$
3	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
4	$\frac{1}{4}$	0	0	0

Hình 2.2 Phân phối hợp $P(X, Y)$

Phân phối thành phần của X là $(1/2, 1/4, 1/8, 1/8)$ và Y là $(1/4, 1/4/, 1/4, 1/4)$. Do đó: $H(X) = 7/4, H(Y) = 2$. Áp dụng luật xích, ta có:

$$H(X|Y) = H(X, Y) - H(Y)$$

Tính được:

$$\begin{aligned} H(X,Y) &= -\sum_{x=1}^{4}\sum_{y=1}^{4} p(x,y)\log_2 p(x,y) \\ &= -2\frac{1}{8}\log_2\frac{1}{8} - \frac{1}{4}\log_2\frac{1}{4} - 6\frac{1}{16}\log_2\frac{1}{16} \\ &= \frac{27}{8} \end{aligned}$$

Suy ra:

$$H(X|Y) = \frac{27}{8} - 2 = \frac{11}{8}$$

Giá trị cực đại và cực tiểu của Entropy được chứng minh qua định lý sau:

Định lý III.2 *Cho biến ngẫu nhiên X, ta luôn có:*

$$0 \leq H(X) \leq \log_2 |X|$$

Chứng minh. Bắt đẳng thức bên trái có thể suy ra trực tiếp từ định nghĩa về Entropy do tính không âm của các số hạng. Đẳng thức xảy ra khi tồn tại $x \in X, p(x) = 1$, các xác suất còn lại bằng 0.

Ta dễ dàng chứng minh được bất đẳng thức: $\ln(x) \leq x - 1$ với $x > 0$ đẳng thức đúng khi $x = 1$.

Đặt $x = \frac{q_i}{p_i}$ suy ra $\ln \frac{q_i}{p_i} \leq \frac{q_i}{p_i} - 1$ (đẳng thức đúng khi và chỉ khi $p_i = q_i$ với mọi i). Điều này tương đương với:

$$\leftrightarrow \sum_{i=1}^{M} p_i \ln \frac{q_i}{p_i} \leq \sum_{i=1}^{M} (q_i - p_i) = 1 - 1 = 0$$

$$\leftrightarrow -\sum_{i=1}^{M} p_i \ln p_i \leq -\sum_{i=1}^{M} \ln q_i$$

: $\ln x = \log_2 x / \log_2 e$

Suy ra:

$$-\sum_{i=1}^{M} \log_2 p_i \leq -\sum_{i=1}^{M} p_i \log_2 q_i$$

Đặt $q_i = \frac{1}{M}, \forall i$

Từ bổ đề ta có:

$$-\sum_{i=1}^{M} \log_2 p_i \leq -\sum_{i=1}^{M} p_i \log_2 \frac{1}{M} = \log_2 M \sum_{i=1}^{M} p_i = \log_2 M$$

Đẳng thức xảy ra khi và chỉ khi $p_i = \frac{1}{M}$.

IV Entropy tương đối và thông tin tương hỗ

Để định lượng sự khác nhau về Entropy giữa hai phân phối trên cùng một biến ngẫu nhiên ta có định nghĩa về khoảng cách Khullback-Leibler.

Định nghĩa IV.1 *Entropy tương đối hay khoảng cách Kullback-Leibler của biến ngẫu nhiên X trên không gian biến cố $\mathcal{X}$ giữa hai hàm xác suất $p(x)$ và $q(x)$ là:*

$$D(p||q) = \sum_{X \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)} = E_p \log_2 \frac{p(x)}{q(x)} \tag{2.14}$$

Từ định nghĩa ta thấy rằng Entropy tương đối không phải là một *khoảng cách* thực sự vì không có tính chất đối xứng và không thỏa mãn bất đẳng thức tam giác. Tuy nhiên có thể coi nó là một độ đo khoảng cách giữa hai phân phối xác suất trên một biến ngẫu nhiên.

Định nghĩa IV.2 *Giả sử X và Y là các biến ngẫu nhiên với hàm xác suất (pmf) tương ứng là $p(x)$ và $p(y)$. Thông tin tương hỗ $I(X,Y)$ giữa hai biến ngẫu nhiên này được tính bởi:*

$$I(X;Y) = D(p(x,y)||p(x)p(y)) = \sum_x \sum_y p(x,y) \log_2 \frac{p(x,y)}{p(x)P(y)} \quad (2.15)$$

Chú ý rằng thông tin tương tương hỗ cũng có thể gọi là thông tin chung. Ý nghĩa của đại lượng này là biến ngẫu nhiên này chứa bao nhiêu thông tin về biến ngẫu nhiên khác. Thông tin tương hỗ cũng là một trường hợp riêng của khoảng các tương đối, nó thỏa mãn các điều kiện của một khoảng cách.

Ví dụ. Giả sử $X = 0, 1$ và $p(x), q(x)$ là hai hàm xác suất trên X. Giả sử $p(0) = 1 - r, p(1) = r$ và $q(0) = 1 - s, q(1) = s$. Khi đó:

$$D(p||q) = (1-r) \log_2 \frac{1-r}{1-s} + r \log_2 \frac{r}{s} D(q||p) = (1-s) \log_2 \frac{1-s}{1-r} + s \log_2 \frac{s}{r}$$

Định lý dưới đây giới thiệu về mối quan hệ giữa Entropy, Entropy hợp, Entropy có điều kiện và thông tin tương hỗ.

Định lý IV.1

$$I(X,Y) = H(X) - H(X|Y)$$
$$I(X,Y) = H(Y) - H(Y|X)$$
$$I(X,Y) = H(X) + H(Y) - H(X|Y)$$
$$I(X,Y) = I(Y,X)$$
$$I(X,X) = H(X)$$

Chứng minh. Ta có:

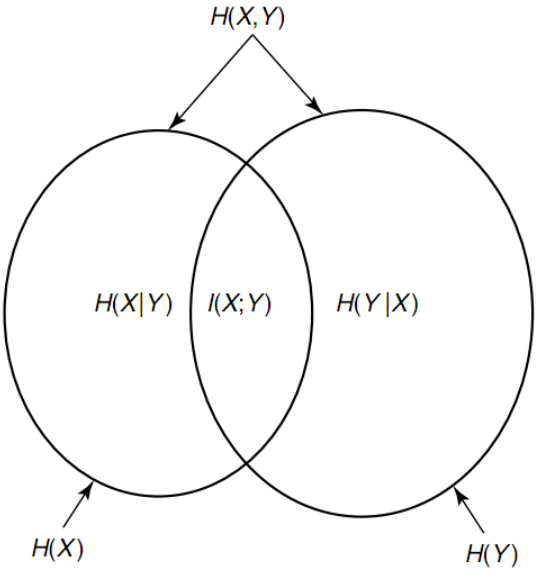
$$\begin{aligned} I(X;Y) &= \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \\ &= \sum_{x,y} p(x,y) \log \frac{p(x|y)}{p(x)} \\ &= -\sum_{x,y} p(x,y) \log p(x) + \sum_{x,y} p(x,y) \log p(x|y) \\ &= -\sum_{x} p(x) \log p(x) - (-\sum_{x,y} p(x,y) \log p(x|y)) \\ &= H(X) - H(X|Y). \end{aligned}$$

Các phần còn lại xem như bài tập cho sinh viên.

Để dễ hình dung mối liên hệ giữa các đại lượng này, ta có thể biểu diễn mỗi đại lượng qua biểu đồ Venn như hình 2.3.

Định lý IV.2 (Luật xích tổng quát) *Gọi $X_1, X_2, .., X_n$ là các biến ngẫu nhiên, ta có:*

$$H(X_1, X_2, .., X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, .., X_1) \tag{2.16}$$



Hình 2.3 Mối liên hệ giữa $H(X,Y), I(X,Y), H(X), H(Y)$

Chứng minh. Bằng phương pháp quy nạp

$$H(X_1, X_2) = H(X_1) + H(X_2|X_1), \tag{2.17}$$

$$H(X_1, X_2, X_3) = H(X_1) + H(X_2, X_3|X_1) \tag{2.18}$$

$$= H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1) \tag{2.19}$$

$$..... \tag{2.20}$$

$$H(X_1, X_2, .., X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1}, ..., X_1) \tag{2.21}$$

$$= \sum_{i=1}^{n} H(X_i|X_{i-1}, .., X_1) \tag{2.22}$$

Định nghĩa về luật xích cho thông tin tương hỗ có điều kiện

Định nghĩa IV.3 *Thông tin tương hỗ có điều kiện của biến ngẫu nhiên*

X và Y với điều kiện Z được định nghĩa bởi

$$I(X;Y|Z) = H(X|Z) - H(X|Y;Z) \tag{2.23}$$

$$= E_{p(x,y,z)} \log \frac{p(X,Y|Z)}{p(X|Z)P(Y|Z)} \tag{2.24}$$

Định lý IV.3 (Thông tin tương hỗ có điều kiện)

$$I(X_1, X_2, ..., X_n; Y) = \sum_{i=1}^{n} I(X_i; Y|X_{i-1}, .., X_2, X_1) \tag{2.25}$$

Chứng minh. Ta có:

$$I(X_1, X_2, .., X_n) = H(X_1, X_2, .., X_n) - H(X_1, X_2, .., X_n|Y) \tag{2.26}$$

$$= \sum_{i=1}^{n} H(X_i|X_{i-1}, .., X_1) - \sum_{i=1}^{n} H(X_i|X_{i-1}, .., X_1, Y) \tag{2.27}$$

$$= \sum_{i=1}^{n} I(X_i; Y|X_1, X_2, .., X_{i-1}) \tag{2.28}$$

V Bất đẳng thức xử lý dữ liệu

Phần này giới thiệu một số bất đẳng thức trong việc xử lý dữ liệu với mục tiêu chính là đánh giá thông tin chung của các biến ngẫu nhiên. Tư tưởng chính của việc đánh giá quá trình xử lý dữ liệu là đánh giá thông tin chung của đầu vào và đầu ra của một quá trình. Trong đó Khoảng cách tương đối là cơ sở cho quá trình đánh giá này.

Bất đẳng thức Jensen được dùng nhiều trong lĩnh vực xử lý thông tin. Một trong các đặc tính của nó là đánh giá trị trung bình của hàm thông qua đặc tính lỗi, lõm của hàm. Trong phần này chúng ta cũng giới thiệu bất đẳng thức này với mục đích làm cơ sở để xây dựng các bất đẳng thức thông tin.

Định nghĩa V.1 *Một hàm số $f(x)$ được gọi là hàm lồi trên khoảng (a, b) nếu với mỗi $x_1, x_2 \in (a, b)$ và $0 < \lambda < 1$, thì*

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2) \tag{2.29}$$

f được gọi là lồi nghiêm ngặt nếu đẳng thức xảy ra khi và chỉ khi $\lambda = 1$ hoặc $\lambda = 0$.

Chú ý rằng nếu f là hàm lồi thì $-f$ là hàm lõm và ngược lại.

Ví dụ. Hàm lồi: $x_2, |x|, e^x, x \log x$, (for $x \geq 0$). Hàm lõm: $\log x, \sqrt{x}$

Một tính chất chung của các hàm lồi (lòi ngặt) là đạo hàm cấp hai là không âm (dương). Tính chất này ngược lại đối với hàm lồi. Dưới đây là tình chất của kỳ vọng đối với hàm f có tính chất lồi hoặc lõm.

Định lý V.1 (Bất đẳng thức Jensen với kỳ vọng) *Nếu hàm f là một hàm lồi và X là một biến ngẫu nhiên thì:*

$$Ef(X) \geq f(EX) \tag{2.30}$$

Nếu f là lồi ngặt thì chỉ xảy ra dấu đẳng thức.

Chứng minh. Ta sử dụng phương pháp quy nạp theo tập không gian mẫu của biến ngẫu nhiên X. Với X là biến ngẫu nhiên nhận hai giá trị với hai xác suất là p_1, p_2, bất đẳng thức trở thành:

$$p_1 f(x_1) + p_2 f(x_2) \geq f(p_1 x_1 + p_2 x_2) \tag{2.31}$$

Điều này là hiển nhiên đúng do f là hàm lồi. Giả sử bđt đúng với $k-1, k > 3$, ta chứng minh bđt đúng với k. Đặt $p_i^{'} = p_i/(1 - p_k), i = 1, .., k - 1$, ta

có:

$$\sum_{i=1}^{k} p_i f(x_i) = p_k f(x_k) + (1-p_k) \sum_{i=1}^{k-1} p_i^{'} f(x_i) \tag{2.32}$$

$$\geq p_k f(x_k) + (1-p_k) f\left(\sum_{i=1}^{k-1} p_i^{'} x_i \right) \tag{2.33}$$

$$\geq f\left(p_k x_k + (1-p_k) \sum_{i=1}^{k-1} p_i^{'} x_i \right) \tag{2.34}$$

$$= f\left(\sum_{i=1}^{k} p_i x_i \right) \tag{2.35}$$

Đến đây ta được phải chứng minh!

Định lý V.2 (Bất đẳng thức xử lý thông tin) *Cho biến ngẫu nhiên X với hai hàm phân phối xác suất là $p(x), q(x)$ thì ta luôn có:*

$$D(p||q) \geq 0 \tag{2.36}$$

Đẳng thức xảy ra khi $p(x) = q(x)$

Chứng minh. Đặt $A = \{x : p(x) \geq 0\}$, ta có:

$$-D(p||q) = -\sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)} \tag{2.37}$$

$$= \sum_{x \in \mathcal{X}} p(x) \log \frac{q(x)}{p(x)} \tag{2.38}$$

$$\leq \log \sum_{x \in \mathcal{X}} p(x) \frac{q(x)}{p(x)}, (\texttt{Bất đẳng thức Jensen}) \tag{2.39}$$

$$= \log \sum_{x \in \mathcal{X}} q(x) \tag{2.40}$$

$$= \log 1 \tag{2.41}$$

$$= 0 \tag{2.42}$$

Đẳng thức xảy khi bất đẳng thức Jensen xảy ra. Vì $\log t$ là hàm lồi ngặt theo t nên đẳng thức xảy ra khi $q(x)/p(x)$ là hằng số. Kết hợp điều kiện xảy ra bất đẳng thức trong (2.40), ta được đẳng thức xảy ra khi $p(x) = q(x)$

Hệ quả V.1 *Với bất kỳ hai biến ngẫu nhiên X, Y thì:*

$$I(X, Y) \geq 0 \tag{2.43}$$

Đẳng thức xảy ra khi và chỉ khi X và Y là độc lập.

Chứng minh.

$$I(X, Y) = D(p(x, y)||p(x)p(y)) \geq 0. \tag{2.44}$$

Đẳng thức xảy ra khi và chỉ khi $p(x, y) = p(x) \cdot p(y)$ tức là X và Y độc lập với nhau

Hệ quả V.2

$$I(X; Y|Z) \geq 0 \tag{2.45}$$

Đẳng thức xảy ra khi và chỉ khi X và Y là độc lập khi có điều kiện Z

Định lý sau nói nên tính giảm Entropy khi đã biết thêm thông tin. Với hai biến ngẫu nhiên X và Y. Tính bất định của X giảm đi khi biết thông thêm thông tin về Y.

Định lý V.3 (Điều kiện giảm Entropy)

$$H(X|Y) \leq H(X) \tag{2.46}$$

Đẳng thức xảy ra khi X và Y độc lập

Chứng minh. Ta có: $0 \leq I(X,Y) = H(X) - H(X|Y)$

Định lý V.4 (Chặn độc lập Entropy) *Cho các biến ngẫu nhiên $X_1, X_2, .., X_n$ với xác suất hợp là $p(x_1, x_2, .., x_n$:*

$$H(X_1, X_2, ..., X_n) \leq \sum_{i=1}^{n} H(X_i) \tag{2.47}$$

Đẳng thức xảy ra khi $X_1, X_2, .., X_n$ độc lập

Chứng minh. Áp dụng luật xích cho Entropy, ta có:

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X_{i-1}, \ldots, X_1) \tag{2.48}$$

$$\leq \sum_{i=1}^{n} H(X_i) \tag{2.49}$$

Đẳng thức xảy ra khi và chỉ khi X_i độc lập với dãy biến ngẫu nhiên $X_{i-1}, \ldots, X_1$, tức là các biến ngẫu nhiên X_i độc lập với nhau

Để có thể ước lượng tính hiệu quả của quá trình truyền tin, ta cần mô hình trình truyền tin dưới dạng toán học. Trong thực tế, người ta thường sử dụng quá trình Markov để diễn tả quá trình truyền tin vì nó phù hợp với một quá trình tuyền tin tổng quát, trong đó trạng thái hiện tại chỉ phụ thuộc vào trạng thái trước đó.

Định nghĩa V.2 *Các biến ngẫu nhiên X, Y, Z là một quá trình Markov (ký hiêu $X \to Y \to Z$) nếu Z chỉ phụ thuộc Y và độc lập với X. X, Y, Z là một dạng của xích Markov $X \to Y \to Z$ nếu điều kiện được thỏa mãn:*

$$p(x, y, z) = p(x)p(y|x)p(z|y). \tag{2.50}$$

Các tính chất:

1. $X \to Y \to Z$, nếu và chỉ nếu X và Z độc lập có điều kiện khi biết Y, tức là: $p(x, z|y) = p(x|y) \cdot p(z|y)$. Thật vậy,

$$p(x, z|y) = \frac{p(x, y, z)}{p(y)} = \frac{p(x, y)p(z|y)}{p(y)} = p(x|y)p(z|y)$$

2. Nếu $X \to Y \to Z$ thì $Z \to Y \to X$

3. Nếu $Z = f(Y)$ thì $X \to Y \to Z$

Dựa trên các tính chất này, chúng ta có thể chứng minh một định lý quan trọng rằng có xử lý thông tin ở Y cũng không làm tăng thêm lượng thông tin có được về X trong quá trình Markov $X \to Y \to Z$.

Định lý V.5 (Bất đẳng thức xử lý dữ liệu) *Nếu $X \to Y \to Z$ thì*

$$I(X, Y) \geq I(X, Z) \tag{2.51}$$

Chứng minh. Ta khai triển thông tin tương hỗ theo hai cách khách nhau:

$$I(X; Y, Z) = I(X; Z) + I(X; Y|Z) \tag{2.52}$$

$$= I(X; Y) + I(X; Z|Y) \tag{2.53}$$

Vì X và Z là độc lập có điều kiện khi biết Y, nên $I(X; Z|Y) = 0$. Vì $I(X; Y|Z) \geq 0$, ta có:

$$I(X; Y) \geq I(X; Z)$$

Đẳng thức xảy khi và chỉ khi $I(X; Y|Z) = 0$ tức là $X \to Z \to Y$

Ý nghĩa của định lý này khi xử lý thông tin thì thông tin ta có được về nguồn luôn giảm.

Hệ quả V.3 *$X \to Y \to Z$ thì: $I(X; Y|Z) \leq I(X; Y)$*

Chứng minh. Vì

$$I(X;Y,Z) = I(X;Z) + I(X;Y|Z) \tag{2.54}$$

$$= I(X;Y) + I(X;Z|Y) \tag{2.55}$$

Và $I(X;Y|Z) = 0$, nên:

$$I(X;Z) + I(X;Y|Z) = I(X;Y) \tag{2.56}$$

Suy ra

$$I(X;Y|Z) \leq I(X;Y) \tag{2.57}$$

VI Bất đẳng thức Fano

Trong một kênh truyền có dạng $X \to Y \to Z$, giả sử ta đã biết được Y và muốn đoán (ước lượng) giá trị của X. Bất đẳng thức Fano cho ta biết giá trị chặn dưới của lỗi khi dự đoán. Lỗi này phụ thuộc vào entropy có điều kiện $H(X|Y)$. Giả sử ta muốn ước lượng biến ngẫu nhiên X với hàm phân phối là $p(x)$. Ta thấy ra với biến ngẫu nhiên Y là ước lượng liên quan đến X với phần phối điều kiện $p(y|x)$. Từ Y chúng ta xây dựng một ước lượng $g(Y) = \hat{X}$, trong đó $\hat{X}$ là một ước lượng của X. Ta thấy rằng $g(Y)$ là một biến ngẫu nhiên và $X \to Y \to \hat{X}$ là một dãy Markov. Gọi xác suất lỗi của ước lượng này là:

$$P_e = Pr\{X \neq \hat{X}\}$$

Chặn dưới của lỗi này được phát biểu dưới định lý sau:

Định lý VI.1 (Bất đẳng thức Fano) *Với mọi ước lượng $\hat{X}$ sao cho $X \to Y \to \hat{X}$, với $P_e = Pr\{X \neq \hat{X}\}$, ta có:*

$$H(P_e) + P_e \log |\mathcal{X}| \geq H(X|\hat{X}) \geq H(X|Y) \tag{2.58}$$

Dạng yếu của bất đăng thức:

$$1 + P_e \log |\mathcal{X}| \geq H(X|Y) \tag{2.59}$$

hoặc

$$P_e \geq \frac{H(X|Y) - 1}{\log |\mathcal{X}|} \tag{2.60}$$

Chứng minh. Định nghĩa biến ngẫu nhiên lỗi như sau:

$$E = \begin{cases} 1 & \text{Nếu } \hat{X} \neq X \\ 0 & \text{Nếu } \hat{X} = X \end{cases} \tag{2.61}$$

Sử dụng luật xích khai triển theo hai cách, ta có:

$$H(E, X|\hat{X}) = H(X|\hat{X}) + \underbrace{H(E|X, \hat{X})}_{=0} \tag{2.62}$$

$$= \underbrace{H(E|\hat{X})}_{\leq H(P_e)} + \underbrace{H(X|E, X)}_{\leq P_e \log \mathcal{X}} \tag{2.63}$$

Vì có điều kiện làm giảm Entropy nên $H(E|\hat{X}) \leq H(E) = H(P_e)$ và khi đã biết X và $\hat{X}$ thì hoàn toàn xác định được E nên $H(E|X, \hat{X}) = 0$, $H(E) = H(P_e)$.

Nên ta có:

$$H(X|E, \hat{X}) = P(E = 0)H(X|\hat{X}, E = 0) + P(E = 1)H(X|\hat{X}, E = 1) \tag{2.64}$$

$$\leq (1 - P_e).0 + P_e \log |\mathcal{X}| \tag{2.65}$$

Do $P(E = 0) = 1 - P_e$, $H(X|\hat{X}, E = 1) \leq H(X) \leq \log |\mathcal{X}|$. Kết hợp hai cách khai triển lại, ta có

$$H(P_e) + P_e \log |X| \geq H(X|\hat{X}) \tag{2.66}$$

Mặc khác, $X \to Y \to \hat{X}$ nên $H(X|\hat{X}) \geq H(X|Y)$, do đó ta có:

$$H(p) + p \log |X| \geq H(X|\hat{X}) \geq H(X|Y) \tag{2.67}$$

Từ bất đẳng thức Fano ta có suy ra hệ quả sau

Hệ quả VI.1 *Với hai biến ngẫu nhiên X và Y, cho $p = \Pr(X \neq Y)$*

$$H(p) + p \log(|\mathcal{X}| - 1) \geq H(X|Y) \tag{2.68}$$

Hệ quả VI.2 *Đặt $P_e = \Pr(X \neq \hat{A})$ và đặt: $\hat{X}$: $\mathcal{Y} \to \mathcal{X}$; thì:*

$$H(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|Y) \tag{2.69}$$

Chứng minh. Việc chứng minh hệ quả này tương tự bất đẳng thức Fano. Tuy vậy, lúc này chú ý rằng

$$H(X|E, \hat{X}) = \Pr(E = 0)H(X|\hat{X}, E = 0) + \Pr(E = 1)H(X|\hat{X}, E = 1) \tag{2.70}$$

$$\leq (1 - P_e)0 + P_e \log(|\mathcal{X}| - 1) \tag{2.71}$$

Vì từ $E = 0$, $X = \mathcal{X}$ và $E = 1$ vùng kết quả cảu X là $|\mathcal{X}| - 1$, chúng ta chặn trên Entropy có điều kiện bằng $\log(|\mathcal{X}| - 1)$

VII Tập điển hình

Một tập hợp lớn gồm nhiều mẫu cho trước thường được chia làm hai thành phần: thành thần điển hình mang những đặc trưng của tập hợp đó, thành phần không điển hình chứa nhiều phần tử không đặc trưng cho tập hợp đó. Trong lý thuyết thông tin, tập điển hình đóng vai trò quan trọng, đặc biệt trong lĩnh vực mã hóa, nén dữ liệu. Trong phần này chúng

ta sẽ áp dụng lý thuyết Entropy để xác định tập điển hình của một tập hợp cho trước. Cách tiếp cận này giống với luật số lớn trong lý thuyết xác suất.

Định nghĩa VII.1 (Sự hội tụ của biến ngẫu nhiên) *Cho dãy biến ngẫu nhiên $X_1, X_2, \ldots$, ta nói rằng dãy biến ngẫu nhiên này hội đến biến ngẫu nhiên X:*

1. *Theo xác suất nếu với mọi $\epsilon > 0, \Pr\{|X_n - X| > \epsilon\} \to 0$*
2. *Theo bình phương trung bình nếu $E(X_n - X)^2 \to 0$*
3. *Với xác suất bằng 1 nếu $\Pr\{\lim_{n\to\infty} = X\} = 1$*

Định lý VII.1 (Luật số lớn yếu) *Nếu các biến ngẫu nhiên $X_1, X_2, .., X_n$ là độc lập có cùng phần phối với X, thì giá trị trung bình gần tới kỳ vọng $E(X)$, tức là:*

$$\lim_{n\to\infty} \Pr\left(\frac{1}{n}\sum_{i=1}^{n}(X_i) - E(X) \geq \epsilon\right) = 0 \tag{2.72}$$

Định lý VII.2 (Tính chất tiệm cận phân hoạch đều) *Nếu $X_1, X_2, ..., X_n$ là các biến ngẫu nhiên độc lập với cùng phân phối xác suất là $p(x)$, thì ta có:*

$$-\frac{1}{n}\log p(X_1, X_2, .., X_n) \to H(X) \tag{2.73}$$

Chứng minh. Ta chú ý rằng nếu các biến ngẫu nhiên $X_1, X_2, .., X_n$ là độc lập thì hàm của các biến ngẫu nhiên này cũng độc lập. Do đó:

$$-\frac{1}{n}\log p(X_1, X_2, ..., X_n) = -\frac{1}{n}\sum_{i}\log p(X_i) \tag{2.74}$$

$$\to -E\log P(X) \tag{2.75}$$

$$= H(X) \tag{2.76}$$

Dựa trên kết quả này, ta định nghĩa tập điển hình như sau:

Định nghĩa VII.2 (Tập điền hình) *Tập điển hình $A_\epsilon^{(n)}$ với phân phối xác suất $p(x)$ là tập các dãy $(x_1, x_2, .., x_n) \in \mathcal{X}^n$ với thuộc tính*

$$2^{-n(H(X)+\epsilon)} \leq p(x_1, x_2, ..., x_n) \leq 2^{-n(H(X)-\epsilon)} \tag{2.77}$$

Định lý VII.3 (Các tính chất của tập điển hình) *Tập điển hình $A_\epsilon^{(n)}$ của X^n có những tính chất sau:*

1. *Nếu $(x_1, x_2, .., x_n) \in A_\epsilon^{(n)}$, thì $H(X) - \epsilon \leq -\frac{1}{n} \log p(x_1, x_2, .., x_n) \leq H(X) + \epsilon$*
2. *$Pr\{A_\epsilon^{(n)}\} > 1 - \epsilon$ với n đủ lớn.*
3. *$A_\epsilon^{(n)} \leq 2^{n(H(X)+\epsilon)}$, với $|A|$ là số phần tử của tập A.*
4. *$|A_\epsilon^{(n)}| \geq (1 - \epsilon)2^{2(H(X)-\epsilon)}$ với n đủ lớn.*

Chứng minh. 1. Tính chất này suy ra trực tiếp từ định nghĩa

2. Theo định lý tiệm cận phân hoạch đều ta có biến cố $(X_1, X_2, \ldots, X_n) \in A_\epsilon^{(n)}$ với xác suất $\to 1$ khi $n \to +\infty$. Do vậy với $\delta > 0$, tồn tại n_0 sao cho với mọi $n \geq n_0$, ta có

$$\Pr\left\{| - \frac{1}{n} p(X_1, X_2, \ldots, X_n)| < \epsilon\right\} > 1 - \delta$$

Đặt $\delta = \epsilon$, chúng ta được điều phải chứng minh

3. Ta có.

$$1 = \sum_{x \in X^n} p(x) \tag{2.78}$$

$$\geq \sum_{x \in A_\epsilon^{(n)}} p(x) \tag{2.79}$$

$$\geq \sum_{x \in A_\epsilon^{(n)}} 2^{-n(H(x)+\epsilon)} \tag{2.80}$$

$$= 2^{-n(H(x)+\epsilon)} |A_\epsilon^{(n)}| \tag{2.81}$$

4. Theo định nghĩa tập điển hình, với n đủ lớn, ta có $\Pr[A_\epsilon^{(n)}] > 1 - \epsilon$, nên:

$$1 - \epsilon < \Pr[A_\epsilon^{(n)}] \tag{2.82}$$

$$\leq \sum_{x \in A_\epsilon^{(n)}} 2^{-n(H(x)-\epsilon)} \tag{2.83}$$

$$= 2^{-n(H(x)-\epsilon)} |A_\epsilon^{(n)}| \tag{2.84}$$

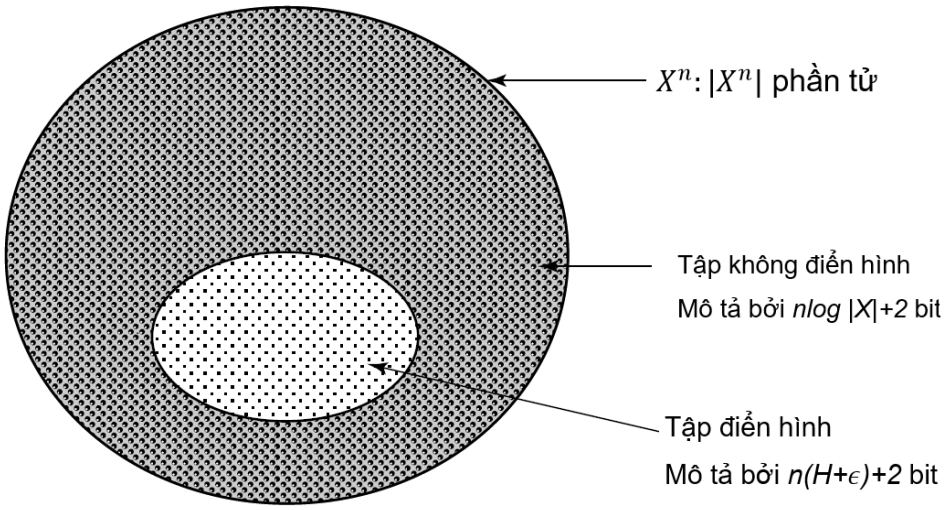
Từ đây suy ra:

$$|A_\epsilon^{(n)}| \geq (1 - \epsilon) 2^{-n(H(x)-\epsilon)} \tag{2.85}$$

Ta có sử dụng tập điển hình của một tập để mô tả lại tập hợp đó. Giả sử X là biến ngẫu nhiên có phân phối $p(X)$. Ta chia các chuỗi trong X^n ra thành hai tập: tập điển hình $A_\epsilon^{(n)}$ và phần còn lại.

Để mã há tập điển hình, ta thường sử dụng một số quy tắc sau:

- Vì không có quá $2^{-n(H(x)+\epsilon)}$ trong tập điển hình $A_\epsilon^{(n)}$, nên có thể dùng tối đa $-n(H(x) + \epsilon) + 1$ bits.



Hình 2.4 Tập điển hình và mô tả tập điển hình

- Thêm tiền tố 0 cho các dãy thuộc tập điển hình, số bits cần thiết là $-n(H(x)+\epsilon)+2$.
- Đối với các dãy không thuộc tập điển hình, số bit tối đa cần thiết là $n\log|X|+1$ bit.

Một số đặc trưng của việc mã hóa dữ liệu trên tập điển hình

- Mã là ánh xạ 1-1 và việc giải mã là dễ dàng. Bit đầu tiên đóng vai trò bit cờ cho biết độ dài từ mã sau nó.
- Sử dụng một cách đánh số brute-force cho tập điển hình cho dù số phần tử trong tập điển hình nhỏ hơn số phần tử trong X_n.
- Các chuỗi điển hình có mô tả ngắn xấp xỉ $nH(X)$.

Ký hiệu x^n biểu diễn cho dãy $x_1, x_2, \ldots, x_n$; $l(x^n)$ là độ dài của từ mã khi mã hóa x^n. Nếu n đủ lớn để $E(l(x^n)) = \sum_{x^n} p(x^n)l(x^n)$. Kỳ vọng độ dài

mã là

$$E(l(x^n)) = \sum_{x^n} p(x^n)l(x^n) = \sum_{x^n \in A_\epsilon^{(n)}} p(x^n)l(x^n) + \sum_{x^n \in A_\epsilon^{(n)c}} p(x^n)l(x^n) \tag{2.86}$$

$$\leq \sum_{x^n \in A_\epsilon^{(n)}} p(x^n)(n(H+\epsilon)+2) + \sum_{x^n \in A_\epsilon^{(n)c}} p(x^n)(n\log|X|+2) \tag{2.87}$$

$$\leq \Pr\{A_\epsilon^{(n)}\}(n(H+\epsilon)+2) + \Pr\{A_\epsilon^{(n)^c}\}(n\log|X|+2) \tag{2.88}$$

$$\leq n(H+\epsilon) + \epsilon n(\log|X|) + 2 = n(H+\epsilon') \tag{2.89}$$

Với $\epsilon' = \epsilon + \epsilon\log|X| + \frac{2}{n}$. Theo phân tích trên, ta có định lý sau:

Định lý VII.4 *Giả sử X^n là dãy biến ngẫu nhiên độc lập và cùng phân phối $p(x)$. Giả sử $\epsilon > 0$. Khi đó tồn tại một mã ánh xạ các chuỗi x^n dài n thành các xâu nhị phân sao cho ánh xạ là 1-1 và:*

$$E(\frac{1}{n}l(X^n)) \leq H(X) + \epsilon \tag{2.90}$$

Câu hỏi ôn tập

1. *Gieo đồng xu* Gieo một đồng xu cân đối đồng chất cho tới khi lần đầu tiên xuất hiện mặt ngửa. Gọi X là số lần gieo cần thiết:

   (a) Tìm độ đo thông tin entropy $H(X)$ bằng bit. Các biểu thức sau có thể được sử dụng:

   $$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r}; \qquad \sum_{n=0}^{\infty} nr^n = \frac{1}{(1-r)^2}.$$

   (b) Một biến ngẫu nhiên X được xác định theo phân bố trên. Hãy tìm một chuỗi *"hiệu quả"* các câu hỏi *có-không* theo mẫu "Có

phải X được chứa trong tập S?". So sánh $H(X)$ với trung bình số câu hỏi để xác định X.

2. *Entropy của hàm:* Giả sử X là một biến ngẫu nhiên nhận một số hữu hạn các giá trị. Quan hệ bất đẳng thức của Xv à Y là gì nếu:

   (a) $Y = 2^X$?

   (b) $Y = \cos X$?

3. *Entropy của các hàm của biến ngẫu nhiên:* Giả sử X là một biến ngẫu nhiên rời rạc. Hãy chỉ ra rằng entropy của một hàm của X là nhỏ hơn hoặc bằng entropy của X bằng cách kiểm chứng các bước sau:

$$\begin{aligned} H(X, g(X)) &\overset{(a)}{=} H(X) + H(g(X)|X) \\ &\overset{(b)}{=} H(X) \\ H(X, g(X)) &\overset{(c)}{=} H(g(X)) + H(X|g(X)) \\ &\overset{(d)}{\geq} H(g(X)). \end{aligned}$$

   Do đó $H(g(X)) \leq H(X)$.

4. *Entropy của một tổng:* Giả sử X và Y là các biến ngẫu nhiên nhận các giá trị lần lượt $x_1, x_2, \ldots, x_r$ và $y_1, y_2, \ldots, y_s$. Giả sử $Z = X + Y$.

   (a) Hãy chứng minh rằng $H(Z|X) = H(Y|X)$. Đồng thời hãy chỉ ra rằng nếu X, Y là độc lập thì $H(Y) \leq H(Z)$ và $H(X) \leq H(Z)$. Do đó phép cộng các biến ngẫu nhiên độc lập sẽ làm tăng thêm tính không chắc chắn của thông tin.

   (b) Hãy lấy ví dụ về các biến ngẫu nhiên trong trường hợp $H(X) > H(Z)$ và $H(Y) > H(Z)$.

(c) Với điều kiện nào thì $H(Z) = H(X) + H(Y)$.

5. *World series:* Trò chơi *World series* tạo bởi một chuỗi gồm 7 trận đấu được kết thúc ngay sau khi một đội nào đó thắng 4 trận. Giả sử X là một biến ngẫu nhiên biểu diễn kết quả của một world series giữa hai đội A và B; Các giá trị có thể có của X là AAAA, BABABAB và BBBAAAA. Giả sử Y là số trận đã đấu, có giá trị từ 4 đến 7. Giả sử A và B là ngang sức nhau và các trận đấu là độc lập. Hãy tính $H(X)$, $H(Y)$, $H(Y|X)$, $H(X|Y)$.

6. *Entropy cực đại:* Tìm hàm mật độ xác suất $p(x)$ làm cực đại entropy $H(X)$ của một biến ngẫu nhiên với giá trị nguyên, không âm X thỏa mãn ràng buộc:

$$EX = \sum_{n=0}^{\infty} np(x) = A, \quad A \text{ là một giá trị xác định nào đó, } A > 0.$$

Hãy tính giá trị cực đại của $H(X)$?

7. *Giá trị của một câu hỏi:* Giả sử $X \sim p(x)$, $x = 1, 2 \ldots, m$. Chúng ta được cho trước một tập $S \subseteq \{1, 2 \ldots, m\}$. Câu hỏi đặt ra liệu $X \in S$ hay không và nhận được câu trả lời $Y = \begin{cases} 1 & \text{nếu } X \in S \\ 0 & \text{nếu } X \notin S. \end{cases}$

Giả sử rằng $\Pr(X \in S) = \alpha$. Tìm độ suy giảm không chắc chắn $H(X) - H(X|Y)$.

8. *Câu hỏi ngẫu nhiên:* Một người mong muốn nhận ra đối tượng ngẫu nhiên $X \sim p(x)$. Một câu hỏi $Q \sim r(q)$ được đưa ra một cách ngẫu nhiên theo $r(q)$. Câu hỏi này dẫn đến một câu trả lời xác định $A = A(x, q) \in \{a_1, a_2, \ldots\}$, giả sử rằng X và Q là độc lập. Khi đó

$I(X; Q, A)$ là lượng không chắc chắn trong X bị loại bỏ bởi cặp câu hỏi-câu trả lời (Q, A).

(a) Chỉ ra rằng $I(X; Q, A) = H(A|Q)$. Hãy giải thích?

(b) Giả sử rằng hai câu hỏi i.i.d $Q_1, Q_2 \sim r(q)$ được hỏi, và các câu trả lời là A_1, A_2. Hãy chỉ ra rằng hai câu hỏi trên là kém giá trị hơn hai lần một câu hỏi đơn theo nghĩa

$$I(X; Q_1, A_1, Q_2, A_2) \leq 2I(X; Q_1, A_1).$$

9. *Thông tin tương hỗ của các mặt sấp và các mặt ngửa:*

(a) Tung một đồng xu cân đối và đồng chất. Thông tin tương hỗ giữa mặt trên và mặt dưới của đồng xu là gì?

(b) Tung một con xúc xắc 6 mặt. Thông tin tương hỗ giữa mặt trên và mặt trước là gì?

10. *Entropy có điều kiện bằng 0:* Hãy chỉ ra rằng nếu $H(Y|X) = 0$ thì Y là một hàm của X, có nghĩa là với mọi x mà $p(x) > 0$ chỉ có một giá trị y sao cho $p(x, y) > 0$.

11. *Xử lí dữ liệu:* Cho xích Markov $X_1 \to X_2 \to X_3 \to \cdots \to X_n$ với điều kiện $p(x_1, \ldots, x_n) = p(x_1)p(x_2|x_1) \cdots p(x_n|x_{n-1})$. Hãy biến đổi $I(X_1; X_2, \ldots, X_n)$ thành biểu thức tối giản.

12. *Thắt cổ chai:* Giả sử $X_1 \to X_2 \to X_3$ là xích Markov với xác suất $p(x_1x_2, x_3) = p(x_1)p(x_2|x_1)p(x_3|x_2)$, trong đó $x_1 \in \{1, 2, \ldots, n\}$, $x_2 \in \{1, 2, \ldots, k\}$, $x_3 = \{1, 2, \ldots, m\}$ và $k < n$, $m > k$ (n, k, m là các số nguyên dương.)

(a) Hãy chỉ ra sự phụ thuộc giữa X_1 và X_3 bị giới hạn bởi *thắt cổ chai* bằng việc chứng minh rằng $I(X_1; X_3) \leq \log k$.

(b) Tính biểu thức $I(X_1; X_3)$ với $k = 1$ và kết luận rằng không sự phụ thuộc nào có thể tiếp tục qua *thắt cổ chai* đó. (2 hình vẽ Hình 3.6 trang 36).

Một ví dụ về thắt cổ chai là lưu lượng giao thông qua quãng đường đang sửa chữa sẽ bị hạn chế. (xem hình 3.7 trang 37)

13. *Fano:* Cho bảng phân phối xác suất đồng thời của (X, Y) như sau:

XY	a	b	c
1	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$
2	$\frac{1}{12}$	$\frac{1}{6}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{6}$

Cho $\hat{X}(Y)$ là một ước lượng của X (dựa vào Y) và cho $P_e = \Pr(\hat{X}(Y) \neq X)$.

(a) Hãy tìm xác suất cực tiểu của ước lượng lỗi $\hat{X}(Y)$ và P_e tương ứng.

(b) Hãy tính bất đẳng thức Fano cho vấn đề này và so sánh.

14. *Entropy rời rạc:* Giả sử X và Y là hai biến ngẫu nhiên nhận giá trị nguyên và độc lập. Giả sử X có phân bố đều trên $\{1, 2, \ldots, 8\}$ và $P\{Y = k\} = 2^{-k}$, $k = 1, 2, 3 \ldots$.

(a) Tính $H(X)$.

(b) Tính $H(Y)$.

(c) Tính $H(X + Y, X - Y)$.

15. *Metric.* Một hàm $\varrho(x, y)$ được gọi là một *metric* nếu với mọi x, y,

- $\varrho(x, y) \geq 0$.

- $\varrho(x, y) = \varrho(y, x)$.
- $\varrho(x, y) = 0$ khi và chỉ khi $x = y$.
- $\varrho(x, y) + \varrho(y, z) \geq \varrho(x, z)$.

(a) Chứng minh rằng $\varrho(X, Y) = H(X|Y) + H(Y|X)$ thỏa mãn hai tính chất đầu tiên và tính chất thứ tư được nêu ở trên. Nếu nói rằng $X = Y$ khi và chỉ khi có một ánh xạ $1 - 1$ từ X vào Y, thì tính chất thứ ba cũng được thỏa mãn, tức là $\varrho(X, Y)$ là một metric.

(b) Kiểm chứng rằng $\varrho(X, Y)$ có thể được biểu diễn dưới dạng

$$\varrho(X, Y) = H(X) + H(Y) - 2I(X; Y) \tag{2.91}$$

$$= H(X, Y) - I(X, Y) \tag{2.92}$$

$$= 2H(X, Y) - H(X) - H(Y). \tag{2.93}$$

16. *Entropy của một phép trộn hỗn hợp.* Cho X_1, X_2 là các biến ngẫu nhiên rời rạc nhận được từ hàm phân bố xác suất $p_1(\cdot)$ và $p_2(\cdot)$ trên các bảng chữ cái tương ứng $\mathcal{X}_1 = \{1, 2, \ldots, m\}$ và $\mathcal{X}_2 = \{m + 1, \ldots, n\}$. Đặt

$$X = \begin{cases} X_1 & \text{với xác suất } \alpha, \\ X_2 & \text{với xác suất } 1 - \alpha. \end{cases}$$

(a) Hãy xác định $H(X)$ theo các thành phần của $H(X_1)$ và $H(X_2)$ và α.

(b) Hãy lấy giá trị lớn nhất theo α để chứng tỏ rằng bất đẳng thức $2^{H(X)} \leq 2^{H(X_1)} + 2^{H(X_2)}$ và từ đó chỉ ra rằng $2^{H(X)}$ là cỡ của bảng chữ cái nêu trên.

17. *Độ đo của sự tương quan.* Cho X_1 và X_2 là các biến ngẫu nhiên cùng phân bố nhưng không nhất thiết độc lập. Đặt

$$\varrho = 1 - \frac{H(X_2|X_1)}{H(X_1)}.$$

(a) Chứng minh rằng $\varrho = \frac{I(X_1;X_2)}{H(X_1)}$.

(b) Chứng minh rằng $0 \leq \varrho \leq 1$.

(c) Khi nào thì $\varrho = 0$? $\varrho = 1$?

18. *Entropy cực đại.* Tìm hàm phân bố xác suất $p(x)$ làm cực đại giá trị của entropy $H(X)$ của một biến ngẫu nhiên không âm giá trị nguyên với hạn chế

$$EX = \sum_{n=0}^{\infty} np(n) = A.$$

CHƯƠNG 3
NÉN DỮ LỆU

Nén dữ liệu còn được gọi là mã hóa dữ liệu, là chuyển dữ liệu ban đầu thành một dữ liệu khác với kích cỡ nhỏ hơn mà vẫn giữ được thông tin của nó. Việc nén dữ liệu có một giới hạn nhất định. Nói cách khác, dữ liệu được nén mà không làm mất mát thông tin sẽ có kích thước không nhỏ hơn một giới hạn nhất định. Để tìm hiểu khả năng của việc nén không mất mát thông tin, trong chương này chúng ta sử dụng định nghĩa và tính chất của Entropy để thiết lập nền tảng cho giới hạn của việc nén thông tin. Nén dữ liệu có thể thực hiện bằng việc gán sự mô tả ngắn gọn cho những tín hiệu có tần số xuất hiện cao, và cần sự mô tả dài cho những dữ liệu ít xuất hiện.

Trong chương này, chúng ta đưa ra các loại nén dữ liệu cơ bản. Từ đó xây dựng việc mã khóa không gây nhập nhằng (mã tiền tố) và tối ưu. Chương này cũng giới thiệu một số loại mã hóa tối ưu, cận tối ưu thường được sử dụng.

I Định nghĩa mã hóa dữ liệu

Đầu tiên, ta giới thiệu định nghĩa tổng quát nhất cho việc mã hóa.

Định nghĩa I.1 *Một mã nguồn C của biến ngẫu nhiên X là một ánh xạ từ miền giá trị của X, $\mathcal{X}$ tới tập các xâu có độ dài hữu hạn thuộc bảng chữ cái D-phân D^**

$$C : \mathcal{X} \to D^* \tag{3.1}$$

Trong đó $C(x)$ là từ mã tương ứng với một giá trị x, $l(x)$ là độ dài của

từ mã $C(x)$.

Không mất tính tổng quá từ đây về sau ta giả sử bảng chữ cái D-phân là $D = \{0, 1, 2, \ldots, D-1\}$. Trong định nghĩa này, mã hóa có thể hiểu như một hàm số. Để hình dung đơn giản ta lấy ví dụ như sau:

Ví dụ. Cho một phép mã hóa từ tập không gian mẫu $\mathcal{X} = \texttt{Xanh}, \texttt{Đỏ}$ tới bảng chữ cái nhị phân $D = \{0, 1\}$ với $C(\texttt{Xanh}) = 11, C(\texttt{Đỏ}) = 00$

Định nghĩa I.2 *Kỳ vọng độ dài của mã hóa nguồn* $L(C)$ *với biến ngẫu nhiên* X *với hàm xác suất là* $p(x)$ *là:*

$$L(C) = \sum_{x \in \mathcal{X}} p(x) l(x) \tag{3.2}$$

Ví dụ. Cho X là biến ngẫu nhiên với phân phối và từ mã được gán như sau:

$$\Pr(X = 1) = \frac{1}{2}, C(1) = 0$$
$$\Pr(X = 2) = \frac{1}{4}, C(2) = 10$$
$$\Pr(X = 3) = \frac{1}{8}, C(3) = 110$$
$$\Pr(X = 4) = \frac{1}{8}, C(4) = 111$$

Độ dài kỳ vọng là $L(C) = 1.75$ bits

Tiếp theo ta định nghĩa các loại mã cơ bản gồm: mã không kỳ dị, mã tách được, mã tiền tố.

Định nghĩa I.3 (Mã không kỳ dị) *Một phép mã hóa được gọi là không kỳ dị (nonsigular) nếu mỗi phần tử của nguồn* $\mathcal{X}$ *đều tương ứng với một phần tử trong* D^**, tức là:*

$$x \neq x' \rightarrow C(x) \neq C(x') \tag{3.3}$$

Tính không kỳ dị là điều kiện cần thiết để mô tả không nhập nhằng cho một từ dữ liệu trong X. Tuy vậy, chúng ta thường xuyên muốn gửi đi một dãy các phần tử của X. Trong trường hợp như vậy có thể xẩy ra nhập nhằng, tức là một từ mã có thể biểu diễn cho nhiều phần tử của nguồn. Do vậy, chúng ta cần những loại mã tốt hơn trong trường hợp này.

Định nghĩa I.4 (Mã mở rộng) *Mã mở rộng C^* của C là một ánh xạ từ một chuỗi hữu hạn của $\mathcal{X}$ vào D, được định nghĩa bởi:*

$$C(x_1x_2\dots x_n) = C(x_1)C(x_2)\dots C(x_n) \tag{3.4}$$

Trong đó $C(x_1)C(x_2)\dots C(x_n)$ chỉ việc nối các từ mã tương ứng.

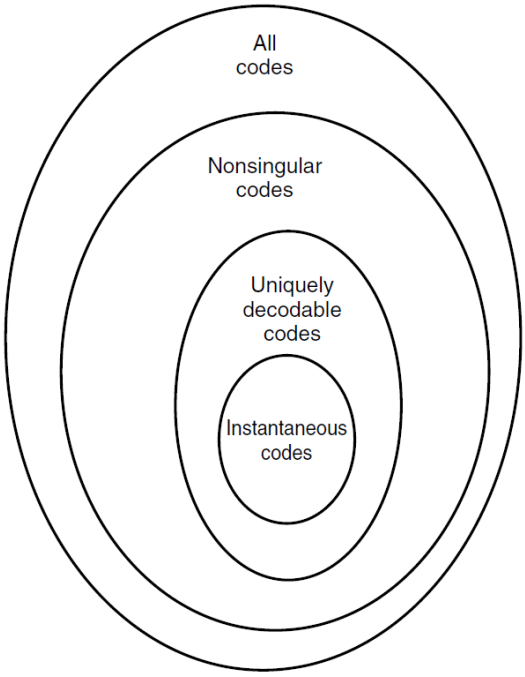
Ví dụ. Nếu $C(x_1) = 00$ và $C(x_2) = 11$, thì $C(x_1x_2) = 0011$

Định nghĩa I.5 (Mã tách được) *Một mã được gọi là giải mã duy nhất (uniquely decodedable) nếu phần mở rộng của nó là không kỳ dị.*

Nói cách khác, bất kỳ một chuỗi thuộc giải mã duy nhất, nó chỉ được tạo duy nhất một dãy nguồn. Tuy vậy, có thể phải xét toàn bộ chỗi đề xác định ký tự đầu tiên trong chuỗi nguồn tương ứng.

Định nghĩa I.6 (Mã tiền tố) *Một mã được gọi là tiền tố (prefix code) hoặc mã tức thời nếu không có từ mã nào là phần trước của từ mã khác.*

Sở dĩ mã tiền tố còn được gọi là mã tức thời (instantaneous) vì nó có thể được giải mã mà không cần tham khảo các từ tương tự trong tương lai và từ khi kết thúc một từ mã có thể được nhận dạng ngay lập tức. Do đó, đối với mã tức thời, ký tự x_i có thể được giải mã ngay khi chúng ta đến cuối mã từ tương ứng với nó. Chúng ta không cần phải chờ đợi để



Hình 3.1 Các loại mã hóa

xem các từ mã sau.

Ví dụ với dãy 01011111010 tạo bởi phép mã hóa nguồn $\mathcal{X} = \{1, 2, 3, 4\}$

$$C(1) = 0, C(2) = 10, C(3) = 110, C(4) = 111 \tag{3.5}$$

là mã tiền tố, do đó ta có thể xác định được dãy nguồn từ dãy được mã hóa. Hình 3.1 biểu diễn mối quan hệ giữa các loại mã hóa.

Để thấy rõ sự khác nhau giữa các loại mã, chúng ta xét các ví dụ các loại mã trong bản I. Trong bảng này, đối với mã không kỳ dị, chuỗi mã 010 ứng với ba ký tự trong nguồn là 2, 14 và 31, do đó nó không thể là mã tác được. Mã tác được không phải là mã tiền tố vì nó không thể giải mã một cách trực tiếp được. Để thấy rõ điều này, chúng ta chọn bất kỳ từ mã nào và bắt đầu từ bit đầu tiên. Nếu hai bit đầu tiên là 00 hoặc 10, thì có thể giải mã ngay lập tức, còn nếu hai bit đầu tiên là 11, chúng ta phải xét tiếp các bit sau. Trong trường hợp này, nếu bit tiếp theo là

1, thì nguồn đầu tiên là 3. Nếu chiều dài của chuỗi bit 0 sau bit 11 là lẻ, thì từ mã đầu tiên là 110 và nguồn đầu tiên là 4. Nếu chiều dài của bit 0 là lẻ thì nguồn đầu tiên là 3. Lặp lại quá trình này, chúng ta có thể tháy từ mã này không phải là tách được. Từ mã cuối cùng trong bảng I là mã tác được vì không có từ mã nào là tiền tố (thành phần trước với số phần tử bất kỳ) của từ mã nào.

X	Mã kỳ dị	Mã ko kỳ dị, không tách được	Mã tách được, không tiền tố	Mã tiền tố
1	0	0	10	0
2	0	010	00	10
3	0	01	11	110
4	0	10	110	111

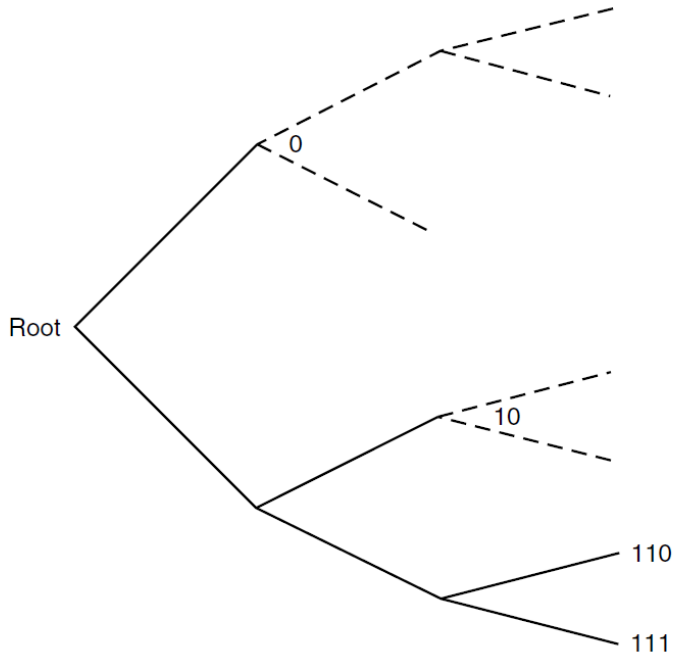
Bảng 3.1 Bảng ví dụ các loại mã

II Mã tối ưu

Một vấn đề cấp thiết đặt ra là dùng các từ mã có độ dài ngắn nhất mà vẫn biểu diễn được mã tiền tố. Vấn đề này có ý nghĩ lớn trong thực tế muốn tối ưu dung lượng để lưu trữ hoặc nâng cao tốc độ truyền thông tin.

1. Bất đẳng thức Kraft

Tập các độ dài mã có thể cho mã tiền tố bị giới hạn bởi sự bất bình đẳng sau.



Hình 3.2 Cây mã T được mở rộng thành cây D-phân T' với độ dài l_{max}

Định lý II.1 (Bất đẳng thức Kraft) *với mã tiền tố trên tập D, độ dài các từ mã là $l_1, l_2, \ldots, l_m$, ta có bất đẳng thức sau:*

$$\sum_i D^{-l_i} \leq 1 \tag{3.6}$$

Bắt đẳng thức này chỉ ra độ dài giới hạn trong việc mã hóa bằng mã tiền tố.

Chứng minh. Ta xây dựng câ mã D- phân tương ứng với mã tiền tố T (từ mã là lá). Sau đó mở rộng cây T thành cây D-phân T' với độ dài l_{max}, tổng số là của cây T' là $D^{l_{max}}$ như hình 3.2.

Ta thấy mỗi là của cây T' là hậu duệ của nhiều nhất một lá của cây T. Lá của cây T tương ứng với từ mã $C(X_i)$ có đúng $D^{l_{max}-l_i}$ hậu duệ trên cây T' (là 1 nếu $l_i = l_{max}$).

Tổng tất cả các là của cây nhỏ hơn cây T' nên ta có:

$$\sum_{i=1}^{|\mathcal{X}|} D^{l_{max}-l_i} \leq D^{l_{max}} \tag{3.7}$$

Chia đều cho $D^{l_{max}}$ ta được điều phải chứng minh.

Đối với bất đẳng thức Kraft, ta có một số chú ý như sau:

1. Đẳng thức trong BĐT Kraft xảy ra khi cây mã là cây D phân cân đối, đầy đủ, hoàn toàn. Từ điều này suy ra $p_i = D^{-l_i}, \forall i = 1..n$
2. Ý nghĩa của BĐT Kraft chỉ ra độ dài giới hạn của các từ mã, mã tiền tố không thể cho độ dài từ mã quá ngắn.
3. Đối với trường hợp tập từ mã là vô hạn có dạng mã tiền tố, ta cũng có:

   $$\sum_{i}^{\infty} D^{-l_i} \leq 1 \tag{3.8}$$

   Đây gọi là bất đẳng thức Kraft mở rộng.

2. Mã tối ưu

Chúng ta xét bài toán: Cho nguồn $\mathcal{X}$ có m phần tử với phân phối $p(x)$, hãy mã hóa nguồn với mã tiền tố trên bảng chữ cái D với độ dài từ mã tương ứng là $l_1, l_2 \ldots, l_m$ sao cho tổng độ dài kỳ vọng là nhỏ nhất. Theo đó ta cần tìm giá trị nhỏ nhất của

$$L = \sum_{i} p_i l_i \tag{3.9}$$

Theo bất đẳng thưc Kraft, các giá trị $l_1, l_2, \ldots, l_m$ thỏa mãn điều kiện sau:

$$\sum_{i} D^{-l_i} \leq 1 \tag{3.10}$$

Ta có thể sử dụng phương pháp nhân tử Lagrange để tìm lời giải tối ưu cho bài toán trên. Cụ thể như sau: Đặt

$$J = \sum p_i l_i + \lambda(\sum_i D^{-l_i}) \tag{3.11}$$

Lấy đạo hàm theo các chiều l_i, ta được

$$\frac{\partial J}{\partial l_i} = p_i - \lambda D^{-l_i} \ln D \tag{3.12}$$

Ta được các điểm tới hạn:

$$D^{-l_i} = \frac{p_i}{\ln D} \tag{3.13}$$

Thay thế đẳng thức trên vào biểu thức ban đầu và giải bài toán theo λ, ta được $\lambda = \frac{1}{\ln D}$, và tìm được lời giải tối ưu là:

$$p_i = D^{-l_i} \tag{3.14}$$

Độ dài của mã tối ưu là:

$$l_i^* = -\log_D p_i \tag{3.15}$$

Độ dài kỳ vọng là:

$$L = \sum p_i l_i = \sum -\log_D p_i = H_D(X) \tag{3.16}$$

Tuy nhiên l_i là số nguyên dương nên ta cần tìm một giá trị nguyên gần với l_i^* nhất.

Định lý II.2 *Độ dài kỳ vọng của mã tiền tố của biến ngẫu nhiên X với phân phối x.s là (X) lớn hơn hoặc bằng Entropy của biến ngẫu nhiên X với cơ số D, tức là*

$$L \geq H_D(X) \tag{3.17}$$

Đẳng thức xảy ra khi $p_i = D^{-l_i}$

Chứng minh.

$$L - H_D(X) = \sum p_i l_i - \sum p_i \log_D \frac{1}{p_i} \tag{3.18}$$

$$= -\sum p_i \log_D D^{-l_i} + \sum p_i \log_D p_i \tag{3.19}$$

Đặt $r_i = \frac{D^{-l_i}}{\sum_j D^{-l_j}}$ và $c = \sum_i D^{-l_i}$, ta có:

$$L - H_D(X) = -\sum p_i \log_D D^{-l_i} + \sum p_i \log_D \sum D^{-l_j} \tag{3.20}$$

$$+ \sum p_i \log_D p_i - \sum p_i \log_D \sum_j D^{-l_j} \tag{3.21}$$

$$= -\sum p_i \log_D \frac{D^{-l_i}}{\sum_j D^{-l_j}} + \sum p_i \log_D p_i - \sum p_i \log_D c \tag{3.22}$$

$$L - H_D(X) = -\sum p_i \log_D r_i + \sum p_i \log_D p_i - \sum p_i \log_D c \tag{3.23}$$

$$= \sum p_i \log_D \frac{p_i}{r_i} - \sum p_i \log_D c \tag{3.24}$$

$$= \sum p_i \log_D \frac{p_i}{r_i} + \sum p_i \log_D \frac{1}{c} \tag{3.25}$$

$$= D(p||r) + \log_D \frac{1}{c} \tag{3.26}$$

$$\geq 0 \tag{3.27}$$

Do khoảng cách tương đối không âm và $c \leq 1$.

Đẳng thức xảy ra khi $p_i = D^{-l_i}$ tức là $-\log_D p_i$ là số nguyên với mọi i. Khi đó ta gọi X có phân phối D-adic

Định lý này chỉ ra rằng điều kiên cần thiết để đẳng thức xảy ra khi nguồn cần có phân phối D-adic. Phần sau chúng ta sẽ chỉ ra giới hạn (chặn trên, dưới) của mã tối ưu.

3. Chặn độ dài mã tối ưu

Từ biểu thức $L - H_D(X) = D(p||r) - \log_D(\sum D^{-l_j})$ và định lý trên ta thấy rằng khi $l_i = \log_D \frac{1}{p_i}$ thì $L = H$.

Tuy nhiên điều kiện này không đảm bảo l_i là nguyên, do vậy ta cần làm tròn để được giá trị gần tối ưu nhất. Do đó, ta chọn:

$$l_i = \lceil \log_D \frac{1}{p_i} \rceil \tag{3.28}$$

Cách chọn này thỏa mãn bất đẳng thức Kraft, vì:

$$\sum D^{-\lceil \log_D \frac{1}{p_i} \rceil} \leq \sum D^{-\log_D \frac{1}{p_i}} \leq \sum_i p_i \leq 1 \tag{3.29}$$

Với cách chọn này, ta có

$$\log_D \frac{1}{p_i} \leq l_i < \log_D \frac{1}{p_i} + 1 \tag{3.30}$$

Nhân với các p_i rồi tổng ta có:

$$H_D(X) \leq L < H_D(X) + 1 \tag{3.31}$$

Vì một mã tối ưu chỉ có thể tốt hơn mã này, chúng ta có định lý sau đây.

Định lý II.3 *Gọi $l_1^*, l_2^*, \ldots, l_m^*$ là độ dài tối ưu của các từ mã với phân phối p trên bảng chữ cái D, gọi L^* là độ dài kỳ vọng của từ mã, ta có:*

$$H_D(X) \leq L < H_D(X) + 1 \tag{3.32}$$

Chứng minh. Đặt $l_i = \lceil \log_D \frac{1}{p_i} \rceil$. Theo phân tích trên thì các số l_i thỏa mãn BĐT Kraft và theo 3.32, ta có:

$$H_D(X) \leq L = \sum p_i l_i < H_D(X) + 1 \tag{3.33}$$

Vì L^* là độ dài mã tối ưu nên nó nhỏ hơn hoặc bằng L và vì $L^* \geq H_D$ nên ta có điều phải chứng minh.

Dựa trên định lý trên ta thấy rằng số bít để mã hóa một ký tự trong nguồn hơn entropy H_D nhiều nhất là 1 bit. Tuy vậy nếu xét một dãy ký tự $(x_1, x_2, \ldots, x_n)$ thì số bit thừa này sẽ được giảm thế nào. Chúng ta sẽ cùng làm rõ hơn trong các phân tích sau. Giả sử chúng ta có n ký tự trên $\mathcal{X}^n$ cần mã hóa. Gọi L_n là độ dài trung bình của dãy ký tự, L là độ dài kỳ vọng của từ mã chia độ dài của dãy, ta có:

$$L_n = \frac{1}{n} \cdot L = \frac{1}{n} \sum p(x_1, x_2, \ldots, x_n) l(x_1, x_2, \ldots, x_n) \tag{3.34}$$

$$= \frac{1}{n} El(x_1, x_2, \ldots, x_n) \tag{3.35}$$

Bằng phân tích như trên ta có:

$$H(X_1, X_2, \ldots, X_n) \leq L \leq H(X_1, X_2, \ldots, X_n) + 1$$

Vì $X_1, X_2, \ldots, X_n$ là độc lập nên ta có

$$H(X_1, X_2, \ldots, X_n) = \sum H(X_i) = nH(X) \tag{3.36}$$

Chia cả hai vế cho n, ta được:

$$H(X_1, X_2, \ldots, X_n) \leq L_n < H(X_1, X_2, \ldots, X_n) + \frac{1}{n} \tag{3.37}$$

Nếu quá trình trên là quá trình dừng thì $H(X_1, X_2, \ldots, X_n)/n \to H(\mathcal{X})$ độ dài kỳ vọng sẽ hội tụ đến tỷ lệ entropy khi $n \to \infty$. Tóm lại kết quả trên chúng ta có định lý sau:

Định lý II.4 *Độ dài kỳ vọng của từ mã trên một ký tự thỏa mãn:*

$$H(X_1, X_2, \ldots, X_n) \leq L_n^* < H(X_1, X_2, \ldots, X_n) + \frac{1}{n} \tag{3.38}$$

Hơn nữa, nếu $X_1, X_2, \ldots, X_n$ là một quá trình dừng, ta có:

$$L_n^* \to H(\mathcal{X}) \tag{3.39}$$

Định lý này giải thích rằng tỷ lệ entropy là số bit dự kiến cho mỗi biểu tượng yêu cầu để mô tả quá trình.

III Mã Huffman

1. Ví dụ và thuật toán

Năm 1952, nhà khoa học người Đức David Albert Huffman [1] đã đưa ra một thuật toán đơn giản tìm mã tiền tố tối ưu cho một phân phối nguồn nhất định. Thuật toán này được sử dụng rộng rãi đến ngày nay vì tính đơn giản của nó.

Ta nhắc lại bài toán tìm mã tối ưu: Cho nguồn $\mathcal{X}$ với phân phối xác suất $p(x)$. Tìm phép mã hóa tiền tố nguồn $\mathcal{X}$ trên bảng chữ cái D sao cho độ dài kỳ vọng của mã hóa $L = \sum l_i p_i$ là nhỏ nhất? Để giải quyết bài toán này, thuật toán Huffman dựa trên ý tưởng cơ bản sau:

- Dựa trên nguyên tắc phần tử nào có xác suất xuất hiện thấp được mã hóa bởi các từ mã dài.
- Trong mỗi bước cộng D xác suất nhỏ nhất rồi sắp xếp lại.
- Đánh số theo cũng một thức tự trong bảng chữ cái D.

[1] David Albert Huffman (9 tháng 8 năm 1925 - 7 tháng 10 năm 1999) là người tiên phong trong khoa học máy tính, được biết đến với mã hóa Huffman đồng thời ông cũng là một trong những người tiên phong trong lĩnh vực toán học origami.

Huffman lấy bằng cử nhân về kỹ thuật điện từ Đại học bang Ohio năm 1944, sau đó phục vụ hai năm làm sĩ quan Hải quân Hoa Kỳ . Ông trở lại Ohio State để lấy bằng thạc sĩ về kỹ thuật điện vào năm 1949. Năm 1953, ông lấy bằng tiến sỹ khoa học về kỹ thuật điện tại Học viện Công nghệ Massachusetts (MIT), với luận án tổng hợp mạch tuần hoàn , được tư vấn bởi Samuel H.

Codeword Length	Codeword	X	Probability
2	01	1	0.25 0.3 0.45 0.55 1
2	10	2	0.25 0.25 0.3 0.45
2	11	3	0.2 0.25 0.25
3	000	4	0.15 0.2
3	001	5	0.15

Hình 3.3 Ví dụ về Mã hóa Huffman

Để rõ hơn về nguyên tắc hoạt động của thuật toán này, ta xét các ví dụ sau:

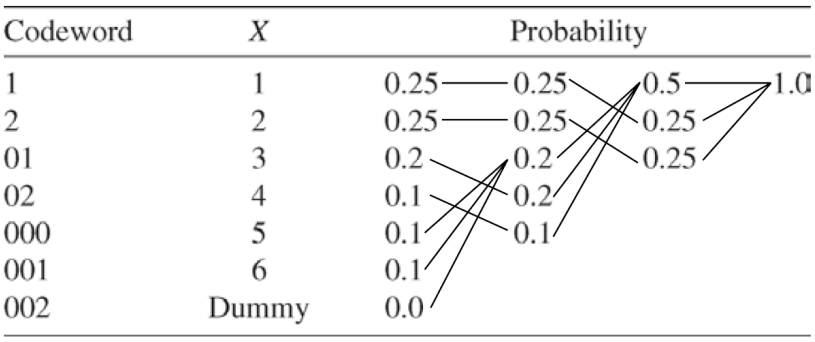
Ví dụ. Cho nguồn $\mathcal{X} = \{1, 2, 3, 4, 5\}$ với phân phối $p_X = \{0.25; 0.25; 0.2; 0.15; 0.15\}$ Thuật toán Huffman hoạt động theo bảng sau: Ta thấy rằng dãy trên đã được sắp xếp theo thứ tự giảm dần xác suất xuất hiện. Hai ký tự nguồn có xác suất nhỏ nhất là 4 và 5, ta cộng xác suất của chúng lại và sắp xếp lại dãy, ta đánh số bit 0 tương ứng với phần tử trên là 4, bit 1 tương ứng với phần từ dưới là 5. Sau đó sắp xếp lại dãy theo xác suất xuất hiện. Làm tương tự như vậy đến khi tổng xác suất bằng 1. Đọc đường đi từ trái sang phải tương ứng ta được mã hóa Huffman tương ứng với nguồn là:

$$C(X) = \{01; 10; 11; 000; 001\}$$

Độ dài trung bình của mã hóa là:

$$L = \sum l_i p_i = 2.3 \; bit$$

Ví dụ. Cho nguồn $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ với phân phối $p_X = \{0.25; 0.25; 0.2; 0.1; 0.1;$ Hay tìm mã Huffman cho bảng chữ cái tam phân $D = \{0, 1, 2\}$. Trong trường hợp này ta thấy rằng nếu làm tương tự như ví dụ trước đến bước



Hình 3.4 Ví dụ về Mã hóa Huffman trong trường hợp khởi tạo phần tử dummy

cuối cùng sẽ còn hai xác suất. Như vậy để đảm bảo mỗi bước có ba xác suất được cộng lại ta thêm một phần tử giả (gọi là dummy) có xác suất bằng 0 từ ban đầu. Chi tiết mã hóa được mô tả trong hình 3.4. Độ dài trung bình của mã hóa là:

$$L = \sum l_i p_i = 1.7 \; bit$$

Cũng theo ví dụ này, khi $D \geq 3$, chúng ta có thể không đủ số ký tự để kết hợp D các xác suất sau một số bước. Trong trường hợp tổng quát ta cần cộng thêm các ký tự giả có xác suất bằng 0 để thuật toán có thể hoạt động được. Vì một bước số lượng các xác suất giảm đi $D-1$, nên để mỗi bước hoạt động bình thường, số lượng ký tự nguồn phải bằng $1 + k(D-1)$, trong đó k là số lần trộn. Dựa vào đây có thể tính được k và số ký tự giả phải thêm vào.

2. Tính tối ưu của mã Huffman

Để chứng minh tính tối ưu của mã Huffman, trước tiên chúng ta chứng minh một số thuộc tính của một mã tối ưu cụ thể.

Không mất tính tổng quát, ta giả sử rằng hàm phân phối xác suất

được được sắp xếp trước, tức là $p_1 \geq p_2 \geq \ldots \geq p_m$. Nhắc lại rằng một mã tối ưu nếu $\sum p_i l_i$ là nhỏ nhất.

Bổ đề III.1 *Với bất kỳ phân phối nào, tồn tại một mã tiền tố tối ưu thỏa mãn các thuộc tính sau:*

1. *Độ dài của từ mã tỷ lệ nghịch với xác suất xuất hiện của chữ cái nguồn tương ứng (tức là nếu $p_j > p_i$, thì $l_j < l_i$).*
2. *Hai từ mã dài nhất có cùng độ dài.*
3. *Hai từ mã dài nhất chỉ khác nhau ở bit cuối cùng.*

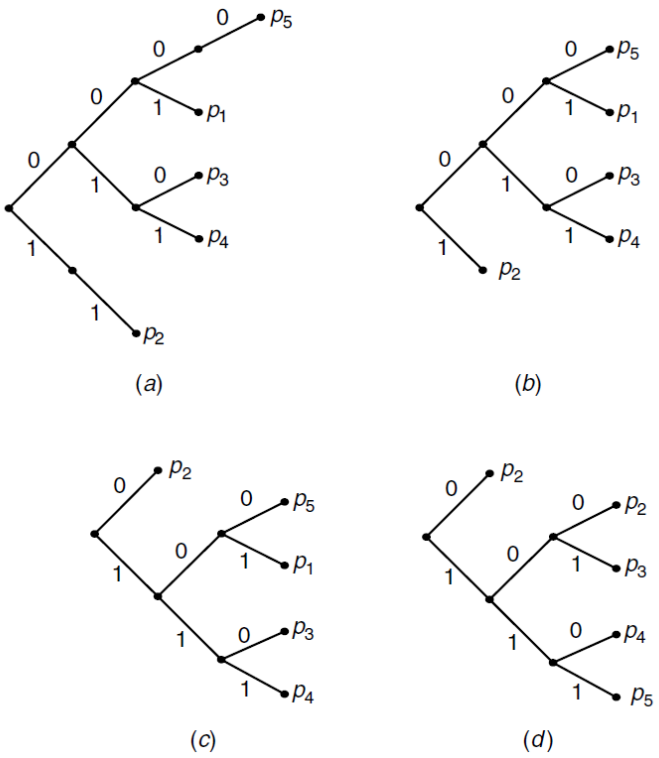
Chứng minh. Đầu tiên, chúng định nghĩa các phép toán: chuyển mã, cắt và sắp xếp lại thứ tự (hình 3.5). Trong hình này, phần (a) là một mã tiền tố cho trước. Bằng cách cắt các nhánh không có anh em, được cây mã như hình (b). Chúng ta tiếp tục sắp xếp lại cây như hình (c), trong đó độ dài sắp xếp theo thứ tự tăng dần từ đỉnh xuống dưới đáy. Cuối cùng, chúng ta đổi chỗ các xác suất để cải tiến độ sâu kỳ vọng của cây, được minh họa ở hình (d). Mọi mã tối ưu có thể được sắp xếp lại và hoán đổi thành dạng chuẩn như trong (d), ở đây $l_1 \leq l_2 \leq \ldots \leq l_{m-1} = l_m$ và 2 từ mã cuối chỉ khác nhau ở bit cuối cùng. Ta xét mã tối ưu C_m. Sau đây ta sẽ lần lượt chứng minh ba tính chất trong bổ đề.

- *Nếu $p_j > p_k$, thì $l_j \leq l_k$*: Xét mã hóa C'_m với từ mã thứ j và thứ k của C_m được đổi chỗ. Ta có:

$$L(C'_m) - L(C_m) = \sum p_i l'_i - \sum p_i l_i \tag{3.40}$$

$$= p_j l_k + p_k l_j - p_j l_j - p_k l_k \tag{3.41}$$

$$= (p_j - p_k)(l_k - l_j) \tag{3.42}$$



Hình 3.5 Các phép toán

Vì $p_j - p_k > 0$, và C_m là mã tối ưu nên $L(C'_m) - L(C_m) \geq 0$. Do đó, ta phải có $l_k \geq l_j$.

- *Hai từ mã dài nhất có cùng độ dài.* Ta sử dụng phép cắt các từ mã. Nếu hai từ mã dài nhất không có cùng độ dài, ta có thể cắt bit cuối cùng của từ mã dài hơn trong khi vẫn đảm bảo tính chất tiền tố và đạt được độ dài kỳ vọng ngắn hơn. Do đó, hai từ mã dài nhất phải có cùng độ dài. Theo tính chất 1, từ mã dài nhất phải thuộc về ký hiệu nguồn ít khả năng nhất.

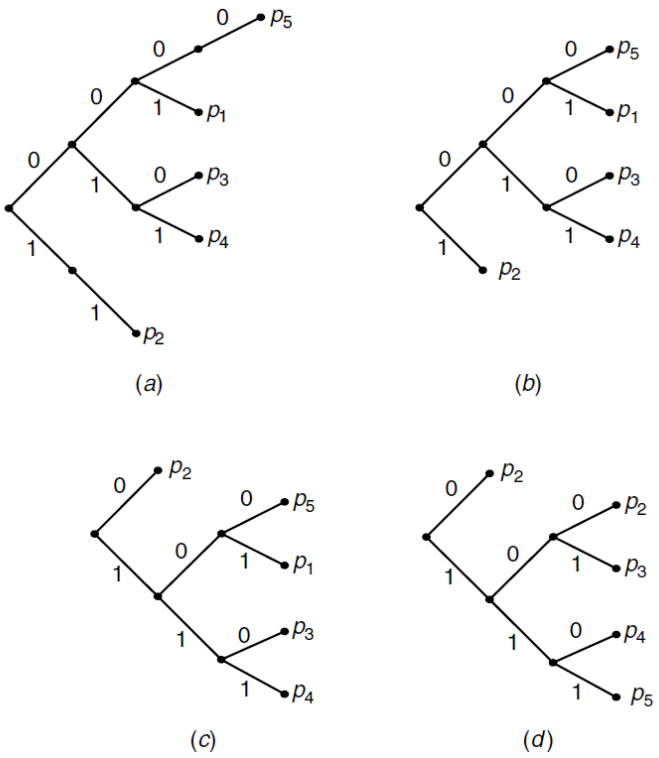
- *Hai từ mã dài nhất chỉ khác nhau ở bit cuối cùng và tương ứng với hai ký hiệu ít khả năng nhất.* Không phải tất cả các mã tối ưu thỏa mãn tính chất này, nhưng bằng phép toán sắp xếp lại, chúng ta có thể tìm ra mã tối ưu thỏa mãn. Nếu như có một từ mã với độ dài lớn nhất mà không có anh em (tức là từ mã chỉ khác ở bit cuối),

chúng ta có thể xóa đi bit cuối cùng của từ mã đó và vẫn thỏa mãn được tính chất tiền tố. Điều này làm giảm độ dài trung bình của từ mã và mâu thuẫn với sự tối ưu của mã. Do vậy, mỗi từ mã có độ dài lớn nhất trong bất cứ mã tối ưu nào đều phải có anh em. Bây giờ, chúng ta có thể hoán đổi các từ mã dài nhất sao cho hai ký hiệu nguồn xác suất thấp nhất được gắn với hai anh em trên cây. Cách làm này không thay đổi độ dài kỳ vọng $\sum p_i l_i$. Do đó, các từ mã cho hai ký hiệu nguồn xác suất thấp nhất sẽ có độ dài lớn nhất và giống nhau toàn bộ ngoại trừ bit cuối.

Tóm lại, chúng ta vừa chỉ ra rằng nếu $p_1 \geq p_2 \geq \ldots \geq p_m$, thì tồn tại một mã tối ưu với $l_1 \leq l_2 \leq \ldots \leq l_{m-1} = l_m$, và từ mã $C(x_{m^\smile 1})$ và $C(x_m)$ chỉ khác nhau duy nhất ở bit cuối cùng.

Theo cách trên, chúng ta vừa chỉ ra rằng tồn tại một mã tối ưu thoả mãn các tính chất ở trong bổ đề III.1. Chúng ta gọi những mã như thế là ***mã chuẩn***.

Với bất kỳ hàm phân phối xác suất cho bảng chữ cái kích cỡ là m, $p = (p_1, p_2, \ldots, p_m)$ với $p_1 \geq p_2 \geq \ldots \geq p_m$, chúng ta định nghĩa phép rút gọn Huffman $p' = (p_1, p_2, \ldots, p_{m-2}, p_{m-1} + p_m)$ trên bảng chữ cái với kích thước $m - 1$ (hình 3.6). Trong hình này gọi $p_1 \geq p_2 \geq \ldots \geq p_5$. Một mã tối ưu tiêu chuẩn được minh họa trong (a). Kết hợp 2 xác suất thấp nhất, chúng ta thu được mã trong (b). Sắp xếp lại các giá trị xác suất theo thứ tự giảm dần, chúng ta được mã tiêu chuẩn trong (c) cho $m-1$ ký tự. Gọi $C^*_{m-1}(p')$ là mã tối ưu cho p', và coi $C^*_m(p)$ là mã tối ưu chuẩn cho p. Việc chứng minh tính chất tối ưu sẽ dẫn theo hai cách xây dựng: Đầu tiên, ta mở rộng mã tối ưu cho p' để xây dựng mã cho p, sau đó chúng ta làm



Hình 3.6 Các phép toán

giảm (condense) mã tối ưu cho p để tạo thành một mã cho rút gọn mã Huffman p'. Sự so sánh độ dài từ trung bình cho 2 mã tạo nên việc một mã tối ưu cho p có thể được tạo ra bằng cách mở rộng mã tối ưu cho p'.

Từ mã tối ưu cho p', chúng ta hình thành mã mở rộng cho m thành phần như sau: Lấy từ mã trong C_m^* tương ứng với trọng số $p_{m-1} + p_m$ mà mở rộng nó bằng cách thêm 0 vào công thức (form) hình thành từ mã cho ký hiệu $m-1$ và thêm 1 vào công thức hình thành từ mã cho ký hiệu m. Cách thức hình thành được minh họa như sau:

	$C^*_{m-1}(p')$		$C_m(p)$	
p_1	w'_1	l'_1	$w_1 = w'_1$	$l_1 = l'_1$
p_2	w'_2	l'_2	$w_2 = w'_2$	$l_2 = l'_2$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
p_{m-2}	w'_{m-2}	l'_{m-2}	$w_{m-2} = w'_{m-2}$	$l_{m-2} = l'_{m-2}$
$p_{m-1} + p_m$	w'_{m-1}	l'_{m-1}	$w_{m-1} = w'_{m-1}0$	$l_{m-1} = l'_{m-1} + 1$
			$w_{m-1} = w'_{m-1}1$	$l_m = l'_{m-1} + 1$

Tính độ dài trung bình $\sum_i p'_i l'_i$, ta có:

$$L(p) = L^*(p') + p_{m-1} + p_m \tag{3.43}$$

Một cách tương tự, từ mã chuẩn cho p, chúng ta tạo ra mã cho p' bằng cách kết hợp các từ mã cho 2 ký hiệu có xác suất thấp nhất $m-1$ và m với xác suất tương ứng là p_{m-1} và p_m, 2 mã này là đoạn anh em bởi các tính chất của mã tiêu chuẩn. Mã mới cho p' có độ dài trung bình là:

$$L(p') = \sum_{i=1}^{m-2} p_i l_i + p_{m-1}(l_{m-1} - 1) + p_m(l_m - 1) \tag{3.44}$$

$$= \sum_{i=1}^{m} p_i l_i - p_{m-1} + p_m \tag{3.45}$$

$$= L^*(p) - p_{m-1} - p_m \tag{3.46}$$

Cộng $L(p)$ với $L(p')$, ta có:

$$L(p') + L(p) = L^*(p') + L^*(p) \tag{3.47}$$

Hoặc:

$$(L(p) + L^*(p')) + (L(p) - L^*(p)) = 0 \tag{3.48}$$

Bây giờ, xem xét hai số hạng trong đẳng thức trên. Bởi vì $L^*(p')$ là độ dài mã tối ưu cho p', chúng ta có $L(p')^\smile L^*(p') \geq 0$. Tương tự, độ dài của mở rộng mã tối ưu cho p' có độ dài trung bình tối thiểu như độ lớn của mã tối ưu cho p (ví dụ: $L(p)^\smile L^*(p) \geq 0$). Nhưng tổng của hai điều kiện không phủ định có thể duy nhất là 0, nếu cả hai đều là 0, điều này chỉ ra rằng $L(p) = L^*(p)$ (ví dụ, phần mở rộng mã tối ưu cho p là tối ưu cho p). Bởi thế, nếu chúng ta bắt đầu với một mã tối ưu cho p' với $m-1$ ký hiệu và tạo ra mã cho m ký hiệu bằng cách mở rộng từ mã tương ứng với $p_{m-1} + p_m$, mã mới cũng sẽ là tối ưu. Bắt đầu với mã cho 2 thành phần, trong đó trường hợp mã tối ưu là rất rõ ràng, chúng ta có thể mở rộng quy nạp kết quả này để chứng minh định lý sau.

Định lý III.1 (Tính tối ưu của mã Huffman) *Nếu C^* là một mã Huffman và C' là một mã tách được bất kỳ, ta có $L(C^*) \leq L(C')$.*

IV Mã Shannon - Fano - Elias

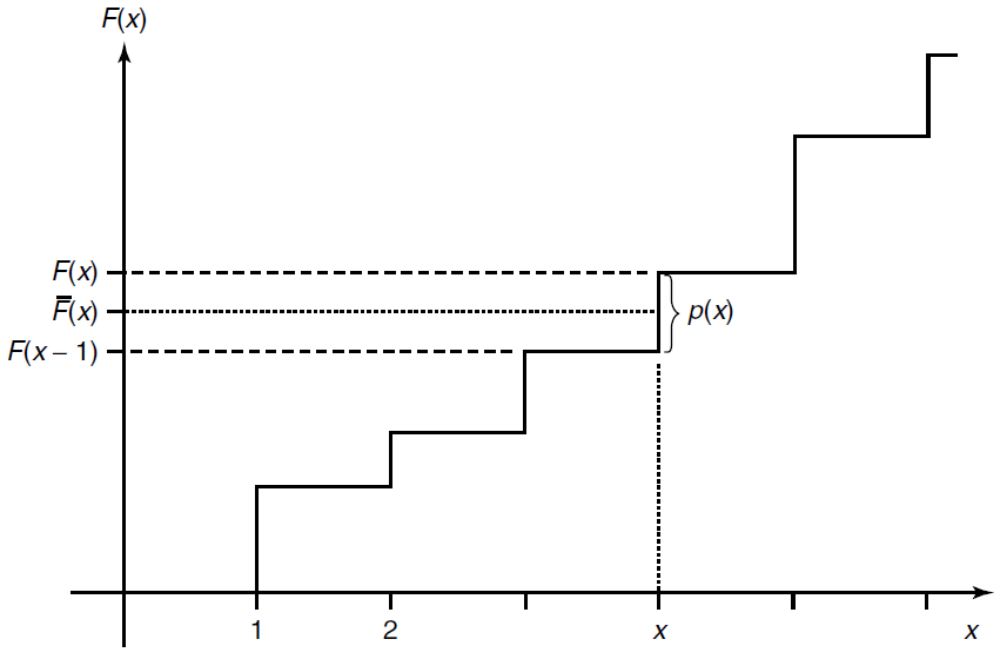
Trong phần trước chúng ta đã chỉ ra rằng từ mã có độ dài $l(x) = \lceil \log \frac{1}{p(x)} \rceil$ thỏa mã bất đẳng thức Kraft. Chúng ta có thể sử dụng điều này để xây dựng mã tách được. Trong phần này chúng ta mô tả một thủ tục xây dựng đơn giản sử dụng hàm phân phối tích lũy để phân chia các từ mã.

Không mất tính tổng quát, chúng ta xét $\mathcal{X} = \{1, 2, \ldots, m\}$. Giả sử $p(x) > 0$ với mọi x. Định nghĩa hàm phân phối tích lũy $F(x)$ như sau:

$$F(x) = \sum_{a \leq x} p(a) \tag{3.49}$$

Hàm này được mô tả trong hình 3.7. Lại xét hàm phân phối như sau:

$$\overline{F}(x) = \sum_{a < x} p(a) + \frac{1}{2} p(x) \tag{3.50}$$



Hình 3.7 Hàm phân phối xác suất cho mã hóa Shannon-Fanno-Elias

Trong đó hàm $\overline{F}(x)$ là tổng các xác suất của tất cả các ký tự trước x. Vì đây là biến ngẫu nhiên rời rạc, hàm phân phối chứa các bước (bậc) có kích thước là xác suất tại các điểm. Giá trị của $\overline{F}(x)$ là điểm giữa của bước tương ứng với x.

Vì các xác suất đều dương, $\overline{F}(a) \neq \overline{F}(b)$ nếu $a \neq b$ nên ta xác định được x nếu biết $\overline{F}(x)$. Ta chỉ cần nhìn biểu đồ của hàm phân phối để biết được x tương ứng. Do đó, giá trị $\overline{F}(x)$ có thể được sử dụng để mã hóa cho x. Tuy vậy, $\overline{F}(x)$ có thể là số thực nên việc biểu diễn bằng các bit sẽ khó khăn. Do đó, chúng ta có thể xấp xỉ chúng để có cách biểu diễn hiệu quả hơn. Giả sử ta cắt bớt từ $\overline{F}(x)$ đến $l(x)$ bits (ký hiệu là $\lfloor \overline{F}(x) \rfloor_{l(x)}$). Do vậy, chúng ta sử dụng $l(x)$ bit đầu của $\overline{F}(x)$ để mã hóa cho x. Bằng định nghĩa của việc làm tròn, ta có

$$\overline{F}(x) - \lfloor \overline{F}(x) \rfloor_{l(x)} < \frac{1}{2^{l(x)}} \tag{3.51}$$

Nếu $l(x) = \left\lceil \log \frac{1}{p(x)} \right\rceil + 1$ thì,

$$\frac{1}{2^{l(x)}} < \frac{p(x)}{2} = \overline{F}(x) - F(x-1) \tag{3.52}$$

Do đó đường biểu diễn $\lfloor \overline{F}(x) \rfloor_{l(x)}$, nằm trong bước tương ứng với x trên biểu đồ biểu diễn $\overline{F}(x)$. Do đó, $l(x)$ bit đủ để biểu diễn x.

Để được một mã tiền tố hợp lệ, ta kiểm việc biểu diễn có vi phạm mã tiền tố hay không? Xét từ mã bất kỳ có dạng $z_1 z_2 \dots z_l$ không chỉ một điểm thuộc khoảng $\left[0.z_1 z_2 \dots z_l, z_1 z_2 \dots z_l + \frac{1}{2^l}\right]$. Phép mã hóa này là tiền tố nếu và chỉ nếu khoảng tương ứng với hai từ mã có sự thoa lẫn giao nhau.

Khoảng tương ứng với mỗi từ mã có độ dài là $2^{-l(x)} < \frac{p(x)}{2}$. Chặn dưới của khoảng nhỏ hơn một nửa của bước, do vậy chặn trên của khoảng sẽ nằm dưới đỉnh của bước đó. Và do đó, khoảng tương ứng với mỗi bước sẽ nằm trong vùng của mỗi bước nên không có khoảng biểu diễn tương ứng nào giao thoa với khoảng khác.

Vì ta sử dụng $l(x) = \left\lceil \log \frac{1}{p(x)} \right\rceil + 1$ để biểu diễn x, kỳ vọng độ dài của từ mã là:

$$L = \sum_x p(x) l(x) = \sum_x p(x) \left(\left\lceil \log \frac{1}{p(x)} \right\rceil + 1 \right) < H(X) + 2 \tag{3.53}$$

Do vậy, việc mã hóa này tạo ra các ký tự không vượt quá entropy là 2 bit. Ta xét các ví dụ sau đây để rõ hơn về các mã hóa này.

Ví dụ. Xét tập nguồn $\mathcal{X} = \{1, 2, 3, 4\}$ thỏa mãn phân phối D-adic được cho dưới bảng sau:

x	$p(x)$	$F(x)$	$\overline{F}(x)$	$\overline{F}(x)$ nhị phân	$l(x)$	Từ mã
1	0.25	0.25	0.125	0.001	3	001
2	0.5	0.75	0.5	0.10	2	10
3	0.125	0.875	0.8125	0.1101	4	1101
4	0.125	1.0	0.9375	0.1111	4	1111

Trong trường hợp này, độ dài trung bình của từ mã là 2.75 và entropy là 1.75. Mã Huffman trong trường hợp này đạt đến Entrop, nhưng đối với phép mã hóa này ta có thể thấy rằng nó không hiệu quả như mã Huffman.

Ví dụ. Chúng ta xét một ví dụ khác của mã hóa Shannon- Fano- Elias. Trong trường hợp này, hàm phân phối không phải là D-adic. Vì phân phối không phải là D-adic nên biểu diễn của $F(x)$ có thể không phải là hữu hạn. Ta định nghĩa $0.010101010\ldots$ bởi $0.\overline{01}$. Ta xây dựng mã hóa như bảng sau:

x	$p(x)$	$F(x)$	$\overline{F}(x)$	$\overline{F}(x)$ dạng nhị phân	$l(x)$	Từ mã
1	0.25	0.25	0.125	0.001	3	001
2	0.25	0.5	0.375	0.011	3	011
3	0.2	0.7	0.6	$0.1\overline{0011}$	4	1001
4	0.15	0.85	0.775	$0.110\overline{0011}$	4	1100
5	0.15	1.0	0.925	$0.111\overline{0110}$	4	1110

Độ dài trung bình của mã hóa là 1.2 bit lớn hơn độ dài của mã Huffman.

Từ các ví dụ trên ta thấy rằng mã hóa Shannon - Fanno - Elias có thể được áp dụng cho các trình tự của các biến ngẫu nhiên. Ý tưởng chính là sử dụng hàm phân phối tích lũy của dãy, biểu diễn lại với độ chính xác thích hợp. Hiệu quả của mã hóa Shannon được đánh giá qua định lý sau:

Định lý IV.1 *Gọi $l(x)$ là chiều dài của từ mã với mã Shannon, $l'(x)$ là chiều dài của mã tách được với bất kỳ. Ta có:*

$$\Pr\left(l(X) \geq l'(X) + c\right) \leq \frac{1}{2^{c-1}} \tag{3.54}$$

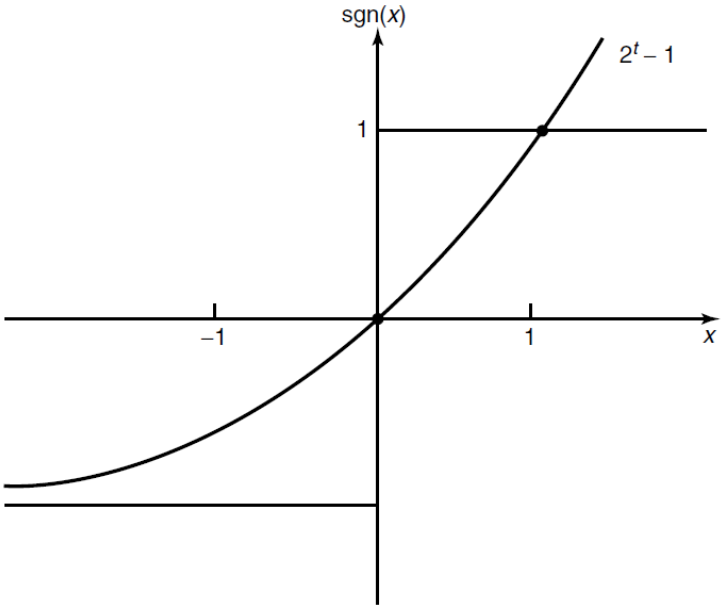
Chứng minh. Ta có

$$\begin{aligned}
\Pr\left(l(x) \geq l'(X) + c\right) \leq \frac{1}{2^{c-1}} &= \Pr\left(\left\lceil \log \frac{1}{p(x)} \right\rceil \geq l'(X) + c\right) && (3.55)\\
&\leq \Pr\left(\left\lceil \log \frac{1}{p(x)} \right\rceil \geq l'(X) + c - 1\right) && (3.56)\\
&\leq \Pr\left(p(x) \leq 2^{-l'(X)-c+1}\right) && (3.57)\\
&\leq \sum_{x:p(x)\leq 2^{-l'(x)-c+1}} p(x) && (3.58)\\
&\leq \sum_{x:p(x)\leq 2^{-l'(x)-c+1}} 2^{-l'(x)-(c-1)} && (3.59)\\
&\leq \sum_{x} 2^{-l'(x)} 2^{-(c-1)} && (3.60)\\
&\leq 2^{-(c-1)} && (3.61)
\end{aligned}$$

$\sum 2^{-l'(x)} \leq 1$ do bất đẳng thức Kraft.

Từ định lý này có thể thấy được tính chất tiệm cận tối ưu của mã Shannon.Tuy vậy, theo định lý trên ta có $l(x) \geq l'(x)+1$ với xác suất $\geq \frac{1}{2}$, điều này không đảm bảo tính tối ưu của mã Shanon. Sau đây chúng ta phân tích một kết quả mạnh hơn đối với mã này dựa trên lý thuyết trò chơi. Mục đích của phương pháp này nhằm chứng minh $l(x) < l'(x)$ thường xuyên hơn $l(x) > l(x)$.

Định lý IV.2 *Với phân phối dyadic $p(x)$, đặt $l(x) = \log \frac{1}{p(x)}$ là chiều dài của mã hóa nhị phân Shannon đối với nguồn, $l'(x)$ là chiều dài của mã*



Hình 3.8 Hàm $sgn(t)$

tách được bất kỳ đối với nguồn. Ta có:

$$\Pr\left[l(X) < l'(X)\right] \geq \Pr\left[l(X) < l'(X)\right] \tag{3.62}$$

Đẳng thức xảy ra nếu $l(x) = l'(x)$ với mọi x. Do vậy $l(x)$ là chiều dài tối ưu của mã hóa.

Chứng minh. Ta định nghĩa hàm $sgn(t)$ như sau:

$$sgn(t) = \begin{cases} 1, & \text{nếu } t > 0 \\ 0 & \text{nếu } t = 0 \\ -1 & \text{nếu } t < 0 \end{cases} \tag{3.63}$$

Hàm này có thể được biểu diễn ở hình 3.8. Từ đây ta thấy rằng $sgn(t) \leq 2^t - 1$ với $t = 0, \pm 1, \pm 2, ...,$. Chú ý rằng bất đẳng thức này thỏa mãn với

mọi t. Ta có:

$$\Pr\Big[l(X) < l'(X)\Big] \geq \Pr\Big[l(X) < l'(X)\Big] = \sum_{x:l'(x)<l(x)} p(x) - \sum_{x:l'(x)<l(x)} p(x) \tag{3.64}$$

$$= \sum_{x} sgn(l(x) - l'(x) \tag{3.65}$$

$$= Esgn(l(x) - l'(x) \tag{3.66}$$

$$\leq \sum_{x} p(x)\Big(2^{l(x)-l'(x)} - 1\Big) \tag{3.67}$$

$$= \sum_{x} 2^{-l(x)}\Big(2^{l(x)-l'(x)} - 1\Big) \tag{3.68}$$

$$= \sum_{x} 2^{-l'(x)} \sum_{x} 2^{-l(x)} \tag{3.69}$$

$$= \sum_{x} 2^{-l'(x)} - 1 \tag{3.70}$$

$$\leq 1 - 1 \tag{3.71}$$

$$= 0 \tag{3.72}$$

Trong đó 3.67 do chặn của hàm $sgn(x)$, và 3.71 là do $l'(x)$ thỏa mãn bất đẳng thức Kraft.

Dẳng thức xảy ra nếu (3.67) và (3.71) cùng xảy ra và hàm $sgn(t)$ nằm giữa khoảng $[0, 1]$ tức là $l(x) = l'(x)$ hoặc $l(x) = l'(x) + 1$. Bắt đẳng thức 3.71 xảy ra khi $l'(x)$ thỏa mãn bất đẳng thức Kraft. Kết hợp lại ta có $l(x) = l'(x)$ với mọi x.

V Mã hóa số học

Mã hóa Huffman giới thiệu trong phần trước là mã tối ưu đối với tập nguồn khi đã biết phân phối xác suất. Tuy vậy trong thực tế độ dài của từ mã trong mã Huffman bị giới hạn do là các số nguyên điều này có thể làm mất tới 1 bit để lưu trữ. Chúng ta có thể giảm bớt sự mất mát này bằng cách sử dụng các khối ký tự đầu vào - tuy nhiên, sự phức tạp của cách tiếp cận này tăng theo cấp số nhân với chiều dài khối. Bây giờ chúng ta mô tả một phương pháp mã hóa khắc phục được nhược điểm này. Trong mã số học số học, thay vì sử dụng một chuỗi các bit để biểu diễn một biểu tượng, chúng ta biểu diễn nó bằng một khoảng con của khoảng đơn vị.

Mã cho một dãy các ký hiệu là một khoảng số thựccó chiều dài giảm khi chúng ta thêm nhiều ký hiệu vào chuỗi. Thuộc tính này cho phép chúng ta có một lược đồ mã theo hệ thức truy hồi (mã cho một phần mở rộng cho một chuỗi có thể được tính đơn giản chỉ từ mã cho chuỗi ban đầu) và mà chiều dài mã vạch không bị giới hạn là một khoảng. Cải tiến này mở rộng phương pháp mã hóa Shano- Fano -Elias trong mục trước.

Xét một dãy biến ngẫu nhiên $X_1, X_2, \ldots$ trong một khoảng không gian hữu hạn $\mathcal{X} = \{1, 2, 3, \ldots, m\}$. Với mỗi dãy, ta đặt 0. trước các dãy và xem các dãy như một số thực (cơ số là $m + 1$) giữa khoảng $[0, 1]$. Gọi X là dãy $X = 0.X_1X_2 \ldots$. X có phân phối là:

$$F_X(x) = \Pr\{X \leq x = 0.x_1x_2 \ldots\} \tag{3.73}$$

$$= \Pr\{0.X_1X_2 \ldots \leq x = 0.x_1x_2 \ldots\} \tag{3.74}$$

$$= \Pr\{X_1 < x_1\} + \Pr\{X_1 = x_1, X_2 < x_2\} + \cdots . \tag{3.75}$$

Xét một dãy X_1, X_2, X_n mà có dùng phâp phối Bernoulli trên tập không

gian mẫu nhị phân $\mathcal{X} = \{0, 1\}$ tức là $p(X = 1) = p, p(X = 0) = q$. Trong trường hợp này, với dãy cụ thể $x^n = 110101$ sẽ ứng với giá trị của hàm F là

$$F(x^n) = \Pr(X_1 < 1) + \Pr(X_1 = 1, X_2 < 1) \tag{3.76}$$
$$+ \Pr(X_1 = 1, X_2 = 1, X_3 < 0) \tag{3.77}$$
$$+ \Pr(X_1 = 1, X_2 = 1, X_3 = 0, X_4 < 1) \tag{3.78}$$
$$+ \Pr(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 < 0) \tag{3.79}$$
$$+ \Pr(X_1 = 1, X_2 = 1, X_3 = 0, X_4 = 1, X_5 = 0, X_6 < 1) \tag{3.80}$$
$$= q + pq + p^2 \cdot 0 + p^2q \cdot q + p^2q \cdot q + p^2qp \cdot 0 + p^2qpqq \tag{3.81}$$
$$= q + pq + p^2q^2 + p^3 + q^3 \tag{3.82}$$

Chú ý rằng $\Pr(X_i < 0) = 0$. Đây là cách chuyển một chuỗi sang một giá trị xác suất tương ứng. Bây giờ giả sử chúng ta muốn biểu diễn dãy nhị phân $X_1, X_2, \ldots, X_n$ có độ dài n và gọi $x_1, x_2, \ldots, x_n$ là kết quả của quá trình chuyển đổi xác suất như trên, thì dãy nhị phân trên có thể được biểu diễn trong nửa khoảng sau:

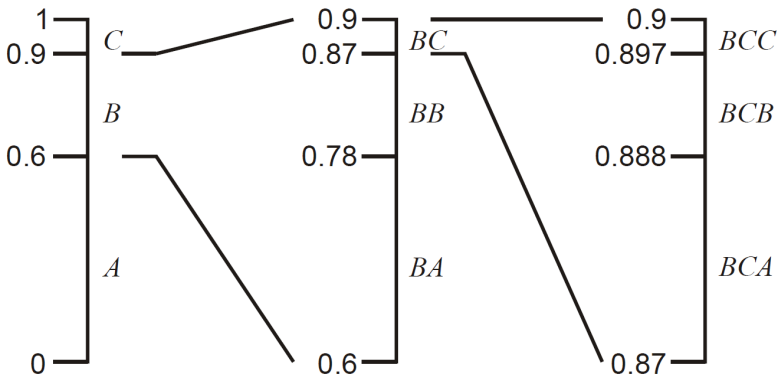
$$[0.x_1x_2 \ldots x_n000 \ldots, x_1x_2 \ldots x_n111 \ldots)$$

Sử dụng phép chuyển xác suất, khoảng này tương ứng với khoảng xác suất sau:

$$[F_Y(0.x_1x_2 \ldots x_n000 \ldots), F_Y(x_1x_2 \ldots x_n111 \ldots))$$

Khoảng này có độ dài bằng tổng tất cả xác suất của các dãy hữu han bắt đầu từ $0.x_1, x_2, \ldots, x_n$.

Ta xét một ví dụ với việc mã hóa nguồn $X = \{A, B, C\}$ với xác suất nguồn như sau: $P(A) = 0.6, p(B) = 0.3, p(C) = 0.1$. Trong ví dụ này ta



Hình 3.9 Biểu diễn khoảng xác suất trong mã hóa số học với nguồn $X = \{A, B, C\}$

thấy mỗi ký tự nguồn A, B, C sẽ được biểu diễn trong các nửa khoảng tương ứng là: $[0, 0.6)$, $[0.6, 0.9)$, và $[0.9, 1)$ Giả sử cần mã hóa dãy BCA, trước hết ta cần xét nửa khoảng biểu diễn cho B là $[0.6, 0.9)$. Trên khoảng này, ta chia tiếp tỷ lệ $0.6 : 0.3 : 0.1$ tương ứng với A, B, C được các khoảng nhỏ là $[0.6, 0.78), [0.78, 0.87), [0.87, 0.9)$. Tương ứng với các khoảng nhỏ này là BA, BB, BC. Ký tự thứ hai là C, nên ta chọn nửa khoảng nhỏ tiếp theo là $[0.87, 0.9)$. Lặp lại quá trình trên nửa khoảng này ta được dãy BCA được biểu diễn bởi nửa khoảng $[0.87, 0.888)$. Biểu diễn lại nhị phân cho nửa khoảng này ta được:

$$0.87 = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^4} + \frac{1}{2^5} + \ldots = 0.11011\cdots \tag{3.83}$$

$$0.888 = \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \frac{1}{2^7} + \ldots = 0.11011\cdots \tag{3.84}$$

Người mã hóa có thể gửi bất kỳ ký tự nào trong khoảng $[0.87, 0.888)$ để chỉ dãy được mã hóa. Ví dụ có thể gửi dãy bit 111 tương ứng với 0.875. Khi bên giải mã nhận được dãy này sẽ hoạt xây dựng các khoảng giống bên mã hóa thu được khoảng $[0.87, 0.888)$ mã hóa cho dãy BCA.

Tuy nhiên 0.875 cũng có thể biểu diễn cho B hoặc BC, do vậy bên mã hóa cần gửi thêm bit để chỉ ra sự kết thúc của chuỗi.

Cũng giống như trong các mã hóa trước, mã hóa số học đòi hỏi bên gửi cần gửi đi bảng phân phối xác suất và từ mã.

VI Mã hóa Lempel Ziv

Trong các phần trước chúng ta đã tìm hiểu về các thuật toán mã hóa dựa trên thống kê, có nghĩa là biết trước phân phối xác suất của nguồn. Tuy vậy trong thực tế các phương pháp này gặp phải một số các nhược điểm như sau:

- Cần biết phân phối xác suất của nguồn. Trong thực tế không phải lúc nào ta cũng biết được phân phối xác suất của nguồn, đặc biệt các nguồn có tính chất động các phân phối có thể dễ thay đổi.
- Khi xác suất của một phần tử thay đổi, cần tính lại phân phối xác suất của cả nguồn.

Trong phần này chúng ta tìm hiểu một phương pháp mã hóa dữ liệu mới mà không cần đến bảng phân phối xác suất, gọi là Lempel -Ziv. Ý tưởng chính của phương pháp này là xây dựng một bộ từ điển trong quá trình mã hóa. Sau đó, thay thế một chuỗi bằng thứ tự từ điển tới các chuỗi con lần xuất hiện trước của nó.

Để có thể mô tả rõ phương pháp này, ta xét ví dụ sau: giả sử ta cần mã hóa chuỗi 1011010100010010001001010010. Từ điển D ban đầu có K dữ liệu là $D(0), D(1), \ldots, D(K-1)$. Cần phải gửi đi $M = \lceil \log K \rceil$ bit để xác định vị trí của dữ liệu. Ban đầu $K = 1, D = \lambda$ và $M = \lceil \log K \rceil = 0$ bit. Các bước mã hóa được mô tả trong bảng 3.2.

- Ban đầu thêm λ vào từ điển, xét từ đầu dãy gặp bit 1, ta thấy $1 \notin D$, gửi đi 1 và đặt $D(1) = 1$.

Từ điển			Mã hóa	Nguồn	K, M
TT	TT nhị phân	Giá trị	(gửi đi)	(giải mã)	
0	0	λ	1	1	$1, 0$
1	01	1	**0**0	0	$2, 1$
2	10	0	**01**1	11	$3, 2$
3	11	11	**10**1	01	$4, 2$
4	100	01	**100**0	010	$5, 3$
5	101	010	**010**0	00	$6, 3$
6	110	00	**001**0	10	$7, 3$
7	111	10	**101**0	0100	$8, 3$
8	1000	0100	**1000**1	01001	$9, 4$
9	1001	01001	**1001**0	010010	$10, 4$

Bảng 3.2 Bảng mô tả mã hóa Lempel - Ziv cho chuỗi 1011010100010010001001010010

- Lúc này từ điển đã được thêm 1, nên ta tính được $K = 2, M = 1$. Tiếp tục xét $0 \notin D$, tách thành $\lambda + 0$ lúc này cần gửi đi M bit của vị trí trong từ điển và bit tiếp theo nên gửi đi 00. Cập nhật từ điển $D(2) = 0$.
- Tiếp theo tính được $K = 3, M = 2$, tiếp tục đọc 1, thấy $1 \in D$ nên đọc bit tiếp theo.
- Giá trị K, M giữ nguyên $K = 3, M = 2$. Đọc 11, thấy $11 \notin D$, tách thành $1 + 1$ gửi đi $M = 2$ bit vị trí của 1 (bit đầu) là 01 và bit tiếp theo là 1. Cập nhật từ điển $D(3) = 11$
- Đọc tiếp $0 \in D$, nên chưa gửi và đọc bit tiếp theo

- Đọc được $01 \notin D$ tách thành $0 + 1$. Tính được $K = 4.M = 2$ nên gửi 2 bit vị trí của 0 là 10 và bit tiếp theo là 1. Cập nhật từ điển Cập nhật từ điển $D(4) = 01$.
- Tiếp tục đọc 0, thấy $0 \in D$ nên chưa gửi, đọc bit tiếp theo.
- Tiếp tục đọc $01 \in D$, đọc bit tiếp theo
- Đọc $010 \notin D$, tách thành $01 + 1$. Tính được $K = 5, M = 3$, gửi đi 3 bit vị trí của 01 là 100 và bit tiếp theo là 1.
- Cứ tiếp tục như vậy ta được chuỗi mã hóa là: 1000111011000

Trong mã hóa này, ta cần chú ý đến số bit chỉ vị trí trong từ điển, công việc này được tính toán thông qua hai chỉ số M, K. Điều này cũng đảm bảo mã hóa là mã tiền tố.

Bộ giải mã một lần nữa bao gồm việc xây dựng từ điển các chuỗi con như là dữ liệu được giải mã. Khi giải mã xong cũng là lúc xây dựng bộ từ điển chuỗi con hệt như quá trình mã hóa.

Có nhiều biến thể của thuật toán Lampel-Ziv cơ bản, khai thác ý tưởng tương tự nhưng sử dụng các thủ tục khác để quản lý từ điển v.v. Các chương trình này nhanh hơn nhưng hiệu quả của chúng trong việc nén văn bản tiếng Anh, mặc dù hữu ích, không cao bằng mã hóa số học.

Ngược lại với các mã khối, mã Huffman, và phương pháp mã hóa số học, thuật toán Lempel-Ziv được định nghĩa không hề dùng đến bất kì mô hình xác suất nào của nguồn. Tuy nhiên với bất kỳ nguồn ergodic nào, thuật toán có thể được chứng minh là tiệm cận tới giá trị entropy của nguồn. Đây là là lý do tại sao nó được goi là thuật toán nén 'vạn năng'. Tuy nhiên, phương pháp Lempel-Ziv đạt được mục đích nén chỉ bằng cách ghi nhớ chuỗi con đã xuất hiện để có kí hiệu tắt trong lần tiếp

theo nó xuất hiện. Khoảng thời gian mà tiệm cận hiệu suất phổ quát này đạt được cho nhiều nguồn khác nhau khi khoảng thời gian dài cố định, bởi vì số lượng các chuỗi con điển hình cần ghi nhớ là rất lớn. Trong thực tế nhiều tập tin có chứa nhiều chuỗi ký tự ngắn lặp lại, một dạng dư thừa mà thuật toán này rất phù hợp.

Bài tập

1. Xét mã $\{0, 10, 01\}$.

   (a) Mã này có phải là mã tiền tố không?

   (b) Mã này có phải là mã không kì dị không?

   (c) Mã này có phải là mã có một cách giải không?

   (d) Liệu có tồn tại mã tiền tố cùng độ dài từ mã với mã này không? Nếu có hãy chỉ ra một mã như vậy.

2. *(Mã Huffman)*

   Cho biến ngẫu nhiên

   $$X = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}.$$

   (a) Tìm mã Huffman nhị phân cho X.

   (b) Tính độ dài kì vọng.

   (c) Tìm mã Huffman tam phân cho X.

3. *(Mã tồi)* Mã nào dưới đây không thể là mã Huffman với bất kì phân phối xác suất nào?

   $$\{0, 10, 01\}, \quad \{00, 01, 10, 110\}, \quad \{01, 10\}$$

4. *(Entropy tương đối là giá của sự chọn mã không phù hợp)* Giả sử X là một biến ngẫu nhiên có giá trị thuộc $\{1, 2, 3, 4, 5\}$, $p(x)$ và $q(x)$ là hai phân pối có thể có của X, được cho theo bảng sau

Symbol	$p(x)$	$q(x)$	$C_1(x)$	$C_2(x)$
1	$\frac{1}{2}$	$\frac{1}{2}$	0	0
2	$\frac{1}{4}$	$\frac{1}{8}$	10	100
3	$\frac{1}{8}$	$\frac{1}{8}$	110	101
4	$\frac{1}{16}$	$\frac{1}{8}$	1110	110
5	$\frac{1}{16}$	$\frac{1}{8}$	1111	111

   (a) Tính $H(p)$, $H(q)$, $D(p||q)$ và $D(q||p)$.

   (b) Kiểm tra rằng C_1 là tối ưu cho p và C_2 là tối ưu cho p.

   (c) Giả sử dùng C_2 cho p, khi đó độ dài từ mã trung bình là bao nhiêu? Nó lớn hơn entropy của p là bao nhiêu?

   (d) Tương tự câu (4c) nhưng dùng C_1 và q.

5. *Mã Huffman* Giả sử X là biến ngẫu nhiên nhận giá trị thuộc tập $\{A, B, C, D, E, F\}$ với xác suất là 0.5, 0.25, 0.1, 0.05, 0.05, và 0.05.

   (a) Xây dựng mã Huffman nhị phân cho biến này. Độ dài trung bình là bao nhiêu?

   (b) Xây dựng mã Huffman tứ phân cho biến này. Giả sử bảng chữ cái là $\{a, b, c, d\}$. Độ dài trung bình là bao nhiêu?

   (c) Một cách khác để xây dựng mã nhị phân là xây dựng mã tứ phân trước rồi chuyển sang mã nhị phần bằng ánh xạ $a \to 00$, $b \to 01$, $c \to 10$, $d \to 11$. Hãy chỉ ra mã nhị phân của biến ngẫu nhiên trên và độ dài trung bình của mã được xây dựng bằng cách này.

(d) Giả sử X là biến ngẫu nhiên bất kì, giả sử L_H là đô dài trung bình của mã Huffman nhị phân của biên ngẫu nhiên đó. Giả sử L_{QB} là độ dài trung bình của mã được xây dựng bằng cách xây dựng mã Huffman tứ phân sau đó chuyển đổi sang nhị phân. Chứng minh rằng

$$L_H \leq L_{QB} \leq L_H + 2.$$

CHƯƠNG 4
TỶ LỆ ENTROPY CỦA QUÁ TRÌNH NGẪU NHIÊN

Tính chất tiệm cận phân hoạch đều trong Chương 3 đã chỉ ra rằng $nH(X)$ bits là đủ về mặt trung bình để mô tả n biến ngẫu nhiên độc lập cùng phân bố. Nhưng nếu các biến này phụ thuộc thì sao? Đặc biệt, nếu các biến ngẫu nhiên có dạng là một quá trình dừng thì sao? Chúng ta sẽ chỉ ra trong trường hợp các biến ngẫu nhiên độc lập cùng phân bố, entropy $H(X_1, X_2, \ldots, X_n)$ tăng (tiệm cận) tuyến tính với n tại tỷ lệ $H(\mathcal{X})$, chúng ta sẽ gọi là *tỷ lệ entropy* của quá trình. Có thể hiểu $H(\mathcal{X})$ như là nén dữ liệu có thể đạt được tốt nhất sẽ được phân tích trong Chương 5.

I Xích Markov

Quá trình ngẫu nhiên $\{X_i\}$ là dãy chỉ số các biến ngẫu nhiên. Trong trường hợp tổng quát, có sự phụ thuộc lẫn nhau giữa các biến ngẫu nhiên. Quá trình này được đặc trưng bởi hàm mật độ xác suất đồng thời $\Pr\{(X_1, X_2, \ldots, X_n) = (x_1, x_2, \ldots, x_n)\} = p(x_1, x_2, \ldots, x_n), (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$, với $n = 1, 2, \ldots$

Định nghĩa I.1 *Quá trình ngẫu nhiên được gọi là* ***dừng*** *nếu phân bố đồng thời của một dãy các tập con bất kỳ của các biến ngẫu nhiên là bất biến qua phép dịch chuyển của chỉ số thời gian; cụ thể,*

$$\Pr\{X_1 = x_1, \ldots, X_n = x_n\} = \Pr\{X_{1+l} = x_1, \ldots, X_{n+l} = x_n\} \qquad (4.1)$$

với mọi n, với mọi phép dịch chuyển l và với mọi giá trị $x_1, \ldots x_n \in \mathcal{X}$.

Một ví dụ đơn giản của quá trình ngẫu nhiên có biến phụ thuộc thể hiện biến ngẫn nhiên chỉ phụ thuộc vào biến ngẫu hnhiên trước đó và độc lập *có điều kiện* với các biến ngẫu nhiên khác. Quá trình như vậy được gọi là Markov.

Định nghĩa I.2 *Quá trình ngẫu nhiên rời rạc* $X_1, X_2, \ldots$ *được gọi là* xích Markov *hauy* quá trình Markov *nếu với mọi* $n = 1, 2, \ldots$ *ta có*

$$\begin{aligned}\Pr(X_{n+1} = x_{n+1} \,|\, X_n = x_n, X_{n-1} = x_{n-1} \ldots X_1 = x_1) \\ = \Pr(X_{n+1} = x_{n+1} \,|\, X_n = x_n)\end{aligned} \tag{4.2}$$

với mọi $x_1, \ldots, x_n, x_{n+1} \in \mathcal{X}$.

Trong trường hợp này, hàm xác suất của các biến ngẫu nhiên sẽ được viết dưới dạng

$$p(x_1, \ldots, x_n) = p(x_1)p(x_2|x_1) \cdots p(x_n|x_{n-1}). \tag{4.3}$$

Định nghĩa I.3 *Quá trình ngẫu nhiên rời rạc* $X_1, X_2, \ldots$ *được gọi là* bất biến theo thời gian *nếu xác suất có điều kiện* $p(x_{n+1}|x_n)$ *không phụ thuộc vào* n. *Điều này có nghĩa là với* $n = 1, 2, \ldots,$

$$\Pr\{X_{n+1} = b|X_n = a\} = \Pr\{X_2 = b|X_1 = a\}, \quad \textit{với mọi } a, b \in \mathcal{X}. \tag{4.4}$$

Nếu không có gì bổ sung, ta sẽ luôn giả thiết rằng xích Markov là bất biến theo thời gian.

Nếu $\{X_i\}$ là một xích Markov, X_n được gọi là *trạng thái* ở thời điểm n. Một xích Markov bất biến với thời gian sẽ được đặc trưng bởi trạng thái ban đầu và một *ma trận chuyển trạng thái* $P = [P_{ij}]$, $i, j \in \{1, 2, \ldots, m\}$, trong đó $P_{ij} = \Pr\{X_{n+1} = j|X_n = i\}$.

Nếu có thể đạt được xác suất dương từ tất cả các trạng thái của xích Markov tới mọi trạng thái khác của xích Markov sau một số hữu hạn bước, thì xích Markov được gọi là *tối giản*. Nếu ước số chung lớn nhất của độ dài của các đường đi từ một trạng thái đến chính nó là bằng 1, thì xích Markov được gọi là *tựa tuần hoàn*.

Nếu ta gọi hàm mật độ xác suất của biến ngẫu nhiên tại thời điểm n là $p(x_n)$ thì hàm mật độ xác suất tại thời điểm $n+1$ sẽ là

$$p(x_{n+1}) = \sum_{x_n} p(x_n) P_{x_n x_{n+1}}. \tag{4.5}$$

Một phân bố của các trạng thái sao cho phân bố tại thời điểm $n+1$ bằng với phân bố của thời điểm n sẽ được gọi là *phân bố dừng*. Sở dĩ ta gọi là phân bố dừng vì nếu trạng thái ban đầu của xích Markov là một phân bố dừng, thì xích Markov sẽ tạo nên một quá trình dừng.

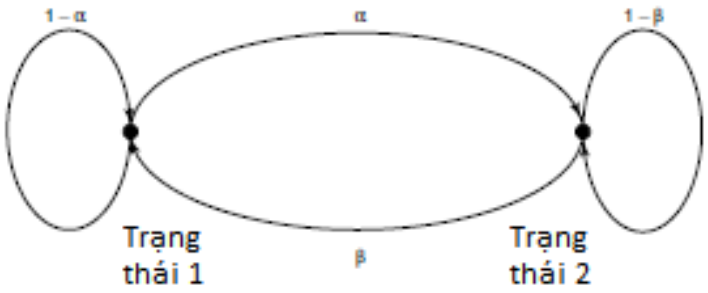
Nếu trạng thái hữu hạn của xích Markov là tối giản và tựa tuần hoàn, thì mật độ dừng là duy nhất, và từ mọi phân bố ban đầu, phân bố của X_n sẽ tiến tới một phân bố dừng khi $n \to \infty$.

Ví dụ. Xét ma trận xác suất chuyển của một xích Markov 2 trạng thái

$$P = \begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix} \tag{4.6}$$

như được mô tả trong Hình 4.1. Giả sử phân bố dừng được mô tả bởi vecto μ gồm các thành phần là các xác suất dừng của trạng thái 1 và 2 tương ứng. Khi đó ta có thể xác định được xác suất dừng từ việc giải phương trình $\mu P = \mu$, hoặc đơn giản hơn là thông qua các xác suất cân bằng. Đối với xác suất dừng, các dòng trạng thái được thể hiện ở đồ thị trạng thái chuyển đều bằng 0. Áp dụng vào Hình 4.1 ta được

$$\mu_1 \alpha = \mu_2 \beta. \tag{4.7}$$



Hình 4.1 Xích Markov hai trạng thái

Vì $\mu_1 + \mu_2 = 1$ nên phân bố dừng sẽ là

$$\mu_1 = \frac{\beta}{\alpha + \beta}, \quad \mu_2 = \frac{\alpha}{\alpha + \beta}. \tag{4.8}$$

Nếu xích Markov có trạng thái ban đầu phụ thuộc vào một trạng thái dừng, quá trình tiếp theo mà ta nhận được sẽ là một quá trình dừng. Entropy của trạng thái X_n tại thời điểm n sẽ là

$$H(X_N) = H\left(\frac{\beta}{\alpha + \beta}, \frac{\alpha}{\alpha + \beta}\right). \tag{4.9}$$

Tuy nhiên, đây không phải là tỷ lệ tăng của entropy đối với $H(X_1, X_2, \ldots, X_n)$. Tính độc lập của các trạng thái X_i vẫn còn là một trở ngại nhất định.

Ta có thể tóm tắt các kết quả trên lại như sau:

- *Tỷ lệ entropy:* Hai định nghĩa cho tỷ lệ entropy đối với quá trình ngẫu nhiên gồm

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, \ldots, X_n), \tag{4.10}$$

$$H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n | X_{n-1}, \ldots, X_1). \tag{4.11}$$

  Đối với quá trình ngẫu nhiên dừng, ta có

$$H(\mathcal{X}) = H'(\mathcal{X}). \tag{4.12}$$

- *Tỷ lệ entropy đối với xích Markov*

$$H(\mathcal{X}) = -\sum_{ij} \mu_i P_{ij} \log P_{ij}. \tag{4.13}$$

- *Nguyên lí thứ hai của nhiệt động lực học:* Với một xích Markov, ta có:

  1. Entropy tương đối $D(\mu_n||\mu'_n)$ là giảm theo thời gian.
  2. Entropy tương đối $D(\mu_n||\mu)$ giữa hàm phân bố và phân bố dừng là giảm theo thời gian.
  3. Entropy $H(X_n)$ tăng nếu hàm phân bố xác suất là chuẩn tắc.
  4. Entropy điều kiện $H(X_n|X_1)$ tăng theo thời gian đối với xích Markov dừng.
  5. Entropy điều kiện $H(X_0|X_n)$ với điều kiện ban đầu X_0 là hàm tăng với mọi xích Markov.

- *Hàm của xích Markov.* Nếu $X_1, \ldots, X_n$ tạo thành một xích Markov dừng và $Y_i = \phi(X_i)$, khi đó

$$H(Y_n|Y_{n-1}, \ldots, Y_1, X_1) \leq H(\mathcal{Y}) \leq H(Y_n|Y_{n-1}, \ldots, Y_1), \tag{4.14}$$

và

$$\lim_{n\to\infty} H(Y_n|Y_{n-1}, \ldots, Y_1, X_1) = H(\mathcal{Y}) = \lim_{n\to\infty} H(Y_n|Y_{n-1}, \ldots, Y_1). \tag{4.15}$$

II Tỷ lệ của Entropy

Giả sử ta có một dãy n biến ngẫu nhiên. Khi đó sẽ xuất hiện một câu hỏi: entropy của dãy tăng trưởng thế nào theo n? Ở mục này ta sẽ định nghĩa *tỷ lệ entropy* để trả lời câu hỏi trên.

Định nghĩa II.1 Entropy *của một quá trình ngẫu nhiên* $\{X_i\}$ *được xác định bởi*

$$H(\mathcal{X}) = \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \tag{4.16}$$

nếu giới hạn đó tồn tại.

Bây giờ ta xét một số ví dụ đơn giản cho tỷ lệ entropy của một số quá trình ngẫu nhiên.

1. *Máy đánh chữ.* Xét trường hợp một máy đánh chữ có thể cho ra m khả năng gõ các kí tự tương đương. Khi đó máy chữ sẽ cho ra m^n chuỗi có độ dài n kí tự. Khi đó $H(X_1, X_2, \ldots, X_n) = \log m^n$ và tỷ lệ entropy sẽ là $H(\mathcal{X}) = \log m$ bit trên mỗi kí tự.

2. $X_1, X_2, \ldots$ *là các biến ngẫu nhiên độc lập cùng phân phối (i.i.d.* Khi đó

$$H(\mathcal{X}) = \lim \frac{H(X_1, X_2, \ldots, X_n)}{n} = \lim \frac{nH(X_1)}{n} = H(X_1), \tag{4.17}$$

   đây chính là tỷ lệ entropy mà ta mong muốn của mỗi kí tự.

3. *Dãy các biến ngẫu nhiên độc lập nhưng không cùng phân phối.* Trong trường hợp này

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i) \tag{4.18}$$

   nhưng các giá trị $H(X_i)$ là không bằng nhau, vì vậy ta có thể chọn được một dãy các phân phối của $X_1, X_2, \ldots$ sao cho giới hạn $\frac{1}{n}\sum H(X_i)$ không tồn tại. Một ví dụ đơn giản là dãy nhị phân ngẫu nhiên, với $p_i = P(X_i = 1)$ không phải là hằng số mà phụ thuộc vào

i, được chọn sao cho giới hạn ở (4.16) không tồn tại. Ví dụ ta chọn

$$p_i = \begin{cases} 0,5 & \text{nếu } 2k < \log\log i < 2k+1, \\ 0 & \text{nếu } 2k+1 < \log\log i < 2k+2 \end{cases} \tag{4.19}$$

với $k = 1, 2, \ldots$. Khi đó ta có thể kéo dài tùy ý với $H(X_i) = 1$, tiếp theo một đoạn ứng với $H(X_i) = 0$ dài hơn đoạn trước theo hàm mũ. Do vậy, giá trị trung bình của $H(X_i)$ sẽ dao động xung quanh 0 và 1 và không có giới hạn. Vậy $H(\mathcal{X})$ không phải làm một quá trình.

Ta cũng có thể định nghĩa giá trị tương ứng với tỷ lệ entropy là đại lượng

$$H'(\mathcal{X}) = \lim_{n\to\infty} H(X_n|X_{n-1}, \ldots, X_1) \tag{4.20}$$

nếu giới hạn đó tồn tại.

Hai đại lượng $H(\mathcal{X})$ và $H'(\mathcal{X})$ tương ứng với hai khái niệm của tỷ lệ entropy. Đại lượng đầu tiên là entropy trên biểu tượng của n biến ngẫu nhiên, còn đại lượng thứ hai là entropy điều kiện của biến ngẫu nhiên cuối cùng trên cơ sở các biến ngẫu nhiên trước. Ta sẽ chứng minh một kết quả quan trọng dưới đây để cho thấy đối với các quá trình dừng, hai giới hạn nói trên là tồn tại và bằng nhau.

Định lý II.1 *Đối với một quá trình dừng, giới hạn* (4.16) *và* (4.20) *là tồn tại và bằng nhau:*

$$H(\mathcal{X}) = H'(\mathcal{X}). \tag{4.21}$$

Đầu tiên ta sẽ chứng minh rằng $H(X_n|X_{n-1}, \ldots, X_1)$ là tồn tại. Ta có định lí.

Định lý II.2 *Đối với một quá trình dừng, đại lượng* $H(X_n|X_{n-1},\ldots,X_1)$ *là không tăng theo* n *và có giới hạn* $H'(\mathcal{X})$.

Chứng minh.

$$H(X_n|X_{n-1},\ldots,X_1) \leq (X_{n+1}|X_n\ldots,X_2) \tag{4.22}$$

$$= H(X_n|X_{n-1},\ldots,X_1), \tag{4.23}$$

ở đây bất đẳng thức được suy từ tính giảm có điều kiện của entropy và đẳng thức sau đó được suy từ tính dừng của quá trình.Vì $H(X_n|X_{n-1},\ldots,X_1)$ là dãy giảm các số không âm nên nó có giới hạn, gọi là $H'(\mathcal{X})$.

Ta nhắc lại kết quả rất kinh điển sau đây của giải tích

Định lý II.3 (Trung bình Cesáro) *Nếu* $a_n \to a$ *và* $b_n = \frac{1}{n}\sum_{i=1}^{n} a_i$, *thì* $b_n \to a$.

Chứng minh. *(Khung chứng minh).* Vì hầu hết các thành phần $\{a_k\}$ đều gần với a nên b_n là giá trị trung bình của n phần tử sẽ cũng gần với a.

Chứng minh chuẩn tắc Xét $\epsilon > 0$. Vì $a_n \to a$ nên tồn tại số $N(\epsilon)$ sao cho $|a_n - a| \leq \epsilon$ với mọi $n \geq N(\epsilon)$. Do đó,

$$|b_n - a| = \left|\frac{1}{n}\sum_{i=1}^{n}(a_i - a)\right| \tag{4.24}$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}|a_i - a| \tag{4.25}$$

$$\leq \frac{1}{n}\sum_{i=1}^{N(\epsilon)}|a_i - a| + \frac{n - N(\epsilon)}{n}\epsilon \tag{4.26}$$

$$\leq \frac{1}{n}\sum_{i=1}^{N(\epsilon)}|a_i - a| + \epsilon \tag{4.27}$$

với mọi $n \geq N(\epsilon)$. Vì số hạng đầu tiên của vế phải tiến về 0 khi $n \to \infty$ nên ta sẽ có $|b_n - a| \leq 2\epsilon$ nếu cho n đủ lớn. Điều này chứng tỏ $b_n \to a$ khi $n \to \infty$. Điều phải chứng minh.

Chứng minh Định lí II.1 Sử dụng quy tắc bắc cầu, ta có

$$\frac{H(X_1, \ldots, X_n)}{n} = \frac{1}{n} \sum_{i=1}^{n} H(X_i | X_{i-1}, \ldots, X_1), \tag{4.28}$$

điều này có nghĩa là tỷ lệ entropy chính là trung bình theo thời gian của các entropy điều kiện. Tuy nhiên, ta đã biết rằng các entropy điều kiện có giới hạn là H'. Vì vậy, sử dụng Định lí II.3 ta suy ra trung bình đó có giới hạn và cũng bằng H'. Vậy, sử dụng Định lí II.2 ta được

$$\begin{aligned} H(\mathcal{X}) &= \lim \frac{H(X_1, \ldots, X_n)}{n} = \lim H(X_n | X_{n-1}, \ldots, X_1) \\ &= H'(\mathcal{X}). \end{aligned} \tag{4.29}$$

Từ đây ta suy ra điều phải chứng minh.

Đặc trưng rõ nét nhất của tỷ lệ entropy của một quá trình ngẫu nhiên được xác định từ tính chất tiệm cận bất đối xứng (AEP) của quá trình dừng ergodic. Ta sẽ chứng minh AEP tổng quát ở Mục 16.8, rằng với mọi quá trình dừng ergodic

$$-\frac{1}{n} \log p(X_1, X_2, \ldots, X_n) \to H(\mathcal{X}) \tag{4.30}$$

với xác suất 1. Từ việc sử dụngt kết quả này, ta có thể mở rộng các định lí ở Chương 3 ra trường hợp quá trình ergodic dừng tổng quát. Ta định nghĩa một tập hợp đặc trưng tương tự như ta làm đối với trường hợp i.i.d. ở Chương 3. Sử dụng các lập luận tương tự, ta chứng minh được rằng tập hợp đặc trưng có xác suất gần với 1 và có khoảng $2^{nH(\mathcal{X})}$ các dãy

đặc trưng có độ dài n, mỗi dãy có xác suất là khoảng $2^{-nH(\mathcal{X})}$. Từ đây ta biểu diễn được các dãy đặc trưng có độ dài n bằng khoảng $nH(\mathcal{X})$ bit. Điều này chỉ ra rằng đặc tính của tỷ lệ entropy được thể hiện bởi trung bình của độ dài biểu diễn của một quá trình dừng ergodic.

Ta đã định nghĩa được tỷ lệ entropy cho các quá trình dừng. Trường hợp xích Markov là một ví dụ cụ thể và dễ dàng tính toán được.

Xích Markov. Tỷ lệ entropy của một xích Markov dừng được định nghĩa bởi

$$\begin{aligned} H(\mathcal{X}) = H'(\mathcal{X}) &= \lim H(X_n|X_{n-1},\dots,X_1) = \lim H(X_n|X_{n-1}) \\ &= H(X_2|X_1), \end{aligned} \tag{4.31}$$

ở đây entropy điều kiện được tính toán bằng cách sử dụng các phân bố dừng. Nhắc lại rằng phân bố dừng μ chính là nghiệm của phương trình

$$\mu_j = \sum_i \mu_i P_{ij} \quad \text{với mọi } j. \tag{4.32}$$

Định lí sau đây cho biết biểu thức của entropy điều kiện.

Định lý II.4 *Cho $\{X_i\}$ là một xích Markov dừng với phân bố dừng μ và ma trận chuyển P. Giả sử $X_1 \sim \mu$. Khi đó tỷ lệ entropy [của xích Markov] sẽ là*

$$H(\mathcal{X}) = -\sum_{ij} \mu_i P_{ij} \log P_{ij}. \tag{4.33}$$

Chứng minh. Sử dụng kết quả

$$H(\mathcal{X}) = H(X_2|X_1) = \sum_i \mu_i \left(\sum_j -P_{ij} \log P_{ij} \right),$$

ta được điều phải chứng minh.

Xích Markov 2 trạng thái Tỷ lệ entropy của xích Markov 2 trạng thái ở Hình 4.1 là

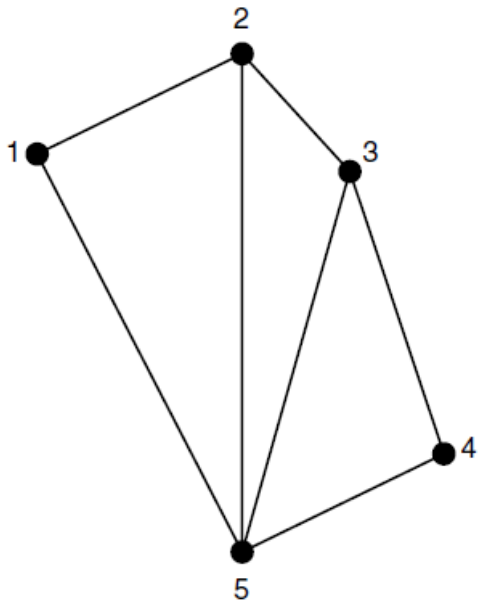
$$H(\mathcal{X}) = H(X_2|X_1) = \frac{\beta}{\alpha+\beta}H(\alpha) + \frac{\alpha}{\alpha+\beta}H(\beta). \tag{4.34}$$

Chú ý. Xích Markov tối giản và hầu tuần hoàn sẽ có một phân bố dừng duy nhất đối với biến trạng thái, và mọi phân bố ban đầu đều hội tụ về phân bố dừng khi $n \to \infty$. Trong trường hợp này, mặc dù phân bố ban đâu không phải là phân bố dừng thì tỷ lệ entropy vẫn được tính bằng $H(\mathcal{X})$ theo công thức (4.31) và (4.33).

III Ví dụ: Tỷ lệ entropy của bước nhảy ngẫu nhiên của một đồ thị có trọng số

Trong mục này ta xét một ví dụ về quá trình ngẫu nhiên là một bước nhảy ngẫu nhiên trên một đồ thị có trọng số (Hình 4.2). Xét đồ thị gồm m đỉnh được đánh nhãn $\{1, 2, \ldots, m\}$, với trọng số tương ứng $W_{ij} \geq 0$ được cho trên cạnh nối đỉnh i và đỉnh j. Ta giả thiết đồ thị không có hướng, tức là $W_{ij} = W_{ji}$. Ở đây ta đặt $W_{ij} = 0$ nếu không có đường nối các đỉnh i và j.

Ta xét chất điểm thực hiện bước nhảy ngẫu nhiên từ đỉnh này sang đỉnh khác của đồ thị. Bước nhảy ngẫu nhiên $\{X_n\}$, $X_n \in \{1, 2, \ldots, m\}$ là dãy các đỉnh của đồ thị. Nếu $X_n = i$ thì đỉnh tiếp theo j sẽ được chọn từ các đỉnh có kết nối với đỉnh i với xác suất tỉ lệ với trọng lượng của đường nối i và j. Khi đó ta có $P_{ij} = W_{ij} / \sum_k W_{ik}$.



Hình 4.2 Chu trình ngẫu nhiên trên một đồ thị

IV Định luật thứ hai của nhiệt động lực học

Luật thứ hai của nhiệt động lực học là một trong các luật cơ bản của vật lí, khẳng định rằng entropy của một hệ cô lập là không tăng. Bây giờ ta sẽ nghiên cứu mối liên hệ giữa định luật thứ hai này và hàm entropy mà ta đã xây dựng trong chương này.

Trong nhiệt động lực học thống kê, entropy thường được định nghĩa là logarit của một số các vi trạng thái của hệ thống. Điều này liên quan trực tiếp đến khái niệm entropy mà ta đã xét khi coi các trạng thái đó là như nhau. Tuy nhiên, câu hỏi đặt ra ở đây là vì sao entropy lại tăng?

Ta mô hình hóa hệ biệt lập này bởi một xích Markov với các phép chuyển tuân thủ các định luật vật lí áp dụng lên hệ. Giả thiết này phù hợp với các trạng thái của hệ, và một thực tế là các trạng thái tương lai của hệ không phụ thuộc vào quá khứ mà chỉ phụ thuộc vào trạng thái hiện tại. Trong các hệ như trên, ta tìm được bốn biểu hiện của định luật

thứ hai, và sẽ cho ta thấy một thực tế là không phải lúc nào entropy của hệ cũng tăng. Tuy nhiên, entropy *tương đối* thì luôn tăng.

1. *Entropy tương đối $D(\mu_n||\mu'_n)$ giảm theo n.* Cho μ_n và μ'_n là hai phân bố xác suất theo trạng thái không gian của xích Markov tai thời điểm n và μ_{n+1} và μ'_{n+1} là các phân bố tương ứng tại thời điểm $n+1$.Kí hiệu các hàm trọng số tương ứng là p và q. Khi đó $p(x_n, x_{n+1}) = p(x_n)r(x_{n+1}|x_n)$ và $q(x_n, x_{n+1}) = q(x_n)r(x_{n+1}|x_n)$, trong đó $r(\cdot|\cdot)$ là hàm chuyển xác suất đối với xích Markov. Khi đó sử dụng nguyên lí bắc cầu cho entropy tương đối, ta có hai biểu thức

$$\begin{aligned} D(p(x_n, x_{n+1})||q(x_n, x_{n+1})) &= D(p(x_n)||q(x_n)) \\ &\quad + D(p(x_{n+1}|x_n)||q(x_{n+1}|x_n)) \\ &= D(p(x_{n+1})||q(x_{n+1})) \\ &\quad + D(p(x_n|x_{n+1})||q(x_n|x_{n+1})). \end{aligned}$$

Vì p và q đều được lấy từ xích Markov, nên các hàm mật độ có điều kiện $p(x_n|x_{n+1})$ và $q(x_n|x_{n+1})$ đều bằng $r(x_{n+1}|x_n)$, và do đó $D(p(x_{n+1}|x_n)||q(x_{n+1}|x_n)) = 0$. Sử dụng tính không âm của $D(p(x_n|x_{n+1})||q(x_n|x_{n+1}))$ (Hệ quả của Định lí 2.6.3), ta có

$$D(p(x_n)||q(x_n)) \geq D(p(x_{n+1})||q(x_{n+1})) \tag{4.35}$$

hoặc

$$D(\mu_n||\mu'_n) \geq D(\mu_{n+1}||\mu'_{n+1}). \tag{4.36}$$

Từ hai biểu thức trên, ta suy ra rằng với mọi xích Markov, khoảng cách giữa các hàm mật độ có điều kiện sẽ giảm theo n.

Một ví dụ thực tế để áp dụng bất đẳng thức ở trên là ta có thể giả thiết hệ thống thuế thu nhập cá nhân của Canada và Anh là như nhau. Khi đó nếu μ_n và μ'_n là phân bố của thu nhập của người dân ở hai quốc gia trên, thì bất đẳng thức trên sẽ cho ta thấy entropy khoảng cách tương đối giữa hai phân bố nói trên sẽ giảm theo thời gian. Phân bố thu nhập của Canada và Anh sẽ ngày càng trở nên tương đồng.

2. *Entropy tương đối $D(\mu_n||\mu)$ giữa phân bố μ_n của trạng thái tại thời điểm n và phân bố dừng μ giảm theo n.* Trong (4.36), μ'_n là phân bố của trạng thái tại n. Nếu cho μ'_n giá trị của một phân bố dừng bất kì μ, thì phân bố μ'_{n+1} tại thời gian tiếp theo sẽ cũng bằng μ. Do đó

$$D(\mu_n||\mu) \geq D(\mu_{n+1}||\mu), \tag{4.37}$$

từ đây suy ra cùng với thời gian, mọi phân bố trạng thái sẽ ngày càng gần với mỗi phân bố dừng. Dãy $D(\mu_n||\mu)$ sẽ là dãy không âm, không tăng và đơn điệu, và do đó giới hạn. Giới hạn đó bằng 0 nếu phân bố dừng là duy nhất, nhưng khó có thể chứng minh được điều này một cách tường minh.

3. *Entropy tăng nếu phân bố dừng là chuẩn tắc.* Trong trường hợp tổng quát, việc entropy tương đối giảm không kéo theo khẳng định entropy tăng. Một phản ví dụ ở đây tương ứng với xích Markov có phân bố dừng không chuẩn tắc. Nếu ta xuất phát xích Markov này từ một phân bố chuẩn tắc, tức là một phân bố với entropy cực đại, thì phân bố đó sẽ tiến dần đến một phân bố dừng có entropy thấp hơn dạng chuẩn tắc. Ở đây entropy sẽ giảm theo thời gian.

Tuy nhiên, nếu phân bố dừng là một phân bố chuẩn tắc thì ta có thể biểu diễn entropy tương đối như sau

$$D(\mu_n||\mu) = \log|\mathcal{X}| - H(\mu_n) = \log|\mathcal{X}| - H(X_n). \tag{4.38}$$

Trong trường hợp này tính đơn điệu giảm của entropy tương đối sẽ kéo theo tính đơn điệu tăng của entropy. Điều này giải thích hiện tượng có trong nhiệt động lực học thống kê, khi mọi vi trạng thái đều sẽ bằng nhau. Ta sẽ nêu đặc trưng cho các quá trình có phân bố dừng chuẩn tắc như sau.

Định nghĩa IV.1 *Một ma trận chuyển xác suất* $[P_{ij}]$, $P_{ij} = \Pr\{X_{n+1} = j|X_n = i\}$ *được gọi là* ngẫu nhiên kép *nếu*

$$\sum_i P_{ij} = 1, \quad j = 1, 2, \ldots \tag{4.39}$$

và

$$\sum_j P_{ij} = 1, \quad i = 1, 2, \ldots \tag{4.40}$$

Chú ý. Phân bố chuẩn tắc là một phân bố dừng theo P nếu và chỉ nếu ma trận chuyển xác suất là ngẫu nhiên kép (xem Bài tập 4.1).

4. *Entropy điều kiện* $H(X_n|X_1)$ *tăng theo* n *cho mọi quá trình Markov.* Nếu quá trình Markov là dừng thì $H(X_n)$ là hằng số. Điều này kéo theo entropy là không tăng. Tuy nhiên, ta sẽ chứng minh rằng $H(X_n|X_1)$ là tăng theo n. Do đó tính bất định điều kiện của tương lai sẽ được tăng lên. Ta sẽ chứng minh điều này theo hai cách. Đầu

tiên, ta sử dụng các tính chất của entropy

$$H(X_n|X_1) \geq H(X_n|X_1, X_2), \quad \text{(điều kiện giảm entropy)} labeleq : 4.50 \tag{4.41}$$

$$= H(X_n|X_2) \quad \text{(Sử dụng tính Markov)} \tag{4.42}$$

$$= H(X_{n-1}|X_1) \quad \text{(sử dụng tính dừng)}. \tag{4.43}$$

Mặt khác, sử dụng bất đẳng thức xử lí số liệu đối với xích Markov $X_1 \to X_{n-1} \to X_n$, ta có

$$I(X_1; X_{n-1}) \geq I(X_1; X_n). \tag{4.44}$$

Thác triển thông tin trên theo các thành phần của entropy, ta có

$$H(X_{n-1}) - H(X_{n-1}|X_1) \geq H(X_n) - H(X_n|X_1). \tag{4.45}$$

Sử dụng tính dừng ta có $H(X_{n-1}) = H(X_n)$, và do đó ta có

$$H(X_{n-1}|X_1) \leq H(X_n|X_1). \tag{4.46}$$

Chú ý rằng người ta cũng dùng các kĩ thuật ở trên cũng dùng để chứng tỏ rằng $H(X_0|X_n)$ tăng theo n với mọi xích Markov.

5. *Entropy hỗn hợp tăng.* Gọi T là một phép xáo trộn (hoán vị) các con bài trên bàn và X là vị trí ngẫu nhiên của các con bài trên bàn. Nếu sự lựa chọn của việc xáo trộn T không phụ thuộc vào X thì ta có

$$H(TX) \geq H(X), \tag{4.47}$$

trong đó TX là phép hoán đổi của các con bài trên bàn từ trạng thái ban đầu X.

V Hàm của xích Markov

Trong mục này chúng tôi xét một ví dụ tương đối phức tạp. Nó sẽ cho thấy khả năng của kĩ thuật mà ta đang áp dụng có thể tiến xa được thế nào. Cho $X_1, X_2, \ldots, X_n, \ldots$ là một xích Markov dừng, và đặt $Y_j = \phi(X_i)$ là một quá trình mà mỗi thành phần của nó là một hàm của xích Markov đang xét. Câu hỏi đặt ra ở đây là tìm tỷ lệ entropy $H(\mathcal{Y})$. Các hàm của xích Markov như trên có ứng dụng trong thực tiễn. Ở nhiều trường hợp, ta chỉ có thể có thông tin về các trạng thái của hệ. Nếu biết được giá trị $Y_1, Y_2, \ldots, Y_n$ thì câu chuyện trở nên dễ dàng hơn rất nhiều, nhưng không phải lúc nào ta cũng có thể có được điều đó. Thật vậy, nếu xích Markov là dừng, thì quá trình Markov $Y_1, Y_2, \ldots, Y_n$ cũng vậy, và từ đó ta tính được ngay tỷ lệ entropy. Tuy nhiên, nếu muốn tính được $H(\mathcal{Y})$ thì ta phải tính được $H(Y_n|Y_{n-1}, \ldots, Y_1)$ với mỗi n và sau đó đi xác định giới hạn. Vì sự hội tụ có thể lâu tùy ý, nên ta rất khó xác định được giới hạn của nó. Trên thực tế, khi n càng lớn thì việc quan sát chênh lệch giữa bước n và $n+1$ là thực sự phức tạp, vì khoảng cách giữa hai giá trị đó là rất nhỏ, ngay cả khi ta còn chưa đạt được đến giới hạn (ví dụ như chuỗi $\sum \frac{1}{n}$).

Trong tính toán thực tế, việc có được các giá trị giới hạn trên và giới hạn dưới là rất quan trọng. Nó giúp ta dừng các bước tính toán một cách hợp lí, khi mà hiệu của cận trên và cận dưới là đủ nhỏ để có thể ước lượng giới hạn một cách đủ tốt.

Ta đã biết rằng $H(Y_n|Y_{n-1}, \ldots, Y_1)$ hội tụ đơn điệu trên tới $H(\mathcal{Y})$. Đối với trường hợp cận dưới, ta sử dụng $H(Y_n|Y_{n-1}, \ldots, Y_1, X_1)$. Khẳng định này được suy ra từ thực tế là X_1 đã chứa nhiều thông tin của Y_n như các đại lượng $Y_1, Y_0, Y_{-1}, \ldots$.

Bổ đề V.1

$$H(Y_n|Y_{n-1},\ldots,Y_1,X_1) \le H(\mathcal{Y}). \tag{4.48}$$

Chứng minh. Ta có, với $k = 1, 2, \ldots$

$$\begin{aligned} H(Y_n|Y_{n-1},\ldots,Y_2,X_1) &\overset{(a)}{=} H(Y_n|Y_{n-1},\ldots,Y_12,Y_1,X_1) && (4.49)\\ &\overset{(b)}{=} H(Y_n|Y_{n-1},\ldots,Y_1,X_1,X_0,X_{-1},\ldots,X_{-k}) && (4.50)\\ &\overset{(c)}{=} H(Y_n|Y_{n-1},\ldots,Y_1,X_1,X_0,X_{-1},\ldots \\ &\qquad X_{-k},Y_0,\ldots,Y_{-k}) && (4.51)\\ &\overset{(d)}{\le} H(Y_n|Y_{n-1},\ldots,Y_1,Y_0,\ldots,Y_{-k}) && (4.52)\\ &\overset{(e)}{=} H(Y_{n+k+1}|Y_{n+k}\ldots,Y_1), && (4.53) \end{aligned}$$

ở đây (a) được suy ra từ chú ý rằng Y_1 là một hàm của X_1, và (b) được suy từ tính Markov của X. Đẳng thức (c) có được nhờ Y_i là các hàm của X_i, bất đẳng thức (d) là do các điều kiện giảm entropy, còn (e) là do tính dừng. Vì bất đẳng thức trên đúng với mọi k, nên giới hạn của nó cũng đúng. Từ đó suy ra

$$\begin{aligned} H(Y_n|Y_{n-1},\ldots,Y_2,X_1) &\le \lim_k H(Y_{n+k+1}|Y_{n+k}\ldots,Y_1), && (4.54)\\ &= H(\mathcal{Y}). && (4.55) \end{aligned}$$

Ta có điều phải chứng minh.

Hệ quả ngay dưới đây sẽ chỉ ra rằng khoảng cách giữa cận trên và cận dưới sẽ giảm theo độ dài.

Bổ đề V.2

$$H(Y_n|Y_{n-1},\ldots,Y_2,Y_1) - H(Y_n|Y_{n-1},\ldots,Y_1,X_1) \to 0. \tag{4.56}$$

Chứng minh. Độ dài của khoảng được viết dưới dạng

$$H(Y_n|Y_{n-1},\ldots,Y_2,Y_1) - H(Y_{n+k+1}|Y_{n+k}\ldots,Y_1,X_1) = I(X_1;Y_n|Y_{n-1},\ldots,Y1). \quad (4.57)$$

Từ tính chất của các thông tin thứ cấp, ta có

$$I(X_1;Y_1,Y_2,\ldots,Y_n) \leq H(X_1), \quad (4.58)$$

và $I(X_1;Y_1,Y_2,\ldots,Y_n)$ tăng theo n. Do đó, giới hạn $\lim I(X_1;Y_1,Y_2,\ldots,Y_n)$ là tồn tại và

$$\lim_{n\to\infty} I(X_1;Y_1,Y_2,\ldots,Y_n) \leq H(X_1). \quad (4.59)$$

Từ đó, sử dụng nguyên lí bắc cầu ta được

$$H(X) \geq \lim_{n\to\infty} I(X_1;Y_1,Y_2,\ldots,Y_n) \quad (4.60)$$

$$= \lim_{n\to\infty} \sum_{i=1}^{n} I(X_1;Y_i|Y_{i-1},\ldots,Y_1) \quad (4.61)$$

$$= \sum_{i=1}^{\infty} I(X_1;Y_i|Y_{i-1},\ldots,Y_1) \quad (4.62)$$

Vì chuỗi trên là hội tụ và các số hạng là không âm nên số hạng tổng quát phải tiến về không, có nghĩa là

$$\lim I(X_1;Y_n|Y_{n-1},\ldots,Y_1) = 0. \quad (4.63)$$

Ta có điều phải chứng minh.

Phối hợp hai Bổ đề V.1 và V.2 ta được định lí sau.

Định lý V.1 *Nếu $X_1, X_2, \ldots, X_n$ là một xích Markov dừng, và $Y_i = \phi(X_i)$ thì*

$$H(Y_n|Y_{n-1},\ldots,Y_1,X_1) = H(\mathcal{Y}) = \lim H(Y_n|Y_{n-1},\ldots,Y_1). \quad (4.64)$$

Trong trường hợp tổng quát, ta cũng có thể xét trường hợp Y_i là hàm ngẫu nhiên của X_i. Xét quá trình Markov $X,1\ldots,X_n$ và tương ứng quá trình mới $Y_1,\ldots,Y_n$ được xác định bởi đại lượng $p(y_i|x_i)$, độc lập có điều kiện với mọi X_j, $j \neq i$, tức là

$$p(x^n, y^n) = p(x_1)\prod_{i=1}^{n-1} p(x_{i+1}|x_i)\prod_{i=1}^{n} p(y_i|x_i). \tag{4.65}$$

Quá trình như trên được gọi là *mô hình Markov ẩn* (HMM - hidden Markov model), được sử dụng rộng rãi trong kĩ thuật nhận diện âm thanh, nhận diện chữ viết, vv. Bằng cách sử dụng các lập luận tương tự như ở trên cho mô hình Markov ẩn, ta có thể chặn trên tỷ lệ entropy của mô hình Markov ẩn bằng cách cho tương ứng với các trạng thái Markov tương ứng. Chi tiết của phần này xin dành cho độc giả.

Bài tập

1. *Đường đi ngẫu nhiên trong khối lập phương.* Một con chim bay từ phòng này sang phòng khác trong một khối lập phương kích thước $3 \times 3 \times 3$. Tỉ lệ entropy là bao nhiêu?

2. *Entropy của đồ thị* Xét đường đi ngẫu nhiên trên đồ thị (liên thông) 3 cạnh.

   (a) Đồ thị nào có tỉ lệ entropy nhỏ nhất? Giá trị đó là bao nhiêu?

   (b) Đồ thị nào có tỉ lệ entropy lớn nhất.

3. *Quá trình dừng.* Giả sử $\ldots, X_{-1}, X_0.X_1, \ldots$ là một quá trình ngẫu nhiên dừng (không nhất thiết là Markov). Phát biểu nào sau đây là đúng? Chứng minh hoặc đưa ra phản ví dụ.

(a) $H(X_n|X_0) = H(X_{-n}|X_0)$.

(b) $H(X_n|X_0) \geq H(X_{n-1}|X_0)$.

(c) $H(X_n|X_1^{n-1}, X_{n+1})$ không tăng theo n.

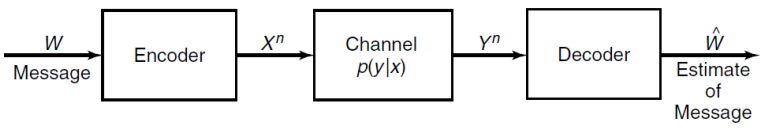
(d) $H(X_{n+1}|X_1^n, X_{n+2}^{2n+1})$ không tăng n.

CHƯƠNG 5
KÊNH TRUYỀN

Việc truyền thông tin này là một quá trình vật lý và chịu sự kiểm soát của nhiễu của môi trường xung quanh và sự không hoàn hảo của quá trình truyền tín hiệu vật lý. Giả sử bên gửi A muốn truyền tín hiệu qua kênh truyền tới bên nhận B, quá trình truyền thành công nếu B nhận đúng những gì A gửi. Trong chương này chúng ta sẽ xây dựng cách đánh giá hiệu quả kênh truyền giữa A và B qua lý thuyết về Entropy. Chương này cũng giới thiệu một số loại kênh cơ bản và các đặc tính của chúng.

I Tổng quan về kênh

Một hệ thống truyền thông đơn giản có thể bao gồm các yếu tố sau: đầu gửi tín hiệu, bộ phận mã hóa, kênh truyền, bộ phận giải mã và đầu nhận (hình 5.1). Mỗi bộ phận có một vai trò khác nhau và không thể thiếu trong các hệ thống truyền thông hiện đại ngày nay. Tuy vậy, bộ phận quan trọng nhất của các hệ thông này là *kênh truyền*, kênh truyền đóng vai trò xương sống quyết định toàn bộ chất lượng, hiệu quả của hệ thống. Trong phần này ta xét những kênh mang các tín hiệu rời rạc, chúng được sử dụng nhiều trong các hệ thống truyền thông hiện đại đang sử dụng các tín hiệu số. Đặc trưng của đầu vào kênh được mô tả bởi các phân



Hình 5.1 Hệ thống truyền thông

phối xác suất nó cũng quyết định phân phối của các tín hiệu đầu ra.

Định nghĩa I.1 (Kênh rời rạc) *Kênh rời rạc bao là một hệ thống bao gồm*

1. *Một bảng chữ cái đầu vào $\mathcal{X}$ và bảng chữ cái đầu ra $\mathcal{Y}$*
2. *Xác suất chuyển ma trận xác suất chuyển $P(Y|X) = \{p(y|x)\}$ là xác suất nhận được chữ cái $y \in \mathcal{Y}$ khi chữ cái đầu vào là $x \in \mathcal{X}$.*
3. *Kênh được gọi là không nhớ nếu phân phối xác suất đầu ra chỉ phụ thuộc vào phân phối xác suất đầu vào tại thời điểm đó và độc lập đó điều kiện với các đầu ra trước đó.*

Giả sử có phần phối đầu vào là $\mathbf{P}_X$, đầu ra là $\mathbf{P}_Y$, ma trận chuyển là $\mathbf{P}$, ta có:

$$\mathbf{P}_Y = \mathbf{P}_X.\mathbf{P}(Y|X) \tag{5.1}$$

Ví dụ. Cho ma trận chuyển như sau

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix} \tag{5.2}$$

Phân phối đầu vào $\mathbf{P}_X = (0.5, 0.25, 0.25)$. Phân phối của đầu ra là $\mathbf{P}_Y = (0.375, 0.3, 0.325)$.

II Dung lượng kênh

Để đánh giá hiệu quả của kênh, chúng ta đưa ra những độ đo để đánh giá lượng thông tin truyền qua kênh gồm: lượng tin truyền qua kênh và dung lượng kênh.

Định nghĩa II.1 *Cho kênh rời rạc $(X, P(Y|X), Y)$, lượng tin và dung lượng truyền qua kênh được định nghĩa như sau:*

- *Lượng tin truyền qua kênh là thông tin tương hỗ giữa đầu vào và đầu ra $I(X, Y)$.*
- *Dụng lượng kênh được định nghĩa bởi lượng tin cực đại theo phân phối đầu vào:*

$$C = \max_{P(X)} I(X, Y) \tag{5.3}$$

Theo định nghĩa trên ta thấy lượng tin quyền qua kênh đối với một phân phối đầu vào chính là thông tin chung hay thông tin giống nhau giữa bên phát và bên nhận.

Dung lượng kênh là giá trị thông tin chung cực đại khi phân phối X biến thiên. Theo đó, lượng tin truyền qua kênh có ý nghĩa tức thời. Còn dung lượng kênh là giá trị toàn cục lớn nhất. Theo đó, dung lượng kênh cũng là đại lượng bất biến với mỗi kênh (kể cả phân phối đầu vào có thay đổi), là đặc trưng truyền tải thông tin qua kênh. Dung lượng kênh có một số tính chất cơ bản như sau:

1. $C \geq 0$ vì $I(X, Y) \geq 0$
2. $C \leq \log |X|$ vì $C = \max I(X, Y) \leq H(X) = \log |X|$
3. $C \leq \log |X|$

III Một số loại kênh thường gặp

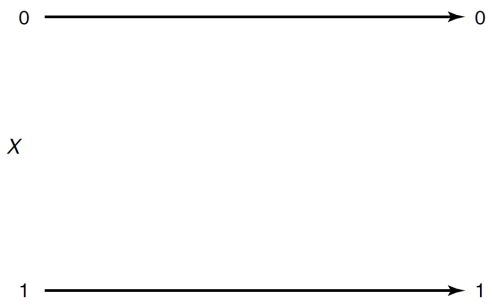
Trong phần này ta sẽ tìm hiểu một số loại kênh truyền thương gặp, từ đơn giản đến phức tạp. Các loại kênh cơ bản này đóng vai trò quan trọng trong việc nghiên cứu quá trình truyền tin nói chung.

1. Kênh nhị phân không nhiễu

Đây là loại kênh cơ bản và cũng là kênh lý tưởng nhất trong truyền tín hiệu nhị phân. Đối với kênh này cả hai đầu vào và đầu ra đều là bảng chữ cái nhị phân (hình 5.2), tức là $X = Y = \{0, 1\}$. Ma trân kênh lúc này là

$$\mathbf{P} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{5.4}$$

Trong trường hợp này, không có sự mất mát tín hiệu khi truyền qua



Hình 5.2 Kênh nhị phân không nhiễu

kênh. Lượng tin truyền qua kênh với phân phối đầu vào là $\mathbf{P}_X = \{p_1, p_2\}$ là:

$$I(X,Y) = H(Y) - H(Y|X) \tag{5.5}$$

$$= -p_1 \log p_1 - p_2 \log p_2 - p_1 H(Y|X=1) - p_2 H(Y|X=2) \tag{5.6}$$

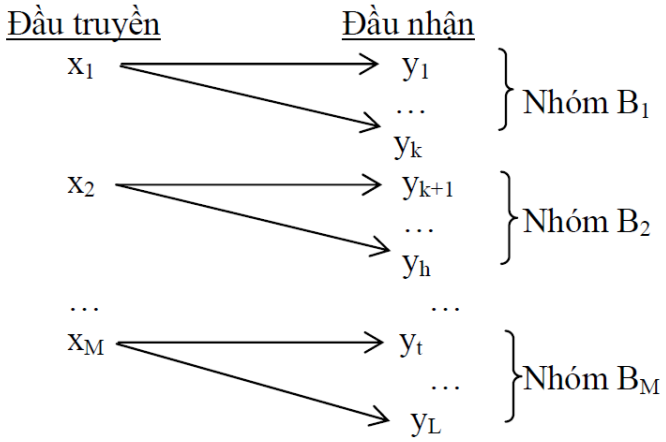
$$= 2(-p_1 \log p_1 - p_2 \log p_2) \tag{5.7}$$

Từ đây suy ra $C = \max_{\mathbf{P}_X} I(X,Y) = 2\ bit$ đạt được khi $p_1 = p_2 = \frac{1}{2}$.

2. Kênh truyền không mất thông tin

Trong kênh truyền không mất thông tin đầu vào và đầu ra gồm nhiều chữ cái trong đó số lượng chữ cái đầu vào nhỏ hơn đầu ra. Ký hiệu

bảng chữ cái đầu vào $X = \{x_1, x_2, \ldots, x_M\}$, bảng chữ cái đầu ra là $Y = \{y_1, y_2, \ldots, y_L\}$ $(M < L)$. Từ tập hợp các giá trị có thể nhận được ở đầu ra $Y = \{y_1, y_2, \ldots, y_L\}$, đầu ra sẽ được được phân thành M nhóm $B_i (i = 1..M)$ tương ứng với các giá trị x_i ở đầu truyền và xác suất để truyền x_i với điều kiện đã nhận y_j là $p(X = x_i | Y = y_j \in B_i) = 1$. Đặc



Hình 5.3 Kênh nhị phân không nhiễu

trưng của kênh truyền không mất tin là $H(X|Y) = 0$. Có nghĩa là lượng tin chưa biết về X khi nhận Y là bằng 0 hay ta có thể hiểu khi nhận được Y thì ta hoàn toàn có thể biết về X.

Định lý III.1 *Dung lượng kênh của kênh truyền không mất thông tin là* $C = \log M$

Chứng minh.

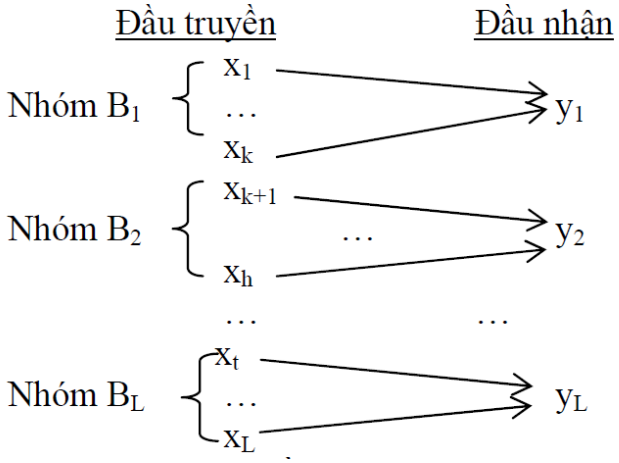
$$I(X,Y) = H(Y) - H(Y|X) = H(X) - H(X|Y) \leq \log |X| = \log M \tag{5.8}$$

3. Kênh truyền xác định

Ngược lại với truyền không mất thông tin, trong kênh truyền này bảng chữ cái đầu ra nhỏ hơn bảng chữ cái đầu vào, tức là $L < M$. Từ tập hợp

các giá trị có thể truyền ở đầu truyền được phân thành L nhóm B_j tương ứng với các giá trị có thể nhận được y_j ở đầu nhận và xác suất để nhận y_j với điều kiện đã truyền x_i là $p(Y = y_j|X = x_i \in B_j) = 1$. Kênh này có đặng trưng là $H(Y|X) = 0$. Có nghĩa là lượng tin chưa biết về Y khi truyền X bằng 0 hay khi truyền X thì ta biết sẽ nhận được Y.

Định lý III.2 *Dung lượng kênh của kênh truyền xác định là* $C = \log L$



Hình 5.4 Kênh nhị phân không nhiễu

4. Kênh đối xứng

Định nghĩa III.1 *Một kênh là đối xứng nếu các hàng của ma trận chuyển* $P(Y|X)$ *là hoán vị của nhau và các cột cũng là hoán vị của nhau.*

Ví dụ. Ví dụ kênh đối xứng với ma trận kênh như sau:

$$\mathbf{P} = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \\ 0.3 & 0.2 & 0.5 \end{pmatrix} \tag{5.9}$$

Trong trường hợp chỉ các hàng là hoán vị của nhau còn các cột có tổng bằng nhau thì kênh gọi là *đối xứng yếu*. Rõ ràng, kênh đối xứng là một

trường hợp đặc biệt của kênh đối xứng yếu. Định lý sau cho ta xác định giá trị dung lượng kênh của một kênh đối xứng yêu

Định lý III.3 *Kênh đối xứng yêu có dung lượng kênh cho bởi công thức:*

$$C = \log |Y| - \log r \tag{5.10}$$

Trong đó r là một hàng của ma trận chuyển. Dung lượng này đạt được khi phân phối trên bảng chữ cái vào là phân phối đều.

Chứng minh. Ta có:

$$I(X,Y) = H(Y) - H(Y|X) = H(Y) - \sum_{x \in X} p(x) H(Y|X = x) \tag{5.11}$$

Vì các hàng là hoán vị của nhau nên $H(Y|X = x)$ là như nhau với mọi giá trị $x \in X$. Vì các biến này có tổng xác suất bằng 1 nên

$$I(X,Y) = H(Y) - H(Y|X) = H(Y) - H(Y|X = x_1) \leq \log |X| - H(r) \tag{5.12}$$

Khi X có phân phối đều thì đăng thức xảy ra, do vậy ta có điều phải chứng minh.

IV Lược đồ giải mã cho kênh truyền

1. Xây dựng lược đồ giải mã

Với một kênh $(X, P(Y|X), Y)$ thì với kênh truyền không nhiễu y_i chính là x_i. Tuy nhiên, trong thực tế luôn có nhiễu xảy ra, do đó nếu nhận được y_i thi chưa chắc đã do x_i được gửi đi. Do đó ta cần tìm cách giải mã y_i về giá trị x_i tương ứng khi kênh truyền có nhiễu sao cho kỳ vọng lỗi là nhỏ nhất. Một phương pháp giải mã như vậy gọi là một *lược đồ giải mã.*

Quá trình xây dựng lược đồ giải mã đồng nghĩa với việc xây dựng một kênh truyền không mất, cũng đồng nghĩa với một phép phân hoạch giá trị nhận đầu nhận theo nhóm tương ứng với mỗi giá trị đầu vào.

- Giả sử giá trị gửi đi là $X = \{x_1, x_2, \ldots, x_M\}$
- Giá trị nhận là $Y = \{y_1, y_2, \ldots, y_L\}$.
- Ta cần xây dựng phép phân hoạch chia giá trị của tập Y thành các nhóm $B_i, (i = 1..M)$, sao cho:

$$B_i \cap B_j = \emptyset \tag{5.13}$$

$$\bigcup_{i=1}^{L} B_i = Y \tag{5.14}$$

  Khi nhận được $y_j \in B_i$ thì giải mã được x_i

Ví dụ. Ta xét một ví dụ đơn giản như sau về phép phân hoạch

- Cho $X = \{0000, 0101, 1110, 1011\}$
- $Y = \{0000, 0001, 0010, 0011, 0100, 0101, 0110, 0111,$
  $1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111\}$

Giả sử ta có thể phân hoạch tập Y thành các tập con B_i như sau:

- $B_1 = \{0000, 1000, 0001, 0010\}$
- $B_2 = \{0101, 1101, 0100, 0111\}$
- $B_3 = \{1110, 0110, 1111, 1100\}$
- $B_4 = \{1011, 0011, 1010, 1001\}$

Với phép phân hoạch này giả sử nhận $y_j = 0011$ thì giải mã về $x_4 = 1011$ vì $y_j \in B_4$.

2. Lược đồ giải mã tối ưu

Xét một kênh truyền trong đó nguồn phát tín hiệu (hay thông báo) với vận tốc R (tín hiệu/giây). Tín hiệu được mã hóa từ bộ ký tự mã. Sau đó tín hiệu mã hóa được truyền trên kênh với vận tốc C (ký tự/giây), C đồng thời là dung lượng của kênh truyền. Trong thực tế ta giả sử tín hiệu truyền trên kênh có thể bị nhiễu với xác suất $P(e)$.

Ví dụ. Giả sử kênh truyền từng bit với $C = 1$, nguồn phát thông báo với tốc độ $R = 2/5$ bit/giây ($R < C$). Để thuận lợi cho mã hóa và giảm nhiễu, ta xét từng khoảng thời gian $n = 5$ giây.

Như vậy trong khoảng thời gian $n = 5$ giây, ta có: Tập hợp các tín hiệu khác nhau là $2nR = 4$. Số bit được phát ra là $nR = 2$ bit và một tín hiệu dạng m_i được kết cấu bởi một dãy các bit.

Quá trình mã hóa các tín hiệu m_1, m_2, m_3, m_4 cần chú ý là: mỗi m_i cần được mã hóa với số bit tối đa là $nC = 5$ bit. Có thể mã hóa hai cách như sau.

Cách 1:

- $m_1 = 00000$
- $m_2 = 01101$
- $m_3 = 11010$
- $m_4 = 10111$

Cách 2:

- $m_1 = 00$
- $m_2 = 01$

- $m_3 = 10$
- $m_4 = 11$

Với cách 1, ta có nhiều khả năng phát hiện và sửa sai do nhiễu.

Ta đánh giá lỗi trong quá trình truyền tin thông qua các dạng sai số cơ bản như sau:

- Xác suất truyền sai từ mã x_i:

$$p(e/x_i) = \sum p(Y = y_j \notin B_i | X = x_i) \tag{5.15}$$

- Xác suất truyền sai trung bình:

$$p(e) = \sum_{i=1}^{M} p(X = x_i) p(e|x_i) \tag{5.16}$$

- Xác suất truyền sai lớn nhất:

$$p_m(e) = \max_{i=1,M} p(e|x_i) \tag{5.17}$$

Bài toán: Xét một quá trình truyền tin, trong đó $X = \{x_1, x_2, \ldots, x_M\}$, $Y = \{y_1, y_2, \ldots, y_M\}$. Mục tiêu của phép giải mã tối ưu tìm một lược đồ giải mã sao cho xác suất truyền sai một từ mã là nhỏ nhất.

Sau đây ta xây dựng cách tiếp cận tham lam sử dụng thuật toán Bayes để xây dựng phép giải mã.

- Theo công thức Bayes ta có:

$$\Pr(x_k|y_j) = p(x_k).p(y_j|x_k)/p(y_j)$$

- Do $p(y_j)$ không đổi nên ta chỉ cần so sánh các giá trị $p(x_k) = p(x_k)p(y_j|x_k)$

- Từ định nghĩa lược đồ giải mã tối ưu, ta cần tìm x_k sao cho $\Pr(x_k|y_j)$ cực đại, tức là các giá trị $p(x_k).p(y_j|x_k)$ cực đại.

Thuật toán xây dựng lược đồ giải mã

- Bước 1: Khởi tạo $B_i = \emptyset, \forall i$
- Bước 2: Lặp
  1. Với mọi $y_j \in Y$, ta tính $p(x_k).p(y_j|x_k), \forall j$
  2. Chọn x_i^* sao cho:

$$p(x_k^*).p(y_j|x_k^*) = \max_{x_k \in X} p(x_k).p(y_j|x_k) \tag{5.18}$$

  3. Cập nhật $B_i = B_i \cup \{y_j\}$ và $g(y_j) = x_i^*$.

Ví dụ. Cho một kênh đầu vào là $X = \{x_1, x_2, x_3\}$, đầu ra là $Y = \{y_1, y_2, y_3\}$, ma trận kênh như sau

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 1/6 & 1/2 \\ 1/6 & 1/2 & 1/3 \end{pmatrix} \tag{5.19}$$

Phân phối đầu vào: $p(x_1) = 1/2; p(x_2) = p(x_3) = 1/4$. Hãy xây dựng lược đồ giải mã tối ưu theo phương pháp Bayes.

Ta thực hiện theo các bước của thuật toán như sau:

- Bước 1: $B_1 = B_2 = B_3 = \emptyset$;
- Bước 2.1: Với y_1, ta tính được
  1. $p(x_1).p(y_1/x_1) = 1/2.1/2 = 1/4$ (Max)
  2. $p(x_2).p(y_1/x_2) = 1/4.1/3 = 1/12$

3. $p(x_3).p(y_1/x_3) = 1/4.1/6 = 1/24$

Do $p(x_1).p(y_1/x_1)$ lớn nhất nên liệt kê y_1 vào tập hợp B_1 tương ứng với x_1, do đo cập nhật $B_1 = \{y_1\}$.

- Bước 2.2: Với y_1, ta tính được

1. $p(x_1).p(y_2/x_1) = 1/2.1/3 = 1/6$ (Max)
2. $p(x_2).p(y_2/x_2) = 1/4.1/6 = 1/24$
3. $p(x_3).p(y_2/x_3) = 1/4.1/2 = 1/8$

Do $p(x_1).p(y_2/x_1)$ lớn nhất nên liệt kê y_2 vào tập hợp B_1 tương ứng với x_1, do đo cập nhật $B_1 = \{y_1, y_2\}$.

- Bước 2.3: Với y_1, ta tính được

1. $p(x_1).p(y_3/x_1) = 1/2.1/6 = 1/12$
2. $p(x_2).p(y_3/x_2) = 1/4.1/2 = 1/8$ (Max)
3. $p(x_3).p(y_3/x_3) = 1/4.1/3 = 1/12$

Do $p(x_2).p(y_3/x_1)$ lớn nhất nên liệt kê y_3 vào tập hợp B_3 tương ứng với x_2, do đo cập nhật $B_2 = \{y_3\}$.

- Kết quả: $B_1 = \{y_1, y_2\}$, $B_2 = \{y_3\}$ và $B_3 = \emptyset$.

Với phép giải mã này, các lỗi trong quá trình giải mã là:

- Xác suất truyền sai từ mã x_1:

$$p(e/x_1) = \sum p(Y = y_j \notin B_1/X = x_1) = p(y_3/x_1) = 1/6$$

- Xác suất truyền sai từ mã x_2:

$$\begin{aligned} p(e/x_2) &= \sum p(Y = y_j \notin B_2/X = x_2) \\ &= p(y_1/x_2) + p(y_2/x_2) = 1/3 + 1/6 = 1/2 \end{aligned}$$

- Xác suất truyền sai từ mã x_3:

$$p(e/x_3) = \sum p(Y = y_j \notin B_3/X = x_3)$$
$$= p(y_1/x_3) + p(y_2/x_3) + p(y_3/x_3) = 1/6 + 1/3 + 1/2 = 1$$

- Xác suất truyền sai trung bình: $p(e) = \sum_{i=1}^{M} p(X = x_i)p(e|x_i) = 11/24$

V Một số phương pháp sửa lỗi cho kênh truyền

1. Phương pháp sử dụng khoảng cách Hamming

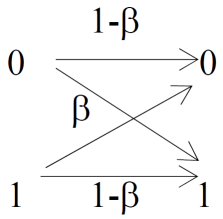
Khoảng cách Hamming giữa hai dãy ký tự có chiều dài bằng nhau là số các ký hiệu ở vị trí tương đương có giá trị khác nhau. Nói một cách khác, khoảng cách Hamming đo số lượng thay thế cần phải có để đổi giá trị của một dãy ký tự sang một dãy ký tự khác, hay số lượng lỗi xảy ra biến đổi một dãy ký tự sang một dãy ký tự khác. Ngoài ra, khoảng cách Hamming có nhiều ý nghía trong lý thuyết tính toán, được áp dụng trong nhiều lĩnh vực như: khoa học máy tinh, sinh học, y học, vv..

Định nghĩa V.1 *Định nghĩa: cho v_1 và v_2 là 2 dãy nhị phân dài n bit, ta gọi khoảng cách Hamming giữa 2 dãy v_1, v_2 là số bit tương ứng khác nhau. Ký hiệu: $d(v_1, v_2)$.*

Ví dụ. Cho hai dãy nhị phân: $v_1 = 10101010, v_2 = 10101111$. Khoảng cách Hamming giữa v_1 và v_2 là 2 hay $d(v_1, v_2) = 2$

Như đã nói ở tren, khoảng cách Hamming cho ta thấy sự tương đồng giữa hai dãy. Xét trong kênh truyền giữa đầu phát và đầu nhận, khoảng cách Hamming cho ta biết số bit lỗi trong quá trình truyền nhận.

Giả sử một từ mã w dài n bit khi được truyền tuần tự từng bit có thể sai e bit. Vấn đề đặt ra là khoảng cách (Hamming) giữa các từ mã và sai số e quan hệ với nhau như thế nào để có thể phân biệt tốt nhất đồng thời tất cả các từ mã? Ta xét trường hợp đơn giản đối với kênh nhị phân như sau: Giả sử ta truyền từ dãy nhị phân có độ dài n bits với xác suất truyền sai là β (như hình 5.5). Xét kênh truyền đối xứng nhị phân. Giả



Hình 5.5 Kênh nhị phân với lỗi là β

sử ta truyền từ dãy nhị phân có độ dài n bits với xác suất truyền sai là β.

Gọi $W = \{w_1, w_2, \ldots, w_s\}$ là tập gồm s tập độ dài từ mã có độ dài n bits. $V = \{v_1, v_2, \ldots, v_{2^n}\}$ là tập các dãy n bit nhận được ở cuối kênh với W có phân phối đều, xác suất để nhận v_j khi truyền là $p(v_j|w_i) = p_{ij}$.

Theo lược đồ giải mã tối ưu ta khi nhận v_j thì giải mã được w_i^* sao cho

$$\Pr[w_i^*|v_j] = \max_{w_i \in W} P(w_k|v_j) \tag{5.20}$$

Theo công thức Bayes, ta có:

$$P(w_k|y_j) = \frac{p(w_k)p(y_j|w_k)}{p(y_j)} \tag{5.21}$$

với $\forall w_k \in W$. Việc tìm $P(w_k|v_j)$ lớn nhất tương đương với $p(w_k) \cdot p(y_j|w_k)$ lớn nhất. Do W có phân phối đều nên $P(w_k|y_j)$ lớn nhất khi và chỉ khi

$p(y_j|w_k)$ đạt giá trị lớn nhất. Do vậy, để tìm w_i^* sao cho $P(w_i^*|v_j) = \max P(w_k|v_j)$ ta chỉ cần tìm w_i^* sao cho $P(v_j|w_i^*) = \max P(v_j|w_k)$ (Xác định qua ma trân truyền tin)

Ví dụ. Xét ma trân truyền tin

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 1/6 & 1/2 \\ 1/6 & 1/2 & 1/3 \end{pmatrix} \tag{5.22}$$

và $p(w_1) = p(w_2) = p(p_3)$. Dựa vào lược đồ giải mãi tối ưu ta thấy nếu nhận v_1 giải mã về w_1, nhận v_2 giải mã về w_3, nhận v_3 giải mã về w_2

Rõ ràng nếu sử dụng khoảng cách Hamming để xây dựng lược đồ giải mã sẽ thuật tiện hơn phương pháp Bayes. Giả sử trong quá trình truyền qua kênh ta nhận được v. Xét hai từ mã w_1 và w_2 cần chọn để giải mã cho v. Gọi $d_1 = d(v, w_1), d_2 = (v, w_2)$, ta có xác suất để nhận v khi truyền w_1 và w_2 tương ứng là:

$$p(v|w_1) = \beta^{d_1}(1-\beta)^{n-d_1} \tag{5.23}$$

$$p(v|w_2) = \beta^{d_2}(1-\beta)^{n-d_2} \tag{5.24}$$

So sánh hai xác suất:

$$\frac{p(v|w_1)}{p(v|w_2)} = \frac{\beta^{d_1}(1-\beta)^{n-d_1}}{\beta^{d_2}(1-\beta)^{n-d_2}} = \left(\frac{1-\beta}{\beta}\right)^{d_2-d_1} \tag{5.25}$$

Nếu nhiễu $\beta \in (0, 1/2)$ thì $\frac{1-\beta}{\beta} > 1$. Do đó $P(v|w_1) > p(v|w_2)$ khi và chỉ khi $d_1 < d_2$. Nếu nhiễu $\beta \in (0, 1/2)$ thì $\frac{1-\beta}{\beta} > 1$. Do đó $P(v|w_1) > p(v|w_2)$ khi và chỉ khi $d_1 < d_2$. Nếu xác suất giải mã càng lớn thì Khoảng cách Hamming càng nhỏ. Kết quả này được phát biểu dưới dạng định lý như sau:

Định lý V.1 *Kênh truyền đối xứng nhị phân với s từ mã độ dài n bit ở đầu truyền, lược đồ giải mã tối ưu có thể thay thế bằng lược đồ giải mã theo khoảng cách Hamming với nguyên lý: nếu nhận được v, sẽ giải ra với* w_i^* *sao cho*

$$d(v, w_i^*) = \min d(v, w_k) \forall w_k \in W \tag{5.26}$$

Ví dụ. Xét bộ mã $W = \{w_1 = 0000, w_2 = 10011, w_2 = 10011, w_3 = 11100, w_4 = 01111\}$. Giả sử ta nhận được dãy $v = 01011$, ta có

$$d(v, w_1) = 3; d(v, w_2) = 2; d(v, w_3); d(v, w_4) = 1.$$

Vậy v được giải về w_4 vì khoảng cách Hamming giữa v và w_4 là nhỏ nhất.

2. Phương pháp kiểm tra chẵn lẻ

Trong phương pháp này, để dễ dàng phát hiện lỗi ta thêm một dãy gồm m bit ở trước dãy bit cần gửi

$$w = \underbrace{r_1 r_2 \dots r_m}_{m \text{ bit kiểm tra}} \underbrace{r_{m+1} r_{m+2} \dots r_{m+k}}_{k \text{ bit thông tin}} \tag{5.27}$$

Ký hiệu:

- r_i: là bit thứ i của từ mã $(1 \leq i \leq n)$
- n: độ dài của từ mã hay số bit của từ mã lẻ
- m: số bit kiểm tra
- $k = n - m$ số bit thông tin, do đó có thể có tối đa 2^k bit thông tin.

Mỗi đoạn mã thông tin có duy nhất một đoạn mã kiểm tra và được xác định bởi hệ phương trình tuyến tính nhị phân sau:

$$\begin{cases} a_{11}r_1 + a_{12}r_2 + \ldots + a_{1n}r_n & = 0 \\ a_{21}r_1 + a_{22}r_2 + \ldots + a_{2n}r_n & = 0 \\ \ldots\ldots\ldots & = 0 \\ a_{n1}r_1 + a_{n2}r_2 + \ldots + a_{nn}r_n & = 0 \end{cases} \tag{5.28}$$

Gọi ma trân $A = \{a_{ij}\}_{m\times n}, a_{ij} \in \{0,1\}, i = 1..m, j = 1..n$. Ma trận A được gọi là ma trận kiểm tra chẵn lẻ có hạng là $Rank(A) = m$. Gọi $w = r_1r_2\ldots r_n$ là từ mã truyền đi, $v = r'_1r'_2\ldots r'_n$ là dãy bit đầu nhận. Ta chia các trường hợp như sau:

- Nếu $A \cdot v = 0$ thì $v = w$, ta gọi v là chẵn (trường hợp nhận đúng).
- Nếu $A \cdot v \neq 0$ thì $v \neq w$, ta gọi v là lẻ (trường hợp nhận sai).

Giả sử ta có ma trận kiểm tra chẵn lẻ là A với $Rank(A) = m$. Tìm bộ mã chẵn lẻ $W = \{w_1, w_2, \ldots, w_s\}$.

Ta xây dựng thuật toán với các bước như sau:

- Bước 1: Xác định các giá trị n, m, k, s như sau
  - Độ dài của từ mã $n =$ số cột của ma trận A.
  - Số bit kiểm tra $m =$ số dòng của ma trận A.
  - Số bit thông tin: $k = n - m$.
  - Số từ mã $s = 2^k$ của bộ mã.
- Tìm các từ mã: giả sử cần tìm từ mã thứ i, ta cần giải hệ phương trình $A \cdot w_i = 0$.

Ví dụ. Cho ma trận kiểm tra chẵn lẻ

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \tag{5.29}$$

Ta xác định bộ mã chẵn lẻ như sau:

- Bước 1: Tính được $Rank(A) = 3, n = 6, m = 3$. Số bit thông tin $k = 3$. Số từ mã $2^k = 8$.
- Bước 2: Xét các từ mã trong 8 từ mã dữ liệu từ 000 đến 111. Với $w_1 = r_1 r_2 r_3 000$ (từ mã ứng với dãy dữ liệu là 000). Giả hệ phương trình $A \cdot w_0$, tức là:

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{cases} r_1 = 0 \\ r_2 = 0 \\ r_3 = 0 \end{cases} \tag{5.30}$$

Xét từ mã $w_1 = r_1 r_2 r_3 001$, ta giải hệ phương trình $A \cdot w_1 = 0$:

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{cases} r_1 = 0 \\ r_2 + r_3 = 0 \\ r_1 + r_3 = 0 \end{cases} \rightarrow \begin{cases} r_1 = 0 \\ r_2 = 0 \\ r_3 = 1 \end{cases} \tag{5.31}$$

Xét tử mã $w_2 = r_1 r_2 r_3$, giải hệ phương trình $A \cdot w_2 = 0$:

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \rightarrow \begin{cases} r_1 = 1 \\ r_2 + r_3 = 0 \\ r_1 + r_3 = 0 \end{cases} \rightarrow \begin{cases} r_1 = 1 \\ r_2 = 1 \\ r_3 = 1 \end{cases} \tag{5.32}$$

Làm tương tự với các từ mã $w3 \ldots w_7$. Ta được bộ mã kiểm tra chẵn lẻ là: $W = \{000000, 001001, 111010, 110011, 110100, 111101, 001110, 000111\}$

Bài tập

1. *Hậu xử lý output* Giả sử ta có một kênh truyền thông với xác suất chuyển là $p(y|x)$ và dung lượng kênh $C = \max_{p(x)} I(X;Y)$. Một nhà thống kê hậu xử lý output bằng cách đưa vào hàm $\tilde{Y} = g(Y)$, nhận dược kênh $p(\tilde{y}|x)$. Anh ta khẳng định rằng phương pháp này cải tiến dung lượng kênh.

   (a) Hãy chỉ ra rằng anh ta SAI.

   (b) Trong những điều kiện nào phương pháp của anh ta không làm giảm dung lượng kênh.

2. *Máy chữ nhiễu* Xét một máy chữ gồm 26 phím.

   (a) Nếu ấn một phím thì chữ cái tương ứng được in ra, dung lượng C tính theo bit là bao nhiêu?

(b) Giả sử rằng ấn một phím không chỉ có thể in ra kí tự tương tự mà còn cả kí tự kế tiếp với xác suất như nhau (xem Hình 3 trang 81). Ví dụ ấn A thì có thể in ra A hoặc B, ấn Z thì có thể in ra Z hoặc A. Tính dung lượng kênh.

(c) Với kênh được mô tả trong phần (2b), với độ dài khối là 1 và để đảm bảo xác suất lỗi bằng 0 thì tốc độ truyền cao nhất có thể là bao nhiêu?

3. *Kênh Z* Kênh Z có các bảng chữ cái vào và ra nhị phân và xác suất chuyển $p(y|x)$ được cho bởi

$$Q = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}, \quad x, y \in \{0, 1\}.$$

Tìm dung lượng của kênh Z và phân phối xác suất vào cực đại.

4. Cho ma trận kênh

$$\mathbf{A} = \begin{pmatrix} 1/2 & 1/3 & 1/6 \\ 1/3 & 1/6 & 1/2 \\ 1/6 & 1/2 & 1/3 \end{pmatrix} \tag{5.33}$$

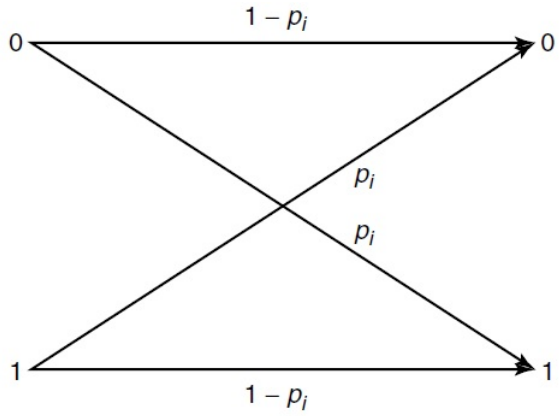
Biết xác suất ở đầu truyền: $p(x_1) = 5/10, p(x_2) = 3/10, p(x_3) = 2/10$.

(a) Tính dung lượng kênh truyền.

(b) Xây dựng lược đồ giải mã tối ưu.

(c) Tính các sai số $p(e)$ và $p_m(e)$.

5. *Kênh phụ thuộc thời gian.* Xét một kênh phụ thuộc rời rạc vào thời gian với bộ nhớ hạn chế (time-varying discret *memoryless* channel).



Đặt $Y_1, Y_2, \ldots, Y_n$ là các biến độc lập có điều kiện với các biến ngẫu nhiên $X_1, X_2, \ldots, X_n$, với các phân bố điều kiện được cho bởi

$$p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p_i(y_i|x_i).$$

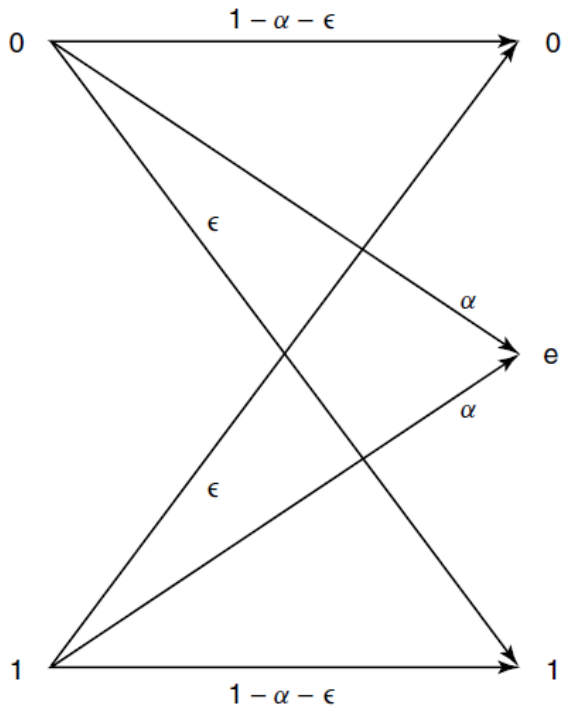
Đặt $\mathbf{X} = (X_1, X_2, \ldots, X_n)$, $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$. Hãy tìm giá trị $\max_{p(\mathbf{x})} I(\mathbf{X}; \mathbf{Y})$.

6. *Các kí tự không sử dụng.* Chứng minh rằng năng lực của kênh truyền với ma trận xác suất chuyển

$$P_{y|x} = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{pmatrix} \tag{5.34}$$

được biểu diễn bằng một hàm phân bố nhận giá trị xác suất 0 tại một trong các kí tự đầu vào. Năng lực của kênh truyền bằng bao nhiêu? Hãy nêu lí do vì sao kí tự đó lại không được sử dụng?

7. *Xóa bỏ và lỗi trong kênh truyền nhị phân.* Xét một kênh truyền có đầu vào nhị phân và có cả phần xóa bỏ và các lỗi. Kí hiệu xác suất của lỗi là ϵ và xác suất của phần xóa bỏ là α, khi đó kênh truyền sẽ có dạng



(a) Hãy tìm dung lượng của kênh.

(b) Hãy xét trường hợp đặc biệt: kênh nhị phân đối xứng ($\alpha = 0$).

(c) Hãy xét trường hợp kênh nhị phân có phần xóa bỏ ($\epsilon = 0$).

8. *Dung lượng kênh.* Tính toán dung lượng kênh sau đây với các ma trận chuyển tương ứng.

(a) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$,

$$p(y|x) = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} \tag{5.35}$$

(b) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$,

$$p(y|x) = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \tag{5.36}$$

(c) $\mathcal{X} = \mathcal{Y} = \{0, 1, 2, 3\}$,

$$p(y|x) = \begin{pmatrix} p & 1-p & 0 & 0 \\ 1-p & p & 0 & 0 \\ 0 & 0 & q & 1-q \\ 0 & 0 & 1-q & q \end{pmatrix} \tag{5.37}$$

9. *Kênh có hai đầu ra tại Y.* Cho Y_1, Y_2 độc lập có điều kiện và cùng nhận...

TÀI LIỆU THAM KHẢO

Tiếng Anh

1. Thomas M. Cover, Joy A. Thomas, *Elements of Information Theory*, second edition, NXB John Wiley & Sons (2006).

Tiếng Việt

1. Phạm Văn Cảnh, Phạm Thị Hằng, *Bài giảng lý thuyết thông tin*, Học viện An ninh nhân dân (2015)
2. Lê Quyết Thắng, Phan Tấn Tài, Dương Văn Hiếu, *Giáo trình lý thuyết thông tin*, Đại học Cần Thơ (2014).
3. Nguyễn Văn Chuyết, Nguyễn Tuấn Anh, *Cơ sở lý thuyết truyền tin*, NXB Giáo Dục (2003).
4. Nguyễn Bình, Ngô Đức Thiện, *Giáo trình Lý thuyết thông tin*, Học viện Công nghệ Bưu Chính Viễn Thông (2013)