

TÓM TẮT CÔNG THỨC

- Độ đo thông tin:

$$\log \frac{1}{p(x_i)} = -\log p(x_i)$$

Đơn vị đo: bit (lb), nat (ln), hart (lg)

$$1 \text{ nat} = \log_2(e) = 1.4427 \text{ bit}$$

$$1 \text{ hart} = \log_2(10) = 3.3219 \text{ bit}$$

- Lượng tin riêng của 1 tin rời rạc:

$$I(x_i) = \log \frac{1}{p(x_i)} = -\log p(x_i) \quad (\text{đơn vị tt})$$

- Lượng tin riêng của 1 nguồn rời rạc:

$$I(X) = \sum_{i=0}^N p(x_i) \cdot \log \frac{1}{p(x_i)} = -\sum_{i=0}^N p(x_i) \cdot \log p(x_i)$$

- Entropy của 1 tin rời rạc:

$$H(x_i) = I(x_i) = -\log p(x_i)$$

- Entropy của 1 nguồn rời rạc:

$$H(X) = -\sum_{i=0}^N p(x_i) \cdot \log p(x_i)$$

- ~~Entropy của nguồn liên tục:~~

~~$$H(X) = -\int_{-\infty}^{+\infty} w(x) \log w(x) dx, \quad w(x) \text{ là hàm mđxs}$$~~

- Lượng tin riêng, entropy của tin rời rạc đồng thời:

$$I(x_i, y_j) = H(x_i, y_j) = -\log p(x_i, y_j)$$

- Lượng tin riêng, entropy của nguồn rời rạc đồng thời:

$$I(X, Y) = H(X, Y) = -\sum_{i,j} p(x_i, y_j) \cdot \log p(x_i, y_j)$$

- ~~Entropy của nguồn liên tục đồng thời:~~

~~$$H(X, Y) = -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(x, y) \cdot \log w(x, y) dx dy$$~~

- Entropy của tin rời rạc có điều kiện:

$$H(x_i|y_j) = \log \frac{1}{p(x_i|y_j)}$$

- Entropy của nguồn rời rạc có điều kiện:

$$H(X|Y) = -\sum_{i,j} p(x_i, y_j) \cdot \log p(x_i|y_j)$$

$$H(Y|X) = -\sum_{i,j} p(x_i, y_j) \cdot \log p(y_j|x_i)$$

- ~~Entropy của nguồn liên tục có điều kiện:~~

~~$$H(X|Y) = -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(x, y) \cdot \log w(x|y) dx dy$$~~

~~$$H(Y|X) = -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(x, y) \cdot \log w(y|x) dx dy$$~~

- Tính chất của các entropy

+ Các entropy đều không âm.

+ Quan hệ giữa các entropy:

$$H(X, Y) = H(X) + H(Y|X) \\ = H(Y) + H(X|Y)$$

+ Nếu X, Y độc lập thống kê:

$$H(Y|X) = H(Y); \quad H(X|Y) = H(X)$$

$$+ 0 \leq H(X|Y) \leq H(X); \quad 0 \leq H(Y|X) \leq H(Y)$$

+ Đối với nguồn rời rạc có N tin: $H(X) \leq \log N$

- Lượng tin tương hỗ của 2 tin rời rạc:

$$I(x_i; y_j) = H(x_i) - H(x_i|y_j) = \log \frac{p(x_i|y_j)}{p(x_i)}$$

$$= \log \frac{p(x_i, y_j)}{p(x_i) \cdot p(y_j)} = \log \frac{p(y_j|x_i)}{p(y_j)} = H(y_j) - H(y_j|x_i)$$

$$I(x_i; y_j) = I(x_i) + I(y_j) - I(x_i, y_j)$$

- Lượng tin tương hỗ TB giữa 2 nguồn rời rạc:

$$I(X; Y) = \sum_{i,j} p(x_i, y_j) \cdot \log \frac{p(x_i|y_j)}{p(x_i)}$$

$$= \sum_{i,j} p(x_i, y_j) \cdot \log \frac{p(y_j|x_i)}{p(y_j)}$$

$$= \sum_{i,j} p(x_i, y_j) \cdot \log \frac{p(x_i, y_j)}{p(x_i) \cdot p(y_j)}$$

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

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

- ~~Lượng tin tương hỗ giữa 2 nguồn liên tục:~~

~~$$I(X; Y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(x, y) \cdot \log \frac{w(x, y)}{w(x) \cdot w(y)} dx dy$$~~

~~$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(x, y) \cdot \log \frac{w(x|y)}{w(x)} dx dy$$~~

~~$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} w(x, y) \cdot \log \frac{w(y|x)}{w(y)} dx dy$$~~

- Tốc độ lập tin của nguồn rời rạc:

$$R(X) = n_0 \cdot H(X) \quad \left(\frac{\text{đv thông tin}}{\text{đv thời gian}} \right)$$

+ n_0 : số tin trung bình nguồn có thể tạo ra trong 1 đơn vị thời gian (tần số tạo tin của nguồn).

+ Nếu nguồn đẳng xác suất: $p(x_i) = \frac{1}{N} \quad \forall i$:

$$R = n_0 \cdot H(X)_{\max} = n_0 \cdot \log N = F \cdot \log N$$

- ~~Tốc độ lập tin của nguồn liên tục:~~

~~$$R = 2F_{\max} \cdot H(X)$$~~

+ Đối với nguồn có công suất định hữu hạn:

~~$$R = 2F_{\max} \cdot \log(x_{\max} - x_{\min})$$~~

+ Đối với nguồn có công suất trung bình hạn chế:

~~$$R = 2F_{\max} \cdot \log \sqrt{2\pi e P_x}$$~~

BÀI TẬP LÝ THUYẾT THÔNG TIN – IT4590

Bài 1. Cho nguồn tin $X = \{x_0, x_1, x_2, x_3, x_4, x_5\}$; $P_X = \left[\frac{1}{2}; \frac{1}{4}; \frac{1}{8}; \frac{1}{16}; \frac{1}{32}; \frac{1}{32}\right]$. Tính Entropy của nguồn X.

Entropy của nguồn X:

$$H(X) = - \sum_{i=0}^5 p(x_i) \log p(x_i) = - \left(\frac{1}{2} \cdot (-1) + \frac{1}{4} \cdot (-2) + \frac{1}{8} \cdot (-3) + \frac{1}{16} \cdot (-4) + \frac{1}{32} \cdot (-5) + \frac{1}{32} \cdot (-5) \right) \\ = 1.9375 \text{ bit/kh}$$

Bài 2. Cho 2 nguồn đồng thời X, Y với ma trận xác suất:

$$P(X, Y) = \begin{bmatrix} \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} \end{bmatrix}$$

Tính entropy đồng thời $H(X, Y)$ theo bit/tin, nat/tin, hart/tin.

GIẢI:

$$\text{Ta có: } H(X, Y) = - \sum_{i,j} p(x_i, y_j) \cdot \log p(x_i, y_j) = \frac{1}{3} \cdot \log 3 + \frac{1}{6} \log 6 + \frac{1}{6} \log 6 + \frac{1}{3} \log 3 \\ = 1.918 \text{ (bit/tin)} = 1.3297 \text{ (nat/tin)} = 0.5775 \text{ (hart/tin)}$$

Nếu tính theo bit thì dùng log cơ số 2, nat – cơ số e, hart – cơ số 10.

Bài 3. Hệ thống truyền tin có nguồn tin vào X gồm 2 tin a, b đẳng xác suất. Hai tin này được mã hóa bằng mã nhị phân và truyền trên kênh nhị phân đối xứng, nguồn ra Y, có xác suất truyền đúng là 0.8, xác suất truyền sai là 0.2.

- Tính các xác suất $P(X)$, $P(Y|X)$, $P(X|Y)$, $P(X, Y)$, $P(Y)$.
- Tính $H(X)$, $H(X, Y)$, $H(Y|X)$, $H(X|Y)$, $H(Y)$, $I(X; Y)$.

GIẢI:

- a. Giả sử nguồn ra Y gồm 2 tin y_0, y_1 , từ giả thiết suy ra:

$$p(y_0|a) = p(y_1|b) = 0.8; \quad p(y_0|b) = p(y_1|a) = 0.2$$

Vì nguồn vào X gồm 2 tin a, b đẳng xác suất nên $p(a) = p(b) = \frac{1}{2} \Rightarrow P(X) = \left(\frac{1}{2}; \frac{1}{2}\right)$

Xác suất $P(Y|X)$ xác định bởi ma trận:

$$P(Y|X) = \begin{bmatrix} p(y_0|a) & p(y_0|b) \\ p(y_1|a) & p(y_1|b) \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

Ma trận xác suất đồng thời:

$$P(X, Y) = \begin{bmatrix} p(a, y_0) & p(a, y_1) \\ p(b, y_0) & p(b, y_1) \end{bmatrix} = \begin{bmatrix} p(y_0|a)p(a) & p(y_1|a)p(a) \\ p(y_0|b)p(b) & p(y_1|b)p(b) \end{bmatrix} = \begin{bmatrix} 0.4 & 0.1 \\ 0.1 & 0.4 \end{bmatrix}$$

Từ đây tính được $P(Y)$ (cộng tương ứng theo cột): $p(y_0) = p(y_1) = 1/2$

Ma trận $P(X|Y)$ xác định bởi:

$$P(X|Y) = \begin{bmatrix} p(x_0|y_0) & p(x_0|y_1) \\ p(x_1|y_0) & p(x_1|y_1) \end{bmatrix} = \begin{bmatrix} \frac{p(x_0, y_0)}{p(y_0)} & \frac{p(x_0, y_1)}{p(y_1)} \\ \frac{p(x_1, y_0)}{p(y_0)} & \frac{p(x_1, y_1)}{p(y_1)} \end{bmatrix} = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.8 \end{bmatrix}$$

- b. Dùng công thức dễ dàng tính được:

Ma Katz

$$H(X) = - \sum_i p(x_i) \cdot \log p(x_i) = 1; \quad H(Y) = - \sum_j p(y_j) \cdot \log p(y_j) = 1$$

$$H(X, Y) = - \sum_{i,j} p(x_i, y_j) \cdot \log p(x_i, y_j) = 2 \left(\frac{2}{5} \cdot \log \frac{5}{2} + \frac{1}{10} \cdot \log \frac{10}{1} \right) \approx 1.7219 \frac{\text{bit}}{\text{kh}}$$

$$H(X|Y) = H(X, Y) - H(Y) = 0.7219; \quad H(Y|X) = H(X, Y) - H(X) = 0.7219$$

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = 0.2781 \text{ bit/kh}$$

Bài 4. Giả sử kênh nhị phân được sử dụng để truyền nguồn tin nhị phân đẳng xác suất có ma trận kênh là:

$$P(Y|X) = \begin{bmatrix} 0.75 & 0.25 \\ 0.25 & 0.75 \end{bmatrix}$$

x_0, x_1 là 2 giá trị 0 và 1 trên đầu vào kênh; y_0, y_1 là 2 giá trị 0 và 1 trên đầu ra của kênh.

a. Tính $P(X, Y)$, $P(X|Y)$, $P(Y)$.

b. Tính $H(X, Y)$, $H(Y|X)$, $H(X|Y)$, $H(X)$, $H(Y)$.

GIẢI:

a. Vì nguồn tin là nguồn nhị phân đẳng xác suất nên: $p(x_0) = p(x_1) = 1/2$

Ma trận xác suất đồng thời:

$$P(X, Y) = \begin{bmatrix} p(x_0, y_0) & p(x_0, y_1) \\ p(x_1, y_0) & p(x_1, y_1) \end{bmatrix} = \begin{bmatrix} p(y_0|x_0)p(x_0) & p(y_1|x_0)p(x_0) \\ p(y_0|x_1)p(x_1) & p(y_1|x_1)p(x_1) \end{bmatrix}$$

$$= \begin{bmatrix} 0,75 \cdot 0,5 & 0,25 \cdot 0,5 \\ 0,25 \cdot 0,5 & 0,75 \cdot 0,5 \end{bmatrix} = \begin{bmatrix} \frac{3}{8} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{8} \end{bmatrix}$$

Từ đây suy ra: $P(Y) = \left(\frac{1}{2}; \frac{1}{2}\right)$

Ma trận xác suất có điều kiện $P(X|Y)$:

$$P(X|Y) = \begin{bmatrix} p(x_0|y_0) & p(x_0|y_1) \\ p(x_1|y_0) & p(x_1|y_1) \end{bmatrix} = \begin{bmatrix} p(x_0, y_0)/p(y_0) & p(x_0, y_1)/p(y_1) \\ p(x_1, y_0)/p(y_0) & p(x_1, y_1)/p(y_1) \end{bmatrix} = \begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix}$$

b. Entropy của nguồn rời rạc đồng thời:

$$H(X, Y) = - \sum_{i,j} p(x_i, y_j) \cdot \log p(x_i, y_j) = 2 \left(\frac{3}{8} \cdot \log \frac{8}{3} + \frac{1}{8} \cdot \log \frac{8}{1} \right) \approx 1,8113 \text{ bit/tin}$$

Entropy có điều kiện:

$$H(X|Y) = - \sum_{i,j} p(x_i, y_j) \log p(x_i|y_j) = 0,8113 \text{ bit/tin}$$

$$H(Y|X) = - \sum_{i,j} p(x_i, y_j) \log p(y_j|x_i) = 0,8113 \text{ bit/tin}$$

Entropy của nguồn rời rạc:

$$H(X) = - \sum_i p(x_i) \cdot \log p(x_i) = 1; \quad H(Y) = - \sum_j p(y_j) \cdot \log p(y_j) = 1$$

Bài 5. Cho nguồn $X = \{a, b, c\}$, xác suất $P(X) = \left(\frac{1}{3}; \frac{1}{3}; \frac{1}{3}\right)$. Ma trận xác suất truyền:

$$P(Y|X) = \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/6 & 2/3 & 1/6 \\ 1/6 & 1/6 & 2/3 \end{bmatrix}$$

Tính $H(X)$, $H(Y)$, $H(X, Y)$, $H(X|Y)$, $I(X; Y)$.

Ma Katz

GIẢI:

Từ giả thiết: $p(a) = p(b) = p(c) = 1/3$. Ma trận xác suất đồng thời $P(X, Y)$ xác định bởi:

$$P(X, Y) = \begin{bmatrix} p(a, y_0) & p(a, y_1) & p(a, y_2) \\ p(b, y_0) & p(b, y_1) & p(b, y_2) \\ p(c, y_0) & p(c, y_1) & p(c, y_2) \end{bmatrix} = \begin{bmatrix} p(y_0|a) \cdot p(a) & p(y_1|a) \cdot p(a) & p(y_2|a) \cdot p(a) \\ p(y_0|b) \cdot p(b) & p(y_1|b) \cdot p(b) & p(y_2|b) \cdot p(b) \\ p(y_0|c) \cdot p(c) & p(y_1|c) \cdot p(c) & p(y_2|c) \cdot p(c) \end{bmatrix}$$

$$= \begin{bmatrix} \frac{2}{3} \cdot \frac{1}{3} & \frac{1}{6} \cdot \frac{1}{3} & \frac{1}{6} \cdot \frac{1}{3} \\ \frac{1}{6} \cdot \frac{1}{3} & \frac{2}{3} \cdot \frac{1}{3} & \frac{1}{6} \cdot \frac{1}{3} \\ \frac{1}{6} \cdot \frac{1}{3} & \frac{1}{6} \cdot \frac{1}{3} & \frac{2}{3} \cdot \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 2/9 & 1/18 & 1/18 \\ 1/18 & 2/9 & 1/18 \\ 1/18 & 1/18 & 2/9 \end{bmatrix}$$

Suy ra: $p(y_0) = p(y_1) = p(y_2) = \frac{2}{9} + \frac{1}{18} + \frac{1}{18} = \frac{1}{3}$.

Và:

$$P(X|Y) = \begin{bmatrix} p(a|y_0) & p(a|y_1) & p(a|y_2) \\ p(b|y_0) & p(b|y_1) & p(b|y_2) \\ p(c|y_0) & p(c|y_1) & p(c|y_2) \end{bmatrix} = \begin{bmatrix} p(a, y_0)/p(y_0) & p(a, y_1)/p(y_1) & p(a, y_2)/p(y_2) \\ p(b, y_0)/p(y_0) & p(b, y_1)/p(y_1) & p(b, y_2)/p(y_2) \\ p(c, y_0)/p(y_0) & p(c, y_1)/p(y_1) & p(c, y_2)/p(y_2) \end{bmatrix}$$

$$= \begin{bmatrix} 2/3 & 1/6 & 1/6 \\ 1/6 & 2/3 & 1/6 \\ 1/6 & 1/6 & 2/3 \end{bmatrix}$$

Từ đó có:

+ Entropy đầu vào:

$$H(X) = - \sum_{j=0,1,2} p(x_j) \cdot \log p(x_j) = 3 \cdot \frac{1}{3} \cdot \log_2 3 = 1.585 \text{ bit/tin}$$

+ Entropy đầu ra:

$$H(Y) = - \sum_{j=0,1,2} p(y_j) \cdot \log p(y_j) = 3 \cdot \frac{1}{3} \cdot \log_2 3 = 1.585 \text{ bit/tin}$$

+ Entropy của 2 nguồn đồng thời:

$$H(X, Y) = - \sum_{i,j} p(x_i, y_j) \cdot \log p(x_i, y_j) = 2.8366 \text{ bit/tin}$$

+ Entropy có điều kiện:

$$H(X|Y) = - \sum_{i,j} p(x_i, y_j) \cdot \log p(x_i|y_j) = 1.2516 \text{ bit/tin}$$

+ Lượng tin tương hỗ giữa 2 nguồn:

$$I(X; Y) = \sum_{i,j} p(x_i, y_j) \cdot \log \frac{p(y_j|x_i)}{p(y_j)} = 3 \cdot \left(\frac{2}{9} \cdot \log_2 \frac{\frac{2}{3}}{\frac{1}{3}} \right) + 6 \cdot \left(\frac{1}{18} \cdot \log_2 \frac{\frac{1}{6}}{\frac{1}{3}} \right) = \frac{1}{3} \text{ bit/tin}$$

Bài 6. Cho hệ thống điều khiển nhiệt độ của lò sấy thuốc lá. Biết người ta sử dụng 20 sensors nhiệt độ. Nhiệt độ trong lò được khống chế ở $40 \pm 0.01^\circ\text{C}$. Nhiệt độ đo các thiết bị cấp nhiệt có thể làm cho nhiệt độ lò biến thiên từ $30 - 50^\circ\text{C}$. Yêu cầu sự sai khác nhiệt độ của lò so với nhiệt độ khống chế là trong thời gian ≤ 20 phút. Giả thiết giá trị nhiệt độ ngẫu nhiên, đẳng xác suất. Tính thông lượng của kênh truyền từ sensors về trung tâm xử lý.

GIẢI:

Từ bài ra ta có, khoảng giá trị nhiệt độ mà lò có thể nhận là: 29.99; 50.01. Nhiệt độ là biến ngẫu nhiên liên tục tuân theo phân phối đều trong đoạn [29,99; 50,01] (do đã giả thiết giá trị nhiệt độ ngẫu nhiên và đẳng xác suất).

Dễ thấy rằng đây là hệ thống truyền tin có công suất định hạn chế, ta xét với 1 kênh truyền ứng với 1 sensor: với thời gian không chế là 20 phút, ta có tần suất tạo tin: $n_0 = 20.60 = 1200 \text{ kh/s}$

Tốc độ lập tin của nguồn 1 sensor là:

$$R_0 = n_0 \cdot \log(x_{Max} - x_{Min}) = 1200 \cdot \log(50.01 - 29.99) \approx 5188.04 \text{ bit/s}$$

Tốc độ lập tin của kênh truyền có 20 sensor là: $R = 20 \cdot R_0 = 20 \cdot 5186.5 \approx 100 \text{ kBit/s}$

$$R = 20R_0 = 20 \cdot 5188.04 \approx 103 \text{ Kbit/s}$$

Thông lượng của kênh truyền cần thiết để sao cho bằng với tốc độ lập tin của nguồn trong trường hợp lí tưởng (khi kênh không nhiễu), như vậy:

$$C = R = 103 \text{ Kbit/s}$$

- Bài 7.** Cho hệ thống truyền hình theo chuẩn CCITT, khung ảnh có kích thước 3x4 được nhìn dưới góc nhìn $\alpha = 20^\circ$. Góc phân biệt độ chói (phân biệt đen trắng) là $\alpha_1 = 2'$; góc phân biệt màu là $\alpha_2 = 5'$. Mắt người có khả năng lưu ảnh trong 1/25 giây. Số ảnh gửi trong 1 giây là 50 ảnh. Ảnh được chia thành pixels thỏa mãn độ phân biệt và giả sử quét thông tin của ảnh theo đường ziczac (từ trái sang phải, từ trên xuống dưới). Thông tin về độ chói của 1 pixel là 1 trong 100 mức đẳng xác suất. Thông tin về màu của 1 pixel là 1 giá trị bộ ba màu cơ bản R-G-B (mỗi màu cơ bản có 256 mức).
- Tính tốc độ lập tin của nguồn.
 - Để truyền ảnh này bằng kênh điện thoại thì thời gian truyền 1 ảnh là bao nhiêu?

GIẢI:

a. D

b. D

Bài 8. Một tín hiệu được tạo thành từ những bit nhị phân. Do nhiễu nên tín hiệu truyền đi có thể bị lỗi ở một vài bit. Qua thống kê, ta thấy 1/4 số bit 0 truyền bị lỗi, và 1/5 số bit 1 truyền bị lỗi. Biết rằng người ra truyền đi tổng cộng 500 bit 0 và 800 bit 1. Tính xác suất nhận đúng tín hiệu.

GIẢI:

Gọi X_0, X_1 lần lượt là sự kiện gặp được bit 0, bit 1. Gọi H là sự kiện nhận đúng tín hiệu.

Ta có: \bar{H} là sự kiện tín hiệu bị lỗi; $P(H) = 1 - P(\bar{H})$.

Từ giả thiết suy ra:

$$P(X_0) = \frac{500}{500 + 800} = \frac{5}{13}; P(X_1) = \frac{800}{500 + 800} = \frac{8}{13}$$

Có 1/4 số bit 0 truyền bị lỗi, 1/5 số bit 1 truyền bị lỗi nên:

$$P(\bar{H}|X_0) = \frac{1}{4}; P(\bar{H}|X_1) = \frac{1}{5}$$

Theo công thức xác suất đầy đủ:

$$P(\bar{H}) = P(X_0) \cdot P(\bar{H}|X_0) + P(X_1) \cdot P(\bar{H}|X_1) = \frac{5}{13} \cdot \frac{1}{4} + \frac{8}{13} \cdot \frac{1}{5} = \frac{57}{260} \Rightarrow P(H) = \frac{203}{260} \approx 78.08\%$$

Vậy xác suất nhận đúng tín hiệu là $\approx 78.08\%$.

Bài 9. Cho nguồn liên tục X tuân theo phân phối đều trong đoạn $[0; a]$ ($a > 0$). Xác định $H(X)$ lần lượt trong các trường hợp $a = 1$; $a = \frac{1}{4}$; $a = 4$.

GIẢI:

Vì X tuân theo phân phối đều trong $[0; a]$ nên ta có hàm mật độ xác suất của biến ngẫu nhiên X :

$$w(x) = \begin{cases} \frac{1}{a}, & x \in [0; a] \\ 0, & \text{otherwise} \end{cases}$$

Do đó, entropy của nguồn liên tục X là:

$$H(X) = - \int_0^a w(x) \cdot \log w(x) dx = \int_0^a \frac{1}{a} \cdot \log a dx = \frac{1}{a} \cdot \log a \cdot \int_0^a dx = \frac{1}{a} \cdot \log(a) \cdot a = \log a$$

+ Với $a = 1$: $H(X) = \log_2 1 = 0$

+ Với $a = \frac{1}{4}$: $H(X) = \log_2 0.25 = -2$ (không tồn tại do entropy luôn không âm)

+ Với $a = 4$: $H(X) = \log_2 4 = 2$ (bit/tin).

Bài 10. Cho nguồn liên tục X có hàm mật độ xác suất xác định bởi:

$$w(x) = \begin{cases} \frac{e^{-x/\lambda}}{\lambda}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

Xác định $H(X)$.

GIẢI:

Entropy của nguồn liên tục X xác định bởi:

$$H(X) = \int_{-\infty}^{+\infty} w(x) \cdot \log \frac{1}{w(x)} dx = \int_0^{+\infty} \frac{e^{-x/\lambda}}{\lambda} \cdot \log \frac{\lambda}{e^{-x/\lambda}} dx = \frac{1}{\lambda} \int_0^{+\infty} e^{-x/\lambda} \cdot (\log \lambda - \log e^{-x/\lambda}) dx = \dots$$

Bài 11. Cho kênh truyền tin gồm 4 đầu vào, 3 đầu ra có ma trận kênh là:

$$P(B, A) = \begin{bmatrix} 0.1 & 0.05 & 0.05 & 0.11 \\ 0.08 & 0.03 & 0.12 & 0.04 \\ 0.13 & 0.09 & 0.14 & 0.06 \end{bmatrix}$$

Ma Katz

- a. Xác định $H(A)$, $H(B)$, $H(A|B)$, $H(B|A)$, $I(A; B)$.
- b. Nguồn A có tốc độ ký hiệu là 100 kh/s. Xác định tốc độ lập tin của nguồn A, độ dư tương đối của nguồn A và thông lượng kênh truyền tin.

GIẢI:

- a. Từ ma trận kênh suy ra:

$$\begin{aligned} P(a_1) &= 0.1 + 0.08 + 0.13 = 0.31; P(a_2) = 0.05 + 0.03 + 0.09 = 0.17; \\ P(a_3) &= 0.05 + 0.12 + 0.14 = 0.31; P(a_4) = 0.11 + 0.04 + 0.06 = 0.21 \\ P(b_1) &= 0.1 + 0.05 + 0.05 + 0.11 = 0.31; P(b_2) = 0.08 + 0.03 + 0.12 + 0.04 = 0.27; \\ P(b_3) &= 0.13 + 0.09 + 0.14 + 0.06 = 0.42 \end{aligned}$$

Do đó:

$$\begin{aligned} H(A) &= - \sum_i P(a_i) \cdot \log P(a_i) \approx 1.955 \text{ bit/kh} \\ H(B) &= - \sum_j P(b_j) \cdot \log P(b_j) \approx 1.559 \text{ bit/kh} \end{aligned}$$

Entropy của 2 nguồn A, B đồng thời:

$$H(A, B) = - \sum_{i,j} P(a_i, b_j) \cdot \log P(a_i, b_j) \approx 3.447 \text{ bit/kh}$$

Suy ra:

$$\begin{aligned} H(A|B) &= H(A, B) - H(B) \approx 3.447 - 1.559 \approx 1.888 \text{ bit/kh} \\ H(B|A) &= H(A, B) - H(A) \approx 3.447 - 1.955 \approx 1.492 \text{ bit/kh} \end{aligned}$$

Lượng tin tương hỗ:

$$I(A; B) = H(A) + H(B) - H(A, B) = 1.955 + 1.559 - 3.447 = 0.067 \text{ bit/kh}$$

- b. Tốc độ lập tin của nguồn A:

$$R = n_0 \cdot H(A) = 100 \cdot 1.955 = 195.5 \frac{\text{bit}}{\text{s}}$$

Độ dư tương đối:

$$d = 1 - \frac{H(A)}{H(A)_{\max}} = 1 - \frac{1.955}{\log_2 4} = 2.25\%$$

($H(A)$ cực đại khi nguồn đẳng xác suất, mà nguồn A gồm 4 tin nên có $H(A)_{\max}$ như trên)

Thông lượng kênh:

$$C = n_0 \cdot I(A; B) = 100 \cdot 0.067 = 6.7 \text{ bit/s}$$

Bài 12. Xét một máy đánh chữ gồm 26 phím (từ A đến Z). Giả sử trong 1 giây có thể gõ được 20 phím.

- a. Trong trường hợp lý tưởng, máy đánh chữ hoạt động chính xác, khi đó thông lượng của kênh truyền bằng bao nhiêu ?
- b. Giả sử máy đánh chữ có thể bị lỗi như sau: ấn một phím không chỉ có thể in ra ký tự tương ứng mà còn cả ký tự kế tiếp với xác suất như nhau. Ví dụ ấn A thì có thể in ra A hoặc B, ấn Z thì có thể sinh ra Z hoặc A. Tính thông lượng kênh truyền.

GIẢI:

- a. Thông lượng của kênh truyền trong trường hợp lý tưởng:

$$C = n_0 \cdot H(A)_{\max} = 20 \cdot \log_2 26 \approx 94 \text{ bit/s}$$

- b. Khi gõ một phím, có hai ký tự có thể được in ra với khả năng như nhau, do đó $H(A|B) = \log 2$, thông lượng của kênh có nhiều:

$$C = n_0 \cdot I(A; B) = 20 \cdot (H(A) - H(A|B)) = 20 \cdot (\log_2 26 - \log_2 2) \approx 74 \text{ bit/s}$$

Ma Katz

Bài 13. Cho kênh nhị phân đối xứng có xác suất truyền lỗi là $p = 10^{-3}$.

a. Tính lượng tin tương hỗ.

b. Nếu chuỗi bit được truyền đi có chiều dài 1KB thì có bao nhiêu bit lỗi?

GIẢI:

a. Gọi nguồn vào là X , nguồn ra là Y , do đây là kênh nhị phân đối xứng nên ta có:

$$X = \{x_0, x_1\}; Y = \{y_0, y_1\}; P(X) = \left(\frac{1}{2}; \frac{1}{2}\right)$$

Ma trận kênh:

$$P(Y|X) = \begin{bmatrix} p(y_0|x_0) & p(y_0|x_1) \\ p(y_1|x_0) & p(y_1|x_1) \end{bmatrix} = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

Entropy có điều kiện:

$$H(Y|X) = 2(1-p) \cdot \log \frac{1}{1-p} + 2p \cdot \log \frac{1}{p}$$

Ma trận xác suất đồng thời:

$$P(X, Y) = \begin{bmatrix} p(x_0, y_0) & p(x_0, y_1) \\ p(x_1, y_0) & p(x_1, y_1) \end{bmatrix} = \begin{bmatrix} (1-p) \cdot \frac{1}{2} & p \cdot \frac{1}{2} \\ p \cdot \frac{1}{2} & (1-p) \cdot \frac{1}{2} \end{bmatrix}$$

Suy ra:

$$P(y_0) = \sum_i p(x_i, y_j) = (1-p) \cdot \frac{1}{2} + p \cdot \frac{1}{2} = \frac{1}{2}; P(y_1) = \dots = \frac{1}{2}$$

Entropy nguồn ra Y : $H(Y) = -\sum_j p(y_j) \cdot \log p(y_j) = 1$

Lượng tin tương hỗ:

$$I(X; Y) = H(Y) - H(Y|X) = 1 - 2(1-p) \cdot \log \frac{1}{1-p} - 2p \cdot \log \frac{1}{p} \approx 0,98859 \text{ bit/kh}$$

b. Ta có:

$$1KB = 1024B = 8192 \text{ bit}$$

Xác suất truyền lỗi là $p = 10^{-3}$ nên số bit lỗi là:

$$8192 \cdot 10^{-3} = 8 \text{ bit (lấy phần nguyên trên)}$$

Bài 14. Xét một bản tin bao gồm họ và tên của bạn, nơi sinh của bạn (gồm 3 thông tin: phường/xã, quận/huyện, tỉnh/TP), bao gồm các chữ cái không dấu, không phân biệt chữ hoa chữ thường, không có khoảng trắng. Nguồn X gồm các tin là các chữ cái khác nhau trong bản tin, xác suất của tin trong nguồn là tần suất xuất hiện của từng chữ cái trong bản tin.

a. Viết bản tin ứng với thông tin của bạn.

b. Xác định mô hình của nguồn X ứng với bản tin trên.

c. Tính entropy của nguồn X .

GIẢI:

a. Bản tin như sau:

LAMMINHANHQUYNHMAIHAIBATRUNGHA NOI

Ma Katz

b. Số kí tự trong bản tin: $N = 33$

Nguồn $X = \{A, B, G, H, L, M, N, O, Q, R, T, U, Y\}$

Xác suất xuất hiện của từng tin trong nguồn:

$$P(A) = \frac{6}{33}; P(B) = \frac{1}{33}; P(G) = \frac{1}{33}; P(H) = \frac{5}{33}; P(L) = \frac{1}{33}; P(M) = \frac{3}{33}; P(N) = \frac{5}{33}; P(O) = \frac{1}{33};$$

$$P(Q) = \frac{1}{33}; P(R) = \frac{1}{33}; P(T) = \frac{1}{33}; P(U) = \frac{2}{33}; P(Y) = \frac{1}{33}.$$

Suy ra:

$$P(X) = \left(\frac{6}{33}; \frac{1}{33}; \frac{1}{33}; \frac{5}{33}; \frac{1}{33}; \frac{3}{33}; \frac{5}{33}; \frac{1}{33}; \frac{1}{33}; \frac{1}{33}; \frac{2}{33}; \frac{1}{33} \right)$$

c. Entropy của nguồn X :

$$H(X) = \sum_{i=1}^{14} P(x_i) \cdot \log_2 \frac{1}{P(x_i)}$$

$$= \frac{6}{33} \log_2 \frac{33}{6} + 8 \left(\frac{1}{33} \cdot \log_2 33 \right) + 2 \left(\frac{5}{33} \cdot \log_2 \frac{33}{5} \right) + \frac{3}{33} \log_2 11 + \frac{2}{33} \log_2 \frac{33}{2}$$

$$= 2.7402 \text{ bit/tin}$$

Bài 15. Nguồn X gồm 2 tin có xác suất lần lượt là p và $1 - p$, với $p = 1/q$, q là giá trị ứng với chữ cái đầu tiên trong họ của bạn, được cho trong bảng sau:

A=1; B=2; C=3; D=4; E=5; F=6; G=7; H=8; I=9; K=10; L=11; M=12; N=13; O=14; P=15;

Q=16; R=17; S=18; T=19; U=20; V=21; X=22; Y=23; Z=24

Ví dụ, nếu bạn họ Nguyễn, chữ cái đầu là N, khi đó $q = 13$; $p = 1/13$.

Mã trận kênh được cho bởi xác suất:

$$P(Y|X) = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$$

Tính $H(X)$, $H(X,Y)$, $I(X; Y)$.

GIẢI:

Họ Lâm $\Rightarrow p = \frac{1}{11} \Rightarrow P(X) = \left(\frac{1}{11}; \frac{10}{11} \right)$

Mã trận xác suất đồng thời:

$$P(X, Y) = \begin{bmatrix} p(x_0, y_0) & p(x_0, y_1) \\ p(x_1, y_0) & p(x_1, y_1) \end{bmatrix} = \begin{bmatrix} p(y_0|x_0)p(x_0) & p(y_1|x_0)p(x_0) \\ p(y_0|x_1)p(x_1) & p(y_1|x_1)p(x_1) \end{bmatrix} = \begin{bmatrix} \frac{2}{3} \cdot \frac{1}{11} & \frac{1}{3} \cdot \frac{1}{11} \\ \frac{1}{3} \cdot \frac{10}{11} & \frac{2}{3} \cdot \frac{10}{11} \end{bmatrix} = \begin{bmatrix} \frac{2}{33} & \frac{1}{33} \\ \frac{10}{33} & \frac{20}{33} \end{bmatrix}$$

Từ đây suy ra: $P(y_0) = \frac{2}{33} + \frac{10}{33} = \frac{12}{33}$; $P(y_1) = \frac{1}{33} + \frac{20}{33} = \frac{21}{33}$; $P(Y) = \left(\frac{12}{33}; \frac{21}{33} \right)$

Entropy của nguồn X :

$$H(X) = - \sum_i p(x_i) \cdot \log p(x_i) = \frac{1}{11} \cdot \log_2 11 + \frac{10}{11} \cdot \log_2 \frac{11}{10} = 0.4395 \text{ bit/tin}$$

Entropy của nguồn rời rạc đồng thời:

$$H(X, Y) = - \sum_{i,j} p(x_i, y_j) \cdot \log p(x_i, y_j) = \frac{2}{33} \cdot \log_2 \frac{33}{2} + \frac{1}{33} \cdot \log_2 33 + \frac{10}{33} \cdot \log_2 \frac{33}{10} + \frac{20}{33} \cdot \log_2 \frac{33}{20}$$

$$= 1.3578 \frac{\text{bit}}{\text{tin}}$$

Entropy của nguồn ra Y :

$$H(Y) = - \sum_i p(y_i) \cdot \log p(y_i) = \frac{12}{33} \cdot \log_2 \frac{33}{12} + \frac{21}{33} \cdot \log_2 \frac{33}{21} = 0.9457 \frac{\text{bit}}{\text{tin}}$$

Ma Katz

Lượng tin tương hỗ:

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = 0.0274 \frac{\text{bit}}{\text{tin}}$$

Bài 16. Cho bộ mã

a – 0000 b – 1002 c – 2100 d – 222 e – 2101 f – 1111
g – 0210 h – 0220 i – 2020 k – 120 l – 221 m – 212

a. Vẽ cây mã

b. Dựa vào cây mã để xác định các đặc tính và tham số cơ bản của bộ mã

GIẢI:

a. Cây mã tự vẽ:

b. Từ cây mã rút ra các đặc tính của bộ mã như sau:

+ Tính đều: bộ mã không đều do các từ mã có độ dài khác nhau (có từ mã độ dài 3, có từ mã độ dài 4).

+ Tính đầy: bộ mã chưa đầy do các nhánh của cây mã chưa tỏa ra hết.

+ Tính prefix: bộ mã có tính prefix do các nút biểu diễn từ mã đều là nút lá.

+ Tính phân tách được: do bộ mã có tính prefix nên cột 1 của bảng thử mã chắn chắn rỗng, nên bộ mã là phân tách được.

Các tham số của bộ mã:

+ Cơ số: từ các nút tỏa ra không quá 3 nhánh, nên bộ mã có cơ số 3.

+ Số từ mã: $N = 12$

Bài 17. Sử dụng bảng thử tính phân tách để kiểm tra xem bộ mã: 11, 201, 110, 021, 011, 1010 có phân tách được hay không? Hãy vẽ cây mã của bộ mã này.

GIẢI:

Bảng thử tính phân tách:

Từ mã	Cột 1	Cột 2
11	0 (11 là phần đầu 110)	21 (0 là phần đầu 021)
201		11 (0 là phần đầu 011) 11 trùng với từ mã
110		
021		
011		
1010		

Vậy bộ mã này không phân tách được.

Bài 18. Cho bản tin 001011011101001010110001. Mã hóa bản tin bằng thuật toán Lempel-Ziv.

GIẢI:

B1: Tách từ thông tin theo thứ tự từ điển:

Ma Katz

0-01-011-0111-010-0101-0110-00-1

B2: Lập từ điển mã hóa

STT	Từ	Từ cũ (STT từ cũ)	Kí tự lấy thêm
0	---	---	---
1	0	rỗng (0)	0
2	01	0 (1)	1
3	011	01 (2)	1
4	0111	011 (3)	1
5	010	01 (2)	0
6	0101	010 (5)	1
7	0110	011 (3)	0
8	00	0 (1)	0
9	1	rỗng (0)	1

B3: Mã hóa: từ mã gồm 2 phần là mã nhị phân cho số thứ tự của phần từ cũ + phần kí tự lấy thêm

Độ dài từ mã: $n = n_1 + n_2$; $n_1 = \lceil \log_2 N \rceil$ với N là số từ trong từ điển, n_2 là số kí tự lấy thêm (1)

Có: $N = 9 \Rightarrow n_1 = 4, n_2 = 1 \Rightarrow n = 5$

STT từ cũ ở dạng nhị phân	Kí tự lấy thêm	Từ mã
0000	0	00000
0001	1	00011
0010	1	00101
0011	1	00111
0010	0	00100
0101	1	01011
0011	0	00110
0001	0	00010
0000	1	00001

Bài 19. X

Bài 20. Hk

Bài 21.

Ma Katz