

Ví dụ: Rút bóng và Lượng thông tin

Nhân Nguyễn Văn

April 2025

Bối cảnh giả định: Rút bóng từ hộp

Có một hộp chứa các quả bóng, mỗi quả có một màu.

Trường hợp A: Hộp 1 (Xác suất đều)

- Gồm 4 quả: đỏ, xanh, vàng, trắng.
- Mỗi màu xuất hiện 1 lần nên xác suất mỗi màu:

$$p(x) = \frac{1}{4} = 0.25$$

Trường hợp B: Hộp 2 (Mất cân bằng)

- Gồm 4 quả: 3 đỏ, 1 trắng.
- Xác suất:

$$p(\text{đỏ}) = 0.75, \quad p(\text{trắng}) = 0.25$$

Ý nghĩa công thức thứ nhất: $I(x) = -\log_2 p(x)$

Công thức này biểu diễn **lượng thông tin (sự bất ngờ)** khi nhận được kết quả x .

So sánh các trường hợp:

Trường hợp	Màu bóng	Xác suất $p(x)$	$I(x)$ (bits)	Giải thích
A	bất kỳ màu nào	0.25	2	Mỗi màu đều bất ngờ như nhau.
B	đỏ	0.75	0.415	Ít bất ngờ (xảy ra thường xuyên).
B	trắng	0.25	2	Rất bất ngờ vì ít xảy ra.

Ghi nhớ: Thông tin càng bất ngờ \rightarrow càng nhiều bit để mã hóa!

Ý nghĩa công thức thứ hai: Entropy $H(x)$

Công thức entropy cho biết **lượng thông tin trung bình** nhận được sau mỗi lần rút bóng.

So sánh các trường hợp:

Trường hợp	Entropy $H(x)$ (bits)	Giải thích
A	2	Các màu đều nhau \rightarrow thông tin tối đa.
B	0.811	Đỏ nhiều \rightarrow ít bất ngờ hơn, thông tin trung bình thấp.

Nhận định:

- Entropy cao \rightarrow thông tin phong phú, khó đoán.
- Entropy thấp \rightarrow có nhiều lặp lại, ít thông tin mới.

Tính cụ thể $H(x)$ trong trường hợp B

Hộp gồm:

- 3 quả đỏ $\rightarrow p(\text{đỏ}) = 0.75$
- 1 quả trắng $\rightarrow p(\text{trắng}) = 0.25$

Tính entropy:

$$\begin{aligned} H(x) &= -[0.75 \cdot \log_2(0.75) + 0.25 \cdot \log_2(0.25)] \\ &= -[0.75 \cdot (-0.415) + 0.25 \cdot (-2)] \\ &= -[-0.31125 - 0.5] = 0.81125 \text{ bits} \end{aligned}$$

Tổng kết

- Công thức $I(x)$ cho biết **mức độ bất ngờ** của 1 sự kiện.
- Công thức $H(x)$ cho biết **trung bình mỗi lần nhận được bao nhiêu thông tin mới**.
- Càng đều nhau \rightarrow càng khó đoán \rightarrow cần nhiều bit để mã hóa \rightarrow entropy cao.