Entropy Codings

Probability, Information, Entropy

August 2018

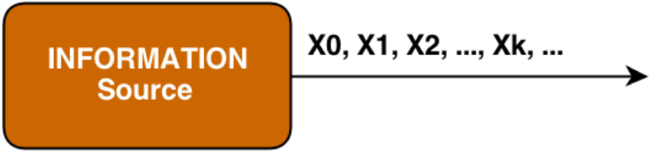
Content

- Information Model

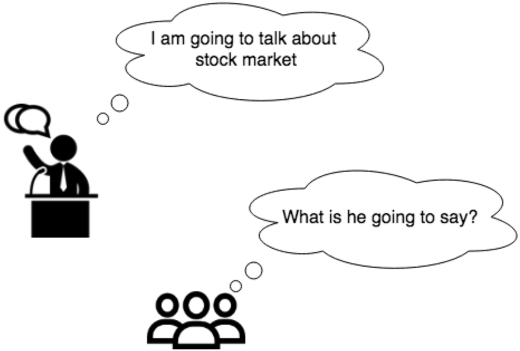
Information Source Model - Nguồn thông tin

- Some examples:
  - Tiếng gà gáy, tiếng trống canh
  - Đèn hiệu, đốt lửa, ...
  - Các file dữ liệu (e.g.: text, audio, image, video files)
  - Các "data stream"(ví dụ: bit stream)
  - A typewritter, mouse clicks of users on a webpage
  - ...
- Some common characteristics
  - Always generate *Something* as *Output* (observable, iterable, finite/infinite)
  - Roughly sequetial (Or can be seen as sequential)
- How to quantitatively characterize them?

Probabilistic Information Source Model



- Object generates "symbols"
- Randomly (Do you believe it?)



- Symbols are generated DETERMINISTICALLY. But its rules may be completely unknown => appears almost completely random to the world.
- Information source can be affected by the world so that output can be changed

Probabilistic Information Source Model

Symbols and Alphabet

- Symbols can be modeled by
  - Apply assumptions to simplify
  - Observe for a long time and figure out symbol set
- Alphabet
  - Symbol set: $X = \{x_i | i = 0..N\}$
  - And a probability distribution over symbol set $p_i = Pr(X = X_i)$
- i.i.d. Information source model
  - Symbols are identical and independent distributed random variables

Concept of independence

- A and B are independent if occurence of A doesn't affect occurence of B and vice versus
  - $P(A \cap B) = P(AB) = P(A)P(B)$
  - $P(A|B) = P(A)$ and $P(B|A) = P(B)$
- If you observe two random variables A and B, can you tell how independent they are?

Correlation

- Covariance, variance, standard deviation

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} \quad (1)$$

$$= \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sqrt{E[(X-\mu_X)^2]}\sqrt{E[(Y-\mu_Y)^2]}} \quad (2)$$

- Sample version

$$r_{X,Y} = \frac{\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i-\bar{y})^2}} \quad (3)$$

- Uncorrected sample standard deviation

$$s_N = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{4}$$

- Corrected sample standard deviation

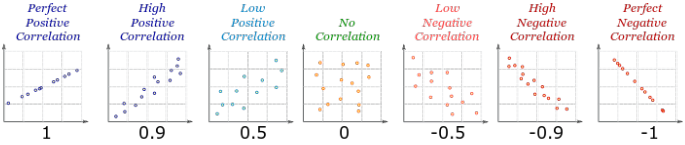
$$\dot{s}_N = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{5}$$

- Why?
- Sample correlation with corrected standard deviation

$$r_{X,Y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{(N-1)\dot{s}_X \dot{s}_Y} \tag{6}$$

Correlation - Application

- Just by observing 2 datasets, we can tell how two ranvars are independent or dependent,
- and how they are independent/dependent



- What is its application, then?

But, be careful

- Let X is a random variable with $\mu_X = 0$ and $Y = X^2$
  - Are they still random variables?
- Let's calculate $\rho_{X,Y}$
- The answer is $\rho_{X,Y} = 0$ (Surprise?)

- Correlation only detect linear dependency

Information content

- Symbol x with probability $Pr(X = x) = p(x)$, amount of information that symbol x carries is

$$I(x) = -log_2(p(x)) = -log(p(x)) \tag{7}$$

- High probability = highly predictable = easy to predict = little information
- $I(x)$ only depends on $p(x)$. We can "code" x with any other symbol without changing I
- *bit* is unit of $I(x)$

Consider 1 flip-flop that can store either '0' or '1' with equal probability (1 bit)

$$\begin{aligned} I('0') &= I('1') \qquad (8) \\ &= -log(0.5) = -log(2^{-1}) = 1 \qquad (9) \end{aligned}$$

This is true for all other similar models:

- Tung xu, xấp ngửa
- Bứt cánh hoá
- Trời mưa hay trời nắng
- ...

Information content - Examples

- Giving 3 coins, toss them.
  - $I(\text{“}HHH'') = ?$
  - $I(\text{“}TTT'') = ?$
  - $I(\text{“}THH'') = ?$
- How about N coins?

Information measurement of a source - Entropy

Let's measure amount of information for an information source instead of a symbol

- Entropy: Average information of all symbols of a source

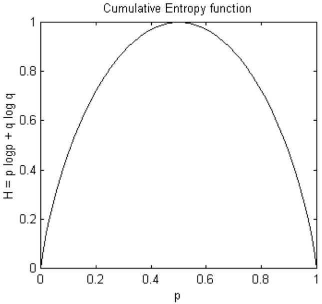
$$
\begin{aligned}
H(X) &= E[I(x)|\forall x \in X] && (10) \\
&= \sum_{x \in X} p(x) I(x) && (11) \\
&= -\sum_{x \in X} p(x) log(p(x)) && (12)
\end{aligned}
$$

Examples: 1-bit info. source

$$X_{1bit} = \{'0', '1'\}, \begin{cases} p('0') & = & p \\ p('1') & = & 1-p \end{cases}$$

- Calculate entropy of 1-bit information source

$$\begin{aligned} H(X_{1bit}) &= -p \times log(p) - (1-p) \times log(1-p) \\ &= 0.5 \times 1 + 0.5 \times 1 \\ &= 1 \end{aligned}$$



Examples: N-bit info. source

- $N = 3$

$$X_{Nbit} = \{'000', '001', '010', ..., '111'\}$$
$$= \{0, 1, 2, ..., 7\}$$
$$p(0) = p(1) = ... = p(k) = \frac{1}{2^3}$$

- Generalize for arbitrarily N

$$X_{Nbit} = \{0, 1, 2, ..., 2^N - 1\}$$
$$p(k) = \frac{1}{2^N}$$

- Entropy

$$H(X_{Nbit}) = \sum_{k=1}^{2^N} \frac{1}{2^N} log(2^N) = N$$

Examples: Binomial info. source

- Binomial info. source with parameter N: Tung N đồng xu, lấy số xu ngửa làm đầu ra

$$\begin{aligned} X_{Nbin} &= \{\text{'0 ngửa'}, \text{'1 ngửa'}, ..., \text{'N ngửa'}\} \\ &= \{0, 1, ..., N\} \\ p(k) &= \binom{N}{k} \frac{1}{2^N} = \frac{N!}{k!(N-k)!} \frac{1}{2^N} \end{aligned}$$

- Entropy

$$H(X_{Nbin}) = \sum_{k=0}^{N} \frac{N!}{k!(N-k)!} \frac{1}{2^N} \log \left(\frac{N!}{k!(N-k)!} \frac{1}{2^N} \right)$$

Binomial distribution - some cases

N	$H(N-bit)$	$H(N-bin)$
2	2	1.5
3	3	1.811
4	4	2.03

- $H(N-bin) \leq H(N-bit)$! Why?

Examples: geometric info. source

- Geometric information source: Tung xu cho đến khi được mặt ngửa. Số lần tung là output

$$X_{geom} = \{1, 2, 3, ...\infty\}$$
$$p(k) = \frac{1}{2^k}$$

- Calculate entropy:

$$H(X_{geom}) = \sum_{k=1}^{\infty} \frac{1}{2^k} \log 2^k$$
$$= \sum_{i=1}^{\infty} \frac{k}{2^k}$$

Examples: Negative binomial info. source

- Negative binomial information source: Tung một đồng xu liên tiếp cho đến khi có r lần ngửa

$$X_{negBin}(r) = \{r, r+1, r+2, ...\infty\}$$

$$p(k) = \binom{k}{k-r+1}\frac{1}{2^k} = \frac{k!}{(k-r+1)!(r-1)!}\frac{1}{2^k}$$

- Calculate Entropy:

$$H(X_{negBin}(r)) = -\sum_{k=r}^{\infty}\binom{k}{k-r+1}\frac{1}{2^k}log\left(\binom{k}{k-r+1}\frac{1}{2^k}\right)$$

Entropy characteristics

- Additive: If X & Y are independent

$$H(X, Y) = H(X) + H(Y) \tag{13}$$

- Given an alphabet $X = \{x_i | i = 1..N\}$

$$H(X) \leq log(N) \tag{14}$$

$$\tag{15}$$

$H(X)$ get maximized iff $p(x_i) = \frac{1}{N}, \forall i$

Data compression

- How many bit are needed to describe the outcome of an symbol (outcome of a random experiment)?

Shanon source coding theorem

Theorem:

- N i.i.d. random variables, each with entropy H(X) can ben compressed into more than $NH(X)$ bits with negligible risk of information loss, as $N \to \infty$

- Compression: compressor function that maps $x \to \dot{x} = c(x)$ which is a bit string
- Decompression: decompressor function that maps $\dot{x} \to x = d(\dot{x})$