

Shannon's source coding theorem

From Wikipedia, the free encyclopedia

This article is about the theory of source coding in data compression. For the term in computer programming, see Source code.

In information theory, **Shannon's source coding theorem** (or **noiseless coding theorem**) establishes the limits to possible data compression, and the operational meaning of the Shannon entropy.

The **source coding theorem** shows that (in the limit, as the length of a stream of independent and identically-distributed random variable (i.i.d.) data tends to infinity) it is impossible to compress the data such that the code rate (average number of bits per symbol) is less than the Shannon entropy of the source, without it being virtually certain that information will be lost. However it is possible to get the code rate arbitrarily close to the Shannon entropy, with negligible probability of loss.

The **source coding theorem for symbol codes** places an upper and a lower bound on the minimal possible expected length of codewords as a function of the entropy of the input word (which is viewed as a random variable) and of the size of the target alphabet.

Contents

- 1 Statements
 - 1.1 Source coding theorem
 - 1.2 Source coding theorem for symbol codes
- 2 Proof: Source coding theorem
- 3 Proof: Source coding theorem for symbol codes
- 4 Extension to non-stationary independent sources
 - 4.1 Fixed Rate lossless source coding for discrete time non-stationary independent sources
- 5 See also
- 6 References

Statements

Source coding is a mapping from (a sequence of) symbols from an information source to a sequence of alphabet symbols (usually bits) such that the source symbols can be exactly recovered from the binary bits (lossless source coding) or recovered within some distortion (lossy source coding). This is the concept behind data compression.

Source coding theorem

In information theory, the **source coding theorem** (Shannon 1948)^[1] informally states that (MacKay 2003, pg. 81,^[2] Cover:Chapter 5^[3]):

N i.i.d. random variables each with entropy $H(X)$ can be compressed into more than $N H(X)$ bits with negligible risk of information loss, as $N \rightarrow \infty$; but conversely, if they are compressed into fewer than $N H(X)$ bits it is virtually certain that information will be lost.

Source coding theorem for symbol codes

Let Σ_1, Σ_2 denote two finite alphabets and let Σ_1^* and Σ_2^* denote the set of all finite words from those alphabets (respectively).

Suppose that X is a random variable taking values in Σ_1 and let f be a uniquely decodable code from Σ_1^* to Σ_2^* where $|\Sigma_2| = a$. Let S denote the random variable given by the word length $f(X)$.

If f is optimal in the sense that it has the minimal expected word length for X , then (Shannon 1948):

$$\frac{H(X)}{\log_2 a} \leq \mathbb{E}S < \frac{H(X)}{\log_2 a} + 1$$

Proof: Source coding theorem

Given X is an i.i.d. source, its time series X_1, \dots, X_n is i.i.d. with entropy $H(X)$ in the discrete-valued case and differential entropy in the continuous-valued case. The Source coding theorem states that for any $\varepsilon > 0$ for any rate larger than the entropy of the source, there is large enough n and an encoder that takes n i.i.d. repetition of the source, $X^{1:n}$, and maps it to $n(H(X) + \varepsilon)$ binary bits such that the source symbols $X^{1:n}$ are recoverable from the binary bits with probability at least $1 - \varepsilon$.

Proof of Achievability. Fix some $\varepsilon > 0$, and let

$$p(x_1, \dots, x_n) = \Pr[X_1 = x_1, \dots, X_n = x_n].$$

The typical set, A_n^ε , is defined as follows:

$$A_n^\varepsilon = \left\{ (x_1, \dots, x_n) : \left| -\frac{1}{n} \log p(x_1, \dots, x_n) - H_n(X) \right| < \varepsilon \right\}.$$

The Asymptotic Equipartition Property (AEP) shows that for large enough n , the probability that a sequence generated by the source lies in the typical set, A_n^ε , as defined approaches one. In particular there for large enough n , $P(A_n^\varepsilon) > 1 - \varepsilon$ (See AEP for a proof):

The definition of typical sets implies that those sequences that lie in the typical set satisfy:

$$2^{-n(H(X)+\varepsilon)} \leq p(x_1, \dots, x_n) \leq 2^{-n(H(X)-\varepsilon)}$$

Note that:

- The probability of a sequence from X being drawn from A_n^ε is greater than $1 - \varepsilon$.
- $|A_n^\varepsilon| \leq 2^{n(H(X)+\varepsilon)}$ since the probability of the whole set A_n^ε is at most one.
- $|A_n^\varepsilon| \geq (1 - \varepsilon)2^{n(H(X)-\varepsilon)}$. For the proof, use the upper bound on the probability of each term in typical set and the lower bound on the probability of the whole set A_n^ε .

Since $|A_n^\varepsilon| \leq 2^{n(H(X)+\varepsilon)}$, $n \cdot (H(X) + \varepsilon)$ bits are enough to point to any string in this set.

The encoding algorithm: The encoder checks if the input sequence lies within the typical set; if yes, it outputs the index of the input sequence within the typical set; if not, the encoder outputs an arbitrary $n(H(X) + \varepsilon)$ digit number. As long as the input sequence lies within the typical set (with probability at least $1 - \varepsilon$), the encoder doesn't make any error. So, the probability of error of the encoder is bounded above by ε .

Proof of Converse. The converse is proved by showing that any set of size smaller than A_n^ε (in the sense of exponent) would cover a set of probability bounded away from 1.

Proof: Source coding theorem for symbol codes

For $1 \leq i \leq n$ let s_i denote the word length of each possible x_i . Define $q_i = a^{-s_i}/C$, where C is chosen so that $q_1 + \dots + q_n = 1$. Then

$$\begin{aligned}
 H(X) &= - \sum_{i=1}^n p_i \log_2 p_i \\
 &\leq - \sum_{i=1}^n p_i \log_2 q_i \\
 &= - \sum_{i=1}^n p_i \log_2 a^{-s_i} + \sum_{i=1}^n p_i \log_2 C \\
 &= - \sum_{i=1}^n p_i \log_2 a^{-s_i} + \log_2 C \\
 &\leq - \sum_{i=1}^n -s_i p_i \log_2 a \\
 &\leq \mathbb{E} S \log_2 a
 \end{aligned}$$

where the second line follows from Gibbs' inequality and the fifth line follows from Kraft's inequality:

$$C = \sum_{i=1}^n a^{-s_i} \leq 1$$

so $\log C \leq 0$.

For the second inequality we may set

$$s_i = \lceil -\log_a p_i \rceil$$

so that

$$-\log_a p_i \leq s_i < -\log_a p_i + 1$$

and so

$$a^{-s_i} \leq p_i$$

and

$$\sum a^{-s_i} \leq \sum p_i = 1$$

and so by Kraft's inequality there exists a prefix-free code having those word lengths. Thus the minimal S satisfies

$$\begin{aligned} \mathbb{E}S &= \sum p_i s_i \\ &< \sum p_i (-\log_a p_i + 1) \\ &= \sum -p_i \frac{\log_2 p_i}{\log_2 a} + 1 \\ &= \frac{H(X)}{\log_2 a} + 1 \end{aligned}$$

Extension to non-stationary independent sources

Fixed Rate lossless source coding for discrete time non-stationary independent sources

Define typical set A_n^ε as:

$$A_n^\varepsilon = \left\{ x_1^n : \left| -\frac{1}{n} \log p(X_1, \dots, X_n) - \overline{H}_n(X) \right| < \varepsilon \right\}.$$

Then, for given $\delta > 0$, for n large enough, $\Pr(A_n^\varepsilon) > 1 - \delta$. Now we just encode the sequences in the typical set, and usual methods in source coding show that the cardinality of this set is smaller than $2^{n(\overline{H}_n(X) + \varepsilon)}$. Thus, on an average, $\overline{H}_n(X) + \varepsilon$ bits suffice for encoding with probability greater than $1 - \delta$, where ε and δ can be made

arbitrarily small, by making n larger.

See also

- Channel coding
- Noisy Channel Coding Theorem
- Error exponent
- Asymptotic Equipartition Property (AEP)

References

1. C.E. Shannon, "A Mathematical Theory of Communication (<http://plan9.bell-labs.com/cm/ms/what/shannonday/shannon1948.pdf>)", *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, July, October, 1948
2. David J. C. MacKay. *Information Theory, Inference, and Learning Algorithms* (<http://www.inference.phy.cam.ac.uk/mackay/itila/book.html>) Cambridge: Cambridge University Press, 2003. ISBN 0-521-64298-1
3. Cover, Thomas M. (2006). "Chapter 5: Data Compression". *Elements of Information Theory*. John Wiley & Sons. ISBN 0-471-24195-4.

Retrieved from "https://en.wikipedia.org/w/index.php?title=Shannon%27s_source_coding_theorem&oldid=657141983"

Categories: Information theory | Coding theory | Data compression | Presentation layer protocols
| Mathematical theorems in theoretical computer science

- This page was last modified on 19 April 2015, at 05:25.
- Text is available under the Creative Commons Attribution-ShareAlike License; additional terms may apply. By using this site, you agree to the Terms of Use and Privacy Policy. Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc., a non-profit organization.