

Noisy channel coding theorem

From Wikipedia, the free encyclopedia
(Redirected from Channel coding)

In information theory, the **noisy-channel coding theorem** establishes that however contaminated with noise interference a communication channel may be, it is possible to communicate digital data (information) error-free up to a given maximum rate through the channel. This surprising result, sometimes called the **fundamental theorem of information theory**, or just **Shannon's theorem**, was first presented by Claude Shannon in 1948.

The **Shannon limit** or **Shannon capacity** of a communications channel is the theoretical maximum information transfer rate of the channel, for a particular noise level.

Contents

- 1 Overview
- 2 Mathematical statement
- 3 Outline of Proof
 - 3.1 Achievability for discrete memoryless channels
 - 3.2 Converse for discrete memoryless channels
- 4 Channel coding theorem for non-stationary memoryless channels
 - 4.1 Outline of the proof
- 5 References
- 6 See also
- 7 External links

Overview

Proved by Claude Shannon in 1948, the theorem describes the maximum possible efficiency of error-correcting methods versus levels of noise interference and data corruption. The theory doesn't describe *how to construct* the error-correcting method, it only tells us how good the *best possible* method can be. Shannon's theorem has wide-ranging applications in both communications and data storage applications. This theorem is of foundational importance to the modern field of information theory.

The Shannon theorem states that given a noisy channel with information capacity C and information transmitted at a rate R , then if

$$R < C$$

there exists a coding technique which allows the probability of error at the receiver to be made arbitrarily small. This means that theoretically, it is possible to transmit information without error up to a limiting rate, C .

The converse is also important. If

$$R > C$$

an arbitrarily small probability of error is not achievable. So, information cannot be guaranteed to be transmitted reliably across a channel at rates beyond the channel capacity. The theorem does not address the rare situation in which rate and capacity are equal.

Simple schemes such as "send the message 3 times and use at best 2 out of 3 voting scheme if the copies differ" are inefficient error-correction methods, unable to asymptotically guarantee that a block of data can be communicated free of error. Advanced techniques such as Reed-Solomon codes and, more recently, Turbo codes come much closer

to reaching the theoretical Shannon limit, but at a cost of high computational complexity. With Turbo codes and the computing power in today's digital signal processors, it is now possible to reach within 1/10 of one decibel of the Shannon limit.

Mathematical statement

Theorem (Shannon, 1948):

1. For every discrete memoryless channel, the channel capacity

$$C = \max_{P_X} I(X;Y)$$

has the following property. For any $\varepsilon > 0$ and $R < C$, for large enough N , there exists a code of length N and rate $\geq R$ and a decoding algorithm, such that the maximal probability of block error is $\leq \varepsilon$.

2. If a probability of bit error p_b is acceptable, rates up to $R(p_b)$ are achievable, where

$$R(p_b) = \frac{C}{1 - H_2(p_b)}.$$

and $H_2(p_b)$ is the *binary entropy function*

$$H_2(p_b) = -[p_b \log p_b + (1 - p_b) \log(1 - p_b)]$$

3. For any p_b , rates greater than $R(p_b)$ are not achievable.

(MacKay (2003), p. 162; cf Gallager (1968), ch.5; Cover and Thomas (1991), p. 198; Shannon (1948) thm. 11)

Outline of Proof

As with several other major results in information theory, the proof of the noisy channel coding theorem includes an achievability result and a matching converse result. These two components serve to bound, in this case, the set of possible rates at which one can communicate over a noisy channel, and matching serves to show that these bounds are tight bounds.

The following outlines are only one set of many different styles available for study in information theory texts.

Achievability for discrete memoryless channels

This particular proof of achievability follows the style of proofs that make use of the Asymptotic equipartition property(AEP). Another style can be found in information theory texts using Error Exponents.

Both types of proofs make use of a random coding argument where the codebook used across a channel is randomly constructed - this serves to reduce computational complexity while still proving the existence of a code satisfying a desired low probability of error at any data rate below the Channel capacity.

By an AEP-related argument, given a channel, length n strings of source symbols X_1^n , and length n strings of channel outputs Y_1^n , we can define a *jointly typical set* by the following:

$$A_\epsilon^{(n)} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$$

$$2^{-n(H(X)+\epsilon)} \leq p(X_1^n) \leq 2^{-n(H(X)-\epsilon)}$$

$$2^{-n(H(Y)+\epsilon)} \leq p(Y_1^n) \leq 2^{-n(H(Y)-\epsilon)}$$

$$2^{-n(H(X,Y)+\epsilon)} \leq p(X_1^n, Y_1^n) \leq 2^{-n(H(X,Y)-\epsilon)},$$

$$2^{-n(I(X;Y)+\epsilon)} \leq p(X_1^n, Y_1^n) \leq 2^{-n(I(X;Y)-\epsilon)}$$

We say that two sequences X_1^n và Y_1^n are *jointly typical* if they lie in the jointly typical set defined above.

Steps

1. In the style of the random coding argument, we randomly generate 2^{nR} codewords of length n from a probability distribution Q .
2. This code is revealed to the sender and receiver. It is also assumed both know the transition matrix $p(y|x)$ for the channel being used.
3. A message W is chosen according to the uniform distribution on the set of codewords. That is, $Pr(W=w) = 2^{-nR}$, $w = 1, 2, \dots, 2^{nR}$.
4. The message W is sent across the channel.
5. The receiver receives a sequence according to $P(y^n|x^n(w)) = \prod_{i=1}^n p(y_i|x_i(w))$
6. Sending these codewords across the channel, we receive Y_1^n , and decode to some source sequence if there exists exactly 1 codeword that is jointly typical with Y . If there are no jointly typical codewords, or if there are more than one, an error is declared. An error also occurs if a decoded codeword doesn't match the original codeword. This is called *typical set decoding*.

The probability of error of this scheme is divided into two parts:

1. First, error can occur if no jointly typical X sequences are found for a received Y sequence
2. Second, error can occur if an incorrect X sequence is jointly typical with a received Y sequence.
 - By the randomness of the code construction, we can assume that the average probability of error averaged over all codes does not depend on the index sent. Thus, without loss of generality, we can assume $W = 1$.
 - From the joint AEP, we know that the probability that no jointly typical X exists goes to 0 as n grows large. We can bound this error probability by ϵ .
 - Also from the joint AEP, we know the probability that a particular $X_1^{n(i)}$ and the Y_1^n resulting from $W = 1$ are jointly typical is $\leq 2^{-n(I(X;Y)-3\epsilon)}$.

Define: $E_i = \{(X_1^n(i), Y_1^n) \in A_\epsilon^{(n)}\}$, $i = 1, 2, \dots, 2^{nR}$

as the event that message i is jointly typical with the sequence received when message 1 is sent.

$$\begin{aligned} P(\text{error}) &= P(\text{error} | W = 1) \leq P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i) \\ &\leq \epsilon + 2^{-n(I(X;Y)-R-3\epsilon)} \end{aligned}$$

We can observe that as n goes to infinity, if $R < I(X;Y)$ for the channel, the probability of error will go to 0.

Finally, given that the average codebook is shown to be "good," we know that there exists a codebook whose performance is better than the average, and so satisfies our need for arbitrarily low error probability communicating across the noisy channel.

Converse for discrete memoryless channels

Suppose a code of 2^{nR} codewords. Let W be drawn uniformly over this set as an index. Let X^n and Y^n be the codewords and received codewords, respectively.

1. $nR = H(W) = H(W|Y^n) + I(W; Y^n)$ using identities involving entropy and mutual information
2. $\leq H(W|Y^n) + I(X^n(W); Y^n)$ since X is a function of W
3. $\leq 1 + P_e^{(n)}nR + I(X^n(W); Y^n)$ by the use of Fano's Inequality
4. $\leq 1 + P_e^{(n)}nR + nC$ by the fact that capacity is maximized mutual information.

The result of these steps is that $P_e^{(n)} \geq 1 - \frac{1}{nR} - \frac{C}{R}$. As the block length n goes to infinity, we obtain $P_e^{(n)}$ is bounded away from 0 if R is greater than C - we can only get arbitrarily low rates of error if R is less than C .

Channel coding theorem for non-stationary memoryless channels

We assume that the channel is memoryless, but its transition probabilities change with time, in a fashion known at the transmitter as well as the receiver.

Then the channel capacity is given by

$$C = \liminf \max_{p(X_1), p(X_2), \dots} \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i)$$

The maximum is attained at the capacity achieving distributions for each respective channel. That is,

$$C = \liminf \frac{1}{n} \sum_{i=1}^n C_i \text{ where } C_i \text{ is the capacity of the } i\text{th channel.}$$

Outline of the proof

The proof runs through in almost the same way as that of channel coding theorem. Achievability follows from random coding with each symbol chosen randomly from the capacity achieving distribution for that particular channel. Typicality arguments use the definition of typical sets for non-stationary sources defined in Asymptotic Equipartition Property.

The technicality of \liminf comes into play when $\frac{1}{n} \sum_{i=1}^n C_i$ does not converge.

References

- C. E. Shannon, The Mathematical Theory of Information. Urbana, IL:University of Illinois Press, 1949 (reprinted 1998).
- 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
- Thomas Cover, Joy Thomas, Elements of Information Theory. New York, NY:John Wiley & Sons, Inc., 1991. ISBN 0-471-06259-6

See also

- Error exponent
- Asymptotic equipartition property (AEP)
- Shannon-Hartley theorem
- Turbo code
- Fano's Inequality

External links

- On Shannon and Shannon's law (<http://www.iet.ntnu.no/projects/beats/Documents/LarsTelektronikk02.pdf>)
- On-line textbook: Information Theory, Inference, and Learning Algorithms (<http://www.inference.phy.cam.ac.uk/mackay/itila/>) , by David MacKay - gives an entertaining and thorough introduction to Shannon theory, including two proofs of the noisy-channel coding theorem. This text also discusses state-of-the-art methods from coding theory, such as low-density parity-check codes, and Turbo codes.

Retrieved from "http://en.wikipedia.org/wiki/Noisy_channel_coding_theorem"

Categories: Articles with sections needing expansion | Information theory | Mathematical theorems | Claude Shannon

- This page was last modified 04:51, 20 September 2006.
 - All text is available under the terms of the GNU Free Documentation License. (See **Copyrights** for details.)
- Wikipedia® is a registered trademark of the Wikimedia Foundation, Inc.