WIKIPEDIA

Noisy-channel coding theorem

In information theory, the **noisy-channel coding theorem** (sometimes **Shannon's theorem** or **Shannon's limit**), establishes that for any given degree of noise contamination of a communication channel, it is possible to communicate discrete data (digital information) nearly error-free up to a computable maximum rate through the channel. This result was presented by Claude Shannon in 1948 and was based in part on earlier work and ideas of Harry Nyquist and Ralph Hartley.

The **Shannon limit** or **Shannon capacity** of a communication channel refers to the maximum rate of error-free data that can theoretically be transferred over the channel if the link is subject to random data transmission errors, for a particular noise level. It was first described by Shannon (1948), and shortly after published in a book by Claude Elwood Shannon and Warren Weaver in 1949 entitled *The Mathematical Theory of Communication*. (ISBN 0252725484). This founded the modern discipline of information theory.

Contents

Overview

Stated by Claude Shannon in 1948, the theorem describes the maximum possible efficiency of error-correcting methods versus levels of noise interference and data corruption. Shannon's theorem has wide-ranging applications in both communications and data storage. This theorem is of foundational importance to the modern field of information theory. Shannon only gave an outline of the proof. The first rigorous proof for the discrete case is due to Amiel Feinstein[1] in 1954.

The Shannon theorem states that given a noisy channel with channel capacity C and information transmitted at a rate R, then if $R < C$ there exist codes that allow the probability of error at the receiver to be made arbitrarily small. This means that, theoretically, it is possible to transmit information nearly without error at any rate below a limiting rate, C.

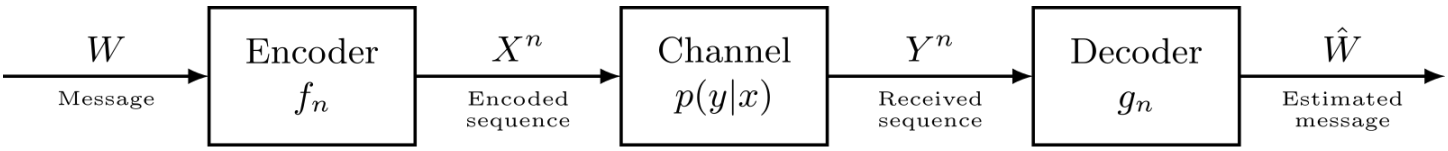
The converse is also important. If $R > C$, an arbitrarily small probability of error is not achievable. All codes will have a probability of error greater than a certain positive minimal level, and this level increases as the rate increases. So, information cannot be guaranteed to be transmitted reliably across a channel at rates beyond the channel capacity. The theorem does not address the rare situation in which rate and capacity are equal.

The channel capacity C can be calculated from the physical properties of a channel; for a band-limited channel with Gaussian noise, using the Shannon–Hartley theorem.

Simple schemes such as "send the message 3 times and use a 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, low-density parity-check (LDPC) codes and turbo codes, come much closer to reaching the theoretical Shannon limit, but at a cost of high computational complexity. Using these highly efficient codes and with the computing power in today's digital signal processors, it is now possible to reach very close to the Shannon limit. In fact, it was shown that LDPC codes can reach within 0.0045 dB of the Shannon limit (for binary Additive white Gaussian noise (AWGN) channels, with very long block lengths).[2]

Mathematical statement

The basic mathematical model for a communication system is the following:



A **message** W is transmitted through a noisy channel by using encoding and decoding functions. An **encoder** maps W into a pre-defined sequence of channel symbols of length n. In its most basic model, the channel distorts each of these symbols independently of the others. The output of the channel –the received sequence– is fed into a **decoder** which maps the sequence into an estimate of the message. In this setting, the probability of error is defined as:

$$P_e = \Pr\left\{\hat{W} \neq W\right\}.$$

Theorem (Shannon, 1948):

1. For every discrete memoryless channel, the channel capacity

   $$C = \sup_{p_X} I(X;Y)$$[3]

   has the following property. For any $\epsilon > 0$ 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 $\leq \epsilon$.

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*

   $$H_2(p_b) = -\left[p_b \log_2 p_b + (1-p_b)\log_2(1-p_b)\right]$$

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 make the analysis simpler 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:

$$\begin{aligned} A_\varepsilon^{(n)} = \{(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n \\ 2^{-n(H(X)+\varepsilon)} \leq p(X_1^n) \leq 2^{-n(H(X)-\varepsilon)} \\ 2^{-n(H(Y)+\varepsilon)} \leq p(Y_1^n) \leq 2^{-n(H(Y)-\varepsilon)} \\ 2^{-n(H(X,Y)+\varepsilon)} \leq p(X_1^n, Y_1^n) \leq 2^{-n(H(X,Y)-\varepsilon)}\} \end{aligned}$$

We say that two sequences X_1^n and 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 that one knows 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, $Pr(W = w) = 2^{-nR}, w = 1, 2, \ldots, 2^{nR}$.
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 $\leq 2^{-n(I(X;Y)-3\varepsilon)}$.

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

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

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

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.

Weak 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 transmitted 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 get arbitrarily low rates of error only if R is less than C.

Strong converse for discrete memoryless channels

A strong converse theorem, proven by Wolfowitz in 1957,[4] states that,

$$P_e \geq 1 - \frac{4A}{n(R-C)^2} - e^{-\frac{n(R-C)}{2}}$$

for some finite positive constant A. While the weak converse states that the error probability is bounded away from zero as n goes to infinity, the strong converse states that the error goes to 1. Thus, C is a sharp threshold between perfectly reliable and completely unreliable communication.

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

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

The maximum is attained at the capacity achieving distributions for each respective channel. That is, $C = \liminf \frac{1}{n} \sum_{i=1}^{n} C_i$ where C_i is the capacity of the i*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 the asymptotic equipartition property article.

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

See also

- Asymptotic equipartition property (AEP)
- Fano's inequality
- Rate–distortion theory
- Shannon's source coding theorem
- Shannon–Hartley theorem
- Turbo code

Notes

1. "A new basic theorem of information theory". Feinstein, Amiel. 1954. Bibcode:1955PhDT.....12F (https://ui.adsabs.harvard.edu/abs/1955PhDT.....12F). hdl:1721.1/4798 (https://hdl.handle.net/1721.1%2F4798).
2. Sae-Young Chung, G. David Forney, Jr., Thomas J. Richardson, and Rüdiger Urbanke, "On the Design of Low-Density Parity-Check Codes within 0.0045 dB of the Shannon Limit (http://www.josephboutros.org/ldpc_vs_turbo/ldpc_Chung_CLfeb01.pdf)", *IEEE Communications Letters*, 5: 58-60, Feb. 2001. ISSN 1089-7798
3. For a description of the "sup" function, see Supremum
4. Robert Gallager. *Information Theory and Reliable Communication.* New York: John Wiley & Sons, 1968. ISBN 0-471-29048-3

References

- Cover T. M., Thomas J. A., *Elements of Information Theory*, John Wiley & Sons, 1991. ISBN 0-471-06259-6
- Fano, R. A., *Transmission of information; a statistical theory of communications*, MIT Press, 1961. ISBN 0-262-06001-9
- Feinstein, Amiel, "A New basic theorem of information theory", *IEEE Transactions on Information Theory*, 4(4): 2-22, 1954.
- MacKay, David J. C., *Information Theory, Inference, and Learning Algorithms (http://www.inference.phy.cam.ac.uk/mackay/itila/book.html)*, Cambridge University Press, 2003. ISBN 0-521-64298-1 [free online]
- Shannon, C. E., *A Mathematical Theory of Communication* (https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6773024). *The Bell System Technical Journal* 27,3: 379–423, 1948.
- Shannon, C. E., *A Mathematical Theory of Communication* (http://cm.bell-labs.com/cm/ms/what/shannonday/paper.html) Urbana, IL: University of Illinois Press, 1948 (reprinted 1998).
- Wolfowitz, J., "The coding of messages subject to chance errors (https://projecteuclid.org/download/pdf_1/euclid.ijm/1255380682)", *Illinois J. Math.*, 1: 591–606, 1957.

External links

- On Shannon and Shannon's law (http://www.cs.miami.edu/home/burt/learning/Csc524.142/LarsTelektronikk02.pdf)
- Shannon's Noisy Channel Coding Theorem (http://cnx.org/content/m10180/latest/)

organization.