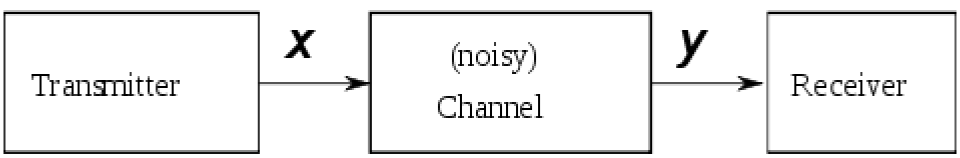
Channel Capacity:

In electrical engineering, computer science and information theory, channel capacity is the tightest upper bound on the rate of information that can be reliably transmitted over a communications channel. By the noisy-channel coding theorem, the channel capacity of a given channel is the limiting information rate (in units of information per unit time) that can be achieved with arbitrarily small error probability.

Information theory, developed by Claude E. Shannon during World War II, defines the notion of channel capacity and provides a mathematical model by which one can compute it. The key result states that the capacity of the channel, as defined above, is given by the maximum of the mutual information between the input and output of the channel, where the maximization is with respect to the input distribution.

Formal Definition:



Let X and Y be the random variables representing the input and output of the channel, respectively. Let $p_{Y|X}(y|x)$ be the conditional distribution function of Y given X, which is an inherent fixed property of the communications channel. Then the choice of the marginal distribution $p_X(x)$ completely determines the joint distribution $p_{X,Y}(x,y)$ due to the identity

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)\, p_X(x)$$

which, in turn, induces a mutual information $I(X;Y)$. The **channel capacity** is defined as

$$C = \sup_{p_X(x)} I(X;Y)$$

where the supremum is taken over all possible choices of $p_X(x)$.

Independent and identically distributed random variables:

In probability theory and statistics, a sequence or other collection of random variables is **independent and identically distributed** (**i.i.d.**) if each random variable has the same probability distribution as the others and all are mutually independent.

Shannon's source coding theorem

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.

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) informally states that:

"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 tends to infinity; but conversely, if they are compressed into fewer than $N\,H(X)$ bits it is virtually certain that information will be lost."

Entropy (information theory)

In information theory, **entropy** is a measure of the uncertainty in a random variable. In this context, the term usually refers to the **Shannon entropy**, which quantifies the expected value of the information contained in a message. Entropy is typically measured in bits, nats, or bans. Shannon entropy is the average unpredictability in a random variable, which is equivalent to its information content. Shannon entropy provides an absolute limit on the best possible lossless encoding or compression of any communication, assuming that the communication may be represented as a sequence of independent and identically distributed random variables.

A single toss of a fair coin has an entropy of one bit. A series of two fair coin tosses has an entropy of two bits. The number of fair coin tosses is its entropy in bits. This random selection between two outcomes in a sequence over time, whether the outcomes are equally probable or not, is often referred to as a Bernoulli process. The entropy of such a process is given by the binary entropy function. The entropy rate for a fair coin toss is one bit per toss. However, if the coin is not fair, then the uncertainty, and hence the entropy rate, is lower. This is because, if asked to predict the next outcome, we could choose the most frequent result and be right more often than wrong. The difference between what we know, or predict, and the information that the unfair coin toss reveals to us is less than one heads-or-tails "message", or bit, per toss.

Data compression

In computer science and information theory, **data compression**, **source coding**, or **bit-rate reduction** involves encoding information using fewer bits than the original representation. Compression can be either lossy or lossless. Lossless compression reduces bits by identifying and eliminating statistical redundancy. No information is lost in lossless compression. Lossy compression reduces bits by identifying unnecessary information and removing it. The process of reducing the size of a data file is popularly referred to as data compression, although its formal name is source coding (coding done at the source of the data before it is stored or transmitted).

Compression is useful because it helps reduce resources usage, such as data storage space or transmission capacity. Because compressed data must be decompressed to use, this extra processing imposes computational or other costs through decompression; this situation is far from being a free lunch. Data compression is subject to a space-time complexity trade-off. For instance, a compression scheme for video may require expensive hardware for the video to be decompressed fast enough to be viewed as it is being decompressed, and the option to decompress the video in full before watching it may be inconvenient or require additional storage. The design of data compression schemes involves trade-offs among various factors, including the degree of compression, the amount of distortion introduced (*e.g.*, when using lossy data compression), and the computational resources required to compress and uncompress the data.

Random variable

In probability and statistics, a **random variable** or **stochastic variable** is a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated probability.

Cumulative distribution function

In probability theory and statistics, the **cumulative distribution function** (**CDF**), or just **distribution function**, describes the probability that a real-valued random variable X with a given probability distribution will be found at a value less than or equal to *x*. In the case of a continuous distribution, it gives the area under the probability density function from minus infinity to *x*. Cumulative distribution functions are also used to specify the distribution of multivariate random variables.

Marginal distribution

In probability theory and statistics, the **marginal distribution** of a subset of a collection of random variables is the probability distribution of the variables contained in the subset. It gives the probabilities of various values of the variables in the subset without reference to the values of the other variables. This contrasts with a conditional distribution, which gives the probabilities contingent upon the values of the other variables.

Conditional probability distribution

In probability theory and statistics, given two jointly distributed random variables X and Y, the **conditional probability distribution** of Y given X is the probability distribution of Y when X is known to be a particular value; in some cases the conditional probabilities may be expressed as functions containing the unspecified value x of X as a parameter. The conditional distribution contrasts with the marginal distribution of a random variable, which is its distribution without reference to the value of the other variable.

Joint probability distribution

In the study of probability, given at least two random variables X, Y, ..., that are defined on a probability space, the **joint probability distribution** for X, Y, ... is a probability distribution that gives the probability that each of X, Y, ... falls in any particular range or discrete set of values specified for that variable. In the case of only two random variables, this is called a **bivariate distribution**, but the concept generalizes to any number of random variables, giving a **multivariate distribution**.

Mutual information

In probability theory and information theory, the **mutual information** (sometimes known by the archaic term **transinformation**) of two random variables is a quantity that measures the mutual dependence of the two random variables.