

In the first 19 pages of, "A Mathematical Theory of Communication," Shannon discusses a variety of topics ranging from encoding with stochastic process to redundancy. We will discuss some of the main points we learned in a progressive fashion. Firstly, the effect of noise in any channel and savings possible are due to statistical structure of original message. Messages are selected from a set of possible meanings which could be small or large. Logarithmic measure is more convenient for a variety of reasons. In general, a communication system consists of: information source, transmitter, channel, receiver, destination and there are three types mainly: discrete, continuous, mixed. The logarithm of the number of possible signals in a discrete channel increases linearly with time and capacity to transmit information can be specified by the bit-per-second rate of increase. For a discrete source, there can be representations stochastically and we can reduce information sent using logical encoding of predictable sequences as in you establish a stochastic model based on the probabilities of getting each symbol. You can use this method to generate English-like sentences with sequences of approximation steps. Moving up in order of approximation moves us closer to real English and the example provided seems extremely convincing. Non-discrete / Ergodic processes are similar but use frequencies and limits. You can measure how much information is "produced" by our Markoff process using entropy. With weighted values, you define conditional entropy as the average of the entropy of y for each value of x weighted on that particular x . The uncertainty / entropy of the joint event x, y is uncertainty of x plus uncertainty of y (given x is known). The ratio of entropy of a source to maximum value it could have while still restricted to the same symbols will be called its relative entropy. One minus the max compression possible when we encode into the same alphabet (again, relative entropy) is the redundancy. To exemplify this concept, the redundancy of English is about 50% meaning half is determined by the language structure and half is chosen freely. Redundancy also closely relates to crossword puzzles as large ones are possible with the 50% in English (redundancy too high is too constraining). One method of encoding is getting messages of length N in order of decreasing probability, divide the series into two groups of nearly equal probability as possible. Group 1 identification first binary bit is 0 and Group 2 is 1. The subsets are defined with the second binary digit. Goes until each subset contains one message. In practicality, we can practically make encoders with transducers that maximize the entropy in the channel and have the same statistical structure as the source.