Error Correcting Codes: Combinatorics, Algorithms and Applications (Fall 2007)

Lecture 8: Shannon's Noise Models

September 14, 2007

Lecturer: Atri Rudra *Scribe: Sandipan Kundu & Atri Rudra*

Till now we have concentrated on the worst case noise model pioneered by Hamming. In today's lecture, we will study some stochastic noise models (most of) which were first studied by Shannon.

1 Overview of Shannon's Result

Shannon introduced the notion of reliable communication over noisy channels. Broadly, there are two types of channels that were studied by Shannon:

- Noisy Channel: This type of channel introduces error in during transmission, which results in incorrect reception of the transmitted signal by the receiver. Redundancy is added at the transmitter to increase reliability of the transmitted data. The redundancy is taken off at the receiver. This process is termed as *Channel Coding*.
- Noise-free Channel: As the name suggests, this channel does not introduce any type of error in transmission. Redundancy in source data is being used for of the compression of the source data in the transmitter. The data is decompressed at the receiver. The process is popularly known as *Source Coding*.

The figure below presents a generic model of a communication system, which combines the two concepts we discussed above:

$$\text{Message} \rightarrow \boxed{\text{Source Encoder}} \rightarrow \boxed{\text{Channel Encoder}} \rightarrow \text{Channel} \rightarrow$$
$$\rightarrow \text{Channel} \rightarrow \boxed{\text{Channel Decoder}} \rightarrow \boxed{\text{Source Decoder}} \rightarrow \text{Decoded Message}$$

In the figure above, source coding and channel coding are coupled. However, Shannon's source coding theorem allows us to decouple both these parts of the communication and study each of these parts separately. Intuitively, this makes sense: if one can have reliable communication over the channel using channel coding, then for the source coding the channel effectively has no noise.

For source coding, Shannon proved a theorem that precisely calculated the amount by which the message can be compressed: this amount is related to the "entropy" of the message. We will however, not talk about source coding in any detail in this course. From now on, we will exclusively focus on the channel coding part of the communication setup. Note that one aspect of channel coding is how we model the channel noise. We have seen Hamming's worst case noise model in some detail. Next, we will study some specific stochastic channels.

2 Shannon's Noise Model

Shannon proposed a stochastic way of modeling noise. The symbols that are input to the channel are assumed to come from some *input alphabet* **X**, while the channel spits out symbols from its *output alphabet* **Y**. The following diagram shows this relationship:

$$\mathbf{X} \ni x \rightarrow \boxed{\text{channel}} \rightarrow y \in \mathbf{Y}$$

The channels considered by Shannon are also *memoryless*, that is, noise acts independently on each transmitted symbol. In this course, well only study *discrete* channels where both the alphabets **X** and **Y** are finite (today we will define one channel that is not discrete, though we will not study it in any detail later on).

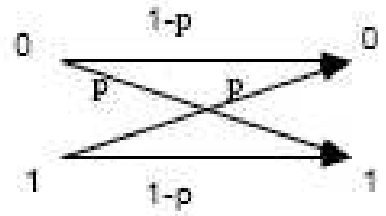
The final piece in specification of a channel is the *transition matrix* **M** that governs the process of how the channel introduces error. In particular, the channel is described in form of a matrix with entries as cross over probability over all combination of the input and output alphabets. For any pair $(x, y) \in \mathbf{X} \times \mathbf{Y}$, let $Pr(y|x)$ denote the probability that when x is input to the channel and y is output by the channel. Then the transition matrix is given by $\mathbf{M}(x, y) = Pr(y|x)$. Specific structure of the matrix is shown below.

$$\mathbf{M} = \begin{pmatrix} & \vdots & \\ \cdots & Pr(y|x) & \cdots \\ & \vdots & \end{pmatrix}$$

Next, we look at some specific instances of channels.

2.1 Binary Symmetric Channel (BSC)

Let $0 \leq p \leq \frac{1}{2}$. The Binary Symmetric Channel with *crossover probability* p or BSC_p is defined as follows. $\mathbf{X} = \mathbf{Y} = \{0, 1\}$. The 2×2 transition matrix can naturally be represented as a bipartite graph where the left vertices correspond to the rows and the right vertices correspond to the columns of the matrix, where $\mathbf{M}(x, y)$ is represented as the weight of the corresponding (x, y) edge. For BSC_p, the graph looks as follows:



Binary Symmetric Channel

In other words, every bit is flipped with probability p.

2.2 q-ary Symmetric Channel (qSC)

We now look at the generalization of BSC_p to alphabets of size $q \geq 2$. Let $0 \leq p \leq 1 - \frac{1}{q}$. The q-ary Symmetric Channel with crossover probability p, or qSC_p is defined as follows. $\mathbf{X} = \mathbf{Y} = [q]$. The transition matrix $\mathbf{M}$ for qSC_p is defined as follows.

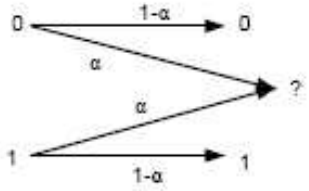
$$M(x,y) = \begin{cases} 1-p & \text{if } y = x \\ \frac{p}{q-1} & \text{if } y \neq x \end{cases}$$

In other words, every symbol is left untouched with probability $1 - p$ and is distorted to each of the $q - 1$ possible different symbols with equal probability.

2.3 Binary Erasure Channel (BEC)

In the previous two examples that we saw, $\mathbf{X} = \mathbf{Y}$. However this need not always be the case.

Let $0 \leq \alpha \leq 1$. The Binary Erasure Channel with *erasure probability* α is defined as follows. $\mathbf{X} = \{0, 1\}$ and $\mathbf{Y} = \{0, 1, ?\}$, where ? denotes an "erasure." The transition matrix is as follows:



Binary Erasure Channel

In the above any edge that is not present represents a transition that occurs with 0 probability. In other words, every bit in BEC_α is erased with probability α (and is left as is with probability $1 - \alpha$).

2.4 Binary Input Additive Gaussian White Noise Channel (BIAGWN)

We now look at a channel that is not discrete. Let $\sigma \geq 0$. The Binary Input Additive Gaussian White Noise Channel with standard deviation σ or $BIAGWN_\sigma$ is defined as follows. $\mathbf{X} = \{-1, 1\}$ and $\mathbf{Y} = \mathbb{R}$. The noise is modeled by continuous Gaussian probability distribution function. The Gaussian distribution has lots of nice properties and is a popular choice for modeling noise of

continuous nature. Given $(x, y) \in \{-1, 1\} \times \mathbb{R}$, the noise $y - x$ is distributed according to the Gaussian distribution of zero mean and standard deviation of σ. In other words,

$$Pr(y \mid x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot \exp\left(-\left(\frac{(y-x)^2}{2\sigma^2}\right)\right)$$

3 Error Correction in Stochastic Noise Models

We now need to revisit the notion of error correction. Note that unlike in Hamming's noise model, we cannot hope to *always* recover the transmitted codeword. As an example, in BSC_p there is always some positive probability that can a codeword can be distorted into another codeword during transmission. In such a scenario no decoding algorithm can hope to recover the transmitted codeword. Thus, in stochastic channels there is always will be some *decoding error probability* (where the randomness is from the channel noise). However, we would like this error probability to be small for every possible transmitted codeword. More precisely, for every message, we would like the decoding algorithm to recover the transmitted message with probability $1 - f(n)$, where $\lim_{n\to\infty} f(n) \to 0$, that is $f(n)$ is $o(1)$. Ideally, we would like to have $f(n) = 2^{-\Omega(n)}$.

3.1 Shannon's General Theorem

Recall that the big question that we are interested in this course is the tradeoff between the rate of the code and the fraction of errors that can be corrected. For stochastic noise models that we have seen, it is natural to think of the fraction of errors to be the parameter that governs the amount of error that is introduced by the channel. For example, for BSC_p, we will think of p as the fraction of errors.

Shannon's remarkable theorem on channel coding was to *precisely* identify when reliable transmission is possible over the stochastic noise models that he considered. In particular, for the general framework of noise models that he considered, Shannon defined the notion of *capacity*, which is a real number such that reliable communication is possible if and only if the rate is less than the capacity of the channel.

We are going to state (and prove) Shannon's general result for the special case of BSC_p. To state the result, we will need the following definition:

Definition 3.1 (q-ary Entropy Function)**.** *Let $q \geq 2$ be an integer and $0 \leq x \leq 1$ be a real. Then the* q-ary entropy function *is defined as follows:*

$$H_q(x) = x\log_q(q-1) - x\log_q(x) - (1-x)\log_q(1-x).$$

See Figure 1 for a pictorial representation of the $H_q(\cdot)$ for the first few values of q. For the special case of $q = 2$, we will drop the subscript from the entropy function and denote $H_2(x)$ by just $H(x)$, that is, $H(x) = -x\log x - (1-x)\log(1-x)$, where $\log x$ is defined as $\log_2(x)$ (we are going to follows this convention for the rest of the course).

We are now ready to state the theorem:

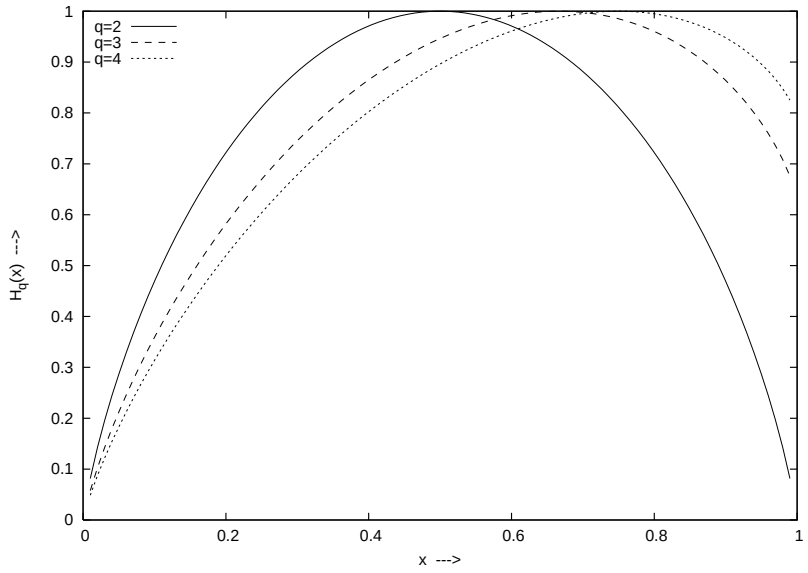


Figure 1: A plot of $H_q(x)$ for $q = 2, 3$ and 4. The maximum value of 1 is achieved at $x = 1 - 1/q$.

Theorem 3.2 (Shannon's Capacity for BSC)**.** *For reals $0 \leq p < \frac{1}{2}$ and $0 \leq \varepsilon \leq \frac{1}{2} - p$, the following hold for large enough n.*

1. *There exists a real $\delta > 0$, an encoding function $E : \{0,1\}^k \to \{0,1\}^n$ and a decoding function $D : \{0,1\}^n \to \{0,1\}^k$ where $k \leq \lfloor 1 - H(p+\varepsilon)n \rfloor$, such that the following holds for every $\mathbf{m} \in \{0,1\}^k$.*
$$\Pr_{\text{Noise } \mathbf{e} \text{ from } BSC}[D(E(\mathbf{m}) + \mathbf{e})) \neq \mathbf{m}] \leq 2^{-\delta n}.$$

2. *If $k \geq \lceil (1 - H(p) + \varepsilon)n \rceil$ then for every pair of encoding and decoding functions, $E : \{0,1\}^k \to \{0,1\{^n$ and $D : \{0,1\}^n \to \{0,1\}^k$, there exists $\mathbf{m} \in \{0,1\}^k$ such that*
$$\Pr_{\text{Noise } \mathbf{e} \text{ from } BSC}[D(E(\mathbf{m}) + \mathbf{e})) \neq \mathbf{m}] \geq \frac{1}{2}.$$

Remark 3.3. *Theorem 3.2 implies that the capacity of BSC_p is $1 - H(p)$. It can also been shown that the capacity of qSC_p and BEC_α are $1 - H_q(p)$ and $1 - \alpha$ respectively.*

The appearance of the entropy function in Theorem 3.2 might surprise the reader who has not seen the theorem before. Without going into the details of the proof for now we remark that the entropy function gives a very good estimate of the volume of a Hamming ball. In particular, recall that $B_q(\mathbf{y}, \rho n)$ is the Hamming Ball of radius ρn, that is, $B_q(\mathbf{y}, \rho n) = \{\mathbf{x} \in [q]^n | \Delta(\mathbf{x}, \mathbf{y}) \leq \rho n\}$. Let $Vol_q(\mathbf{y}, \rho_n) = |B_q(\mathbf{y}, \rho_n)|$ denote the *volume* of the Hamming ball of radius ρn. Note that since the volume of a Hamming ball is translation invariant, $Vol_q(\mathbf{y}, \rho n) = Vol_q(\mathbf{0}, \rho_n)$. We will need the following inequalities in the proof of Theorem 3.2.

Proposition 3.4. *Let* $q \geq 2, n \geq 1$ *be integers and let* $0 \leq \rho \leq 1 - \frac{1}{q}$ *be a real. Then the following inequalities hold:*

1. $Vol_q(\mathbf{0}, \rho n) \leq q^{nH_q(\rho)}$*; and*
2. $Vol_q(\mathbf{0}, \rho n) \geq q^{nH_q(\rho)-o(n)}$.

In the next lecture, we will see the proof of Proposition 3.4 as well as the proof of the "negative" part of Theorem 3.2.