6.2. Probability of Error

It is now convenient to derive a somewhat explicit formula for \( P(\varphi(Y) \neq M) \). We use the laws of total probability and substitution to write

\[
P(\varphi(Y) \neq M) = \sum_{i=1}^{N} P(\varphi(Y) \neq M|M = i)P(M = i)
\]

\[= \frac{1}{N} \sum_{i=1}^{N} P(\varphi(Y) \neq i|M = i)
\]

\[= \frac{1}{N} \sum_{i=1}^{N} P(Y \in \{y : \varphi(y) \neq i\}|M = i)
\]

\[= \frac{1}{N} \sum_{i=1}^{N} W_n(\{y : \varphi(y) \neq i\}|x_i), \text{ by (6.2)}. \tag{6.3}
\]

Although the reader may have reservations about having \( M \) uniformly distributed, in many cases, we will be able to obtain a bound on the terms in (6.3) where this bound does not depend on \( i \), say

\[
W_n(\{y : \varphi(y) \neq i\}|x_i) < \lambda, \quad i = 1, \ldots, N. \tag{6.4}
\]

If we can establish such a bound, then for any probability mass function \( q(i) \) on \( \{1, \ldots, N\} \), if we replace (6.1) with

\[
P(M = i, Y \in B) := q(i)W_n(B|x_i), \quad i = 1, \ldots, N, B \subset \mathcal{Y}^n,
\]

it will follow that

\[
P(\varphi(Y) \neq M) = \sum_{i=1}^{N} P(\varphi(Y) \neq M|M = i)P(M = i)
\]

\[= \sum_{i=1}^{N} P(\varphi(Y) \neq i|M = i)q(i)
\]
\[
= \sum_{i=1}^{N} W_n(\{y : \varphi(y) \neq i\}|x_i)q(i)
\]
\[
< \sum_{i=1}^{N} \lambda q(i) = \lambda.
\]

### 6.3. Introduction to Channel Capacity

Although we have not emphasized it, the error probability \(P(\varphi(Y) \neq M)\) depends on the codeword length \(n\).

**Example 6.1** (Repetition Codes on the BSC). Consider a BSC with crossover probability \(\varepsilon = 0.1\). Suppose we have \(N = 2\) messages, say \(\{0, 1\}\). For \(n = 1\), we might use codewords \(x_0 := 0\) and \(x_1 := 1\). In this case, the probability of error is just the crossover probability \(\varepsilon = 0.1\). For \(n = 3\), we might use codewords \(x_0 := (0, 0, 0)\) and \(x_1 := (1, 1, 1)\) with **majority rule decoding**; i.e., if the number of 1s in \(y\) is two or more, we decode \(\hat{i} = 1\), and zero otherwise. With this decoding rule, an error occurs if two or three of the transmitted bits are flipped by the channel. This probability is 0.028. If we proceed in a similar way with \(n = 5\) bits, the error probability is 0.0086. This analysis suggests that if we fix \(N = 2\) and let \(n \to \infty\), we can drive the error probability to zero. The downside is that the code rate \((\log N)/n\) also goes to zero.

The preceding example suggests that to get an interesting problem, we should allow \(N\) to depend on \(n\) and then ask how the error probability behaves as a function of \(n\). Ideally, we want the error probability to get smaller as \(n\) increases. We also want the code rate to stay away from zero. Amazingly, Shannon showed that there is a number \(C\) called the **channel capacity** such that if \((\log N)/n < C\), then there are codes that make the error probability tend to zero, and if \((\log N)/n > C\), then no code makes the error probability tend to zero. In some cases, it can be shown that if \((\log N)/n > C\), then the error probability actually tends to one [1, pp. 173–175].

To delve into some of these issues, we require more background.

---

*Let \(S_n\) denote the number of flips in \(n\) bits. An error occurs if \(S_n \geq n/2\). Hence, the error probability is \(1 - P(S_n < n/2)\). Since \(S_n\) is a binomial\((n, \varepsilon)\) random variable, the error probability is given by the MATLAB command \(1 - \text{binocdf(floor}(n/2), n, \varepsilon)\) when \(n\) is odd.*
6.4. Codebooks

6.4.1. New Codes from Old Codes by Throwing Away Codewords

Let \( f: \{1, \ldots, N\} \to X^n \) be any encoder, and let \( \varphi: Y^n \to \{1, \ldots, N\} \) be any decoder. Then

\[
D_i := \{ y : \varphi(y) = i \}
\]

contains exactly those outputs \( y \) that are decoded to message \( i \).

If \( G \) is any subset of \( \{1, \ldots, N\} \), we define the modified encoder \( f_G: G \to X^n \) by

\[
f_G(i) := f(i) = x_i, \quad i \in G,
\]

and we define the modified decoder \( \varphi_G: Y^n \to G \) by

\[
\varphi_G(y) := \begin{cases} 
\varphi(y), & y \in \bigcup_{i \in G} D_i, \\
 i_0, & \text{otherwise},
\end{cases}
\]

where \( i_0 \) can be any fixed element of \( G \). Hence, for \( i \in G \) with \( i \neq i_0 \), \( \varphi_G(y) = i \iff y \in D_i \), and we can write

\[
\{ y : \varphi_G(y) \neq i \} = \{ y : \varphi(y) \neq i \}, \quad \text{for } i \in G, i \neq i_0.
\]

(6.5)

Furthermore, if we put

\[
H := \left( \bigcup_{i \in G} D_i \right)^c,
\]

then \( \varphi_G(y) = i_0 \iff y \in D_{i_0} \cup H \). We can therefore write

\[
\{ y : \varphi_G(y) \neq i \} \subset D_i^c = \{ y : \varphi(y) \neq i \}, \quad \text{for all } i \in G.
\]

(6.6)

Suppose that instead of (6.4), we have only

\[
W_n(\{ y : \varphi(y) \neq i \}|x_i) < \lambda, \quad i \in G,
\]

where \( G \) is a proper subset of \( \{1, \ldots, N\} \). Let \( f_G \) and \( \varphi_G \) be the encoder and decoder just described. Then by (6.6), we can write

\[
W_n(\{ y : \varphi_G(y) \neq i \}|x_i) \leq W_n(\{ y : \varphi(y) \neq i \}|x_i) < \lambda, \quad i \in G.
\]
6.4.2. The Random Coding Argument

Let \( g \) be a nonnegative function defined on some set \( Z \), and let \( Z \) be a \( Z \)-valued random variable such that \( E[g(Z)] < \lambda \). Then there is at least one \( z \in Z \) with \( g(z) < \lambda \). To see this, suppose otherwise that \( g(z) \geq \lambda \) for all \( z \). Then we would have \( E[g(Z)] \geq \lambda \), which contradicts the hypothesis that \( E[g(Z)] < \lambda \).

References