A RANDOM CODING BOUND ON FIXED COMPOSITION CODES

6.441 Supplementary Notes 6, 5/5/92

We have derived a lower bound on error probability for fixed composition codes. We saw that at high rates and at the best composition, the exponential rate at which this bound approached zero, i.e., $\max_{\mathbf{Q}} E_{sp}(R,\mathbf{Q})$ was the same as $\max_{\mathbf{Q}} E_r(R,\mathbf{Q})$. In other words, for the best choice of $\mathbf{Q}$, and at high rates, the random coding bound is tight and the error probability, averaged over the ensemble of codes is substantially the same as the error probability of the best code. In this section, we carry out a random coding argument for an ensemble of fixed composition codes. We denote the exponential rate at which the error probability approaches zero as $E_{rf}(R,\mathbf{Q})$. We will see that for large R and arbitrary $\mathbf{Q}$, $E_{sp}(R,\mathbf{Q}) = E_{rf}(R,\mathbf{Q})$. In other words, even for non-optimal $\mathbf{Q}$, the average fixed composition codes is substantially as good as the best code of that composition. We interpret this after deriving the result.

For a given block length N, a given number of code words M, and a given composition $\mathbf{Q}$, we consider the ensemble of codes in which each code word $\mathbf{x}_m$ is independently chosen, with equal probability, to be any of the $T(\mathbf{Q})$ words of composition $\mathbf{Q}$. That is, $\Pr(\mathbf{x}_m=\mathbf{x}) = 1/|T(\mathbf{Q})|$ for all $\mathbf{x}\in T(\mathbf{Q})$. Let $P_{e,m}$ denote the average error probability (over this ensemble of codes) when message m enters the encoder and maximum likelihood decoding is used at the decoder. As in the random coding bound in the text, we first evaluate $\Pr(\text{error} \mid m, \mathbf{x}_m, \mathbf{y})$, the probability of an error conditional on message m entering the encoder, $\mathbf{x}_m$ being the first code word, and $\mathbf{y}$ being the received sequence. Thus, we are evaluating this probability over the random choice of $\mathbf{x}_{m'}$ for each $m'\neq m$ with the channel noise fixed.

For the given $\mathbf{x}_m$ and $\mathbf{y}$, let $\widehat{\mathbf{P}}$ be the noise composition such that $\mathbf{x}_m, \mathbf{y}$ has the joint composition $\mathbf{Q}\widehat{\mathbf{P}}$. Also, let ω be the output composition such that $\omega_j = \sum_k Q_k P(j|k)$ for each j, $0\leq j\leq J-1$. We recall that $\mathbf{P}(\mathbf{y}|\mathbf{x}_m) = \exp[N \sum_{k,j} Q_k \widehat{P}(j|k) \ln P(j|k)]$. An error occurs if, for some $m'\neq m$, $\mathbf{x}_{m'}$ is such that $P(\mathbf{y}|\mathbf{x}_{m'}) \geq P(\mathbf{y}|\mathbf{x}_m)$; we assume, in upper bounding error probability, that an error occurs when the maximum likelihood rule is ambiguous. Now define $T_{\mathbf{y}}(\mathbf{P}')$ as $\{\mathbf{x}: \mathbf{x},\mathbf{y} \in T(\mathbf{QP}')\}$. Thus, $T_{\mathbf{y}}(\mathbf{P}')$ is the set of input sequences that can be carried into $\mathbf{y}$ by the noise composition $\mathbf{P}'$. Notice that since $\mathbf{y}$ has composition ω, we must have

$$\sum_k Q_k P'(j|k) = \omega_j = \sum_k Q_k \widehat{P}(j|k) \quad ; \quad 0\leq j\leq J-1 \tag{38}$$

Note also that if $\mathbf{x}_{m'} \in T_{\mathbf{y}}(\mathbf{P}')$, then $\mathbf{P}(\mathbf{y}|\mathbf{x}_{m'}) = \exp[N \sum_{k,j} Q_k P'(j|k) \ln P(j|k)]$. Thus, an error occurs (conditional on message m, code word $\mathbf{x}_m$, received sequence $\mathbf{y}$) if there is some other code word $\mathbf{x}_{m'}$ in the set $T_{\mathbf{y}}(\mathbf{P}')$ for some $\mathbf{P}'$ such that

$$\exp[N \sum_{k,j} Q_k P'(j|k) \ln P(j|k)] \geq \exp[N \sum_{k,j} Q_k \hat{P}(j|k) \ln P(j|k)] \quad \text{and} \tag{39a}$$

$$\sum_k Q_k P'(j|k) = \sum_k Q_k \hat{P}(j|k) \; ; \; 0 \leq j \leq J-1 \tag{39b}$$

Define $P(\hat{\mathbf{P}})$ as the set of noise compositions $\mathbf{P}'$ such that (39) is satisfied. We then have

$$\Pr(\text{error} \mid m, \mathbf{x}_m, \mathbf{y}) \leq \sum_{m' \neq m} \sum_{\mathbf{P}' \in P(\hat{\mathbf{P}})} \Pr\left[\mathbf{x}_{m'} \in T_{\mathbf{y}}(\mathbf{P}')\right] \tag{40}$$

Note that for a given $m' \neq m$, $\Pr\left[\mathbf{x}_{m'} \in T_{\mathbf{y}}(\mathbf{P}')\right] = |T_{\mathbf{y}}(\mathbf{P}')| \,/\, |T(\mathbf{Q})|$. This is because $\mathbf{x}_{m'}$ is randomly selected (independent of $\mathbf{x}_m$) with equal probability over all sequences in $T(\mathbf{Q})$, and $|T_{\mathbf{y}}(\mathbf{P}')|$ is the number of sequences in $T_{\mathbf{y}}(\mathbf{P}')$; all of the sequences in $T_{\mathbf{y}}(\mathbf{P}')$ are in $T(\mathbf{Q})$ since $\mathbf{x},\mathbf{y} \in T(\mathbf{QP}')$. We also have $|T_{\mathbf{y}}(\mathbf{P}')| = |T(\mathbf{QP}')| \,/\, |T(\omega)|$ and thus

$$\Pr\left[\mathbf{x}_{m'} \in T_{\mathbf{y}}(\mathbf{P}')\right] = \frac{|T(\mathbf{QP}')|}{|T(\mathbf{Q})|\,|T(\omega)|} \leq \exp[-N I(\mathbf{Q};\mathbf{P}')] \tag{41}$$

Lemma 7: For $\mathbf{y} \in T_{\mathbf{x}}(\hat{\mathbf{P}})$,

$$\Pr(\text{error} \mid m, \mathbf{x}_m, \mathbf{y}) \leq \exp\left\{-N \min_{\mathbf{P}' \in P(\hat{\mathbf{P}})} \left[I(\mathbf{Q};\mathbf{P}') - R - \delta_N \right]^+ \right\} \tag{42}$$

where $[z]^+$ means $\max(0, z)$ and $\delta_n \leq (KJ-1)\,[\ln(N+1)]/N$

Proof: There are fewer than $(N+1)^{KJ-1}$ joint compositions $\mathbf{QP}'$ and thus fewer than this many compositions $\mathbf{P}'$ in $P(\hat{\mathbf{P}})$. Since there are M-1 choices of $m' \neq m$, we use (40) to get

$$\Pr(\text{error} \mid m, \mathbf{x}_m, \mathbf{y}) \leq (M-1)(N+1)^{KJ-1} \max_{\mathbf{P}' \in P(\hat{\mathbf{P}})} \Pr\left[\mathbf{x}_{m'} \in T_{\mathbf{y}}(\mathbf{P}')\right]$$

Substituting (41) into this, using $M-1 < e^{NR}$, we get

$$\Pr(\text{error} \mid m, \mathbf{x}_m, \mathbf{y}) \leq \max_{\mathbf{P}' \in P(\hat{\mathbf{P}})} \exp -N[I(\mathbf{Q};\mathbf{P}') - R - \delta_N]$$

where δ_N satisfies the bound in the lemma. Taking the maximization inside, and recognizing that a probability must be at most 1, we have (42).

Lemma 8:

$$P_{e,m} \le \exp\left\{-N \min_{\hat{\mathbf{P}}} \min_{\mathbf{P}' \in P(\hat{\mathbf{P}})} D(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q}) + \left[I(\mathbf{Q};\mathbf{P}') - R - 2\delta_N \right]^+ \right\} \tag{43}$$

Recall from (18) that $\mathbf{P}[\mathbf{y} \in T_{\mathbf{x}_m}(\hat{\mathbf{P}}) \mid \mathbf{x}_m] \le \exp\{-ND(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q})\}$. We then have

$$P_{e,m} \le \sum_{\hat{\mathbf{P}}} \Pr\left[\text{error} \mid m, \mathbf{x}_m, \mathbf{y} \in T_{\mathbf{x}_m}(\hat{\mathbf{P}})\right] \exp\left[-ND(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q})\right]$$

Substituting (42) into this and upper bounding each term in the sum by the maximum, we get (43).

Lemma 9:

$$P_{e,m} \le \exp\left\{-N \min_{\hat{\mathbf{P}}} D(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q}) + \left[I(\mathbf{Q};\hat{\mathbf{P}}) - R - 2\delta_N \right]^+ \right\} \tag{44}$$

Proof: Note that if $I(\mathbf{Q};\mathbf{P}') \ge I(\mathbf{Q};\hat{\mathbf{P}})$, then

$$D(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q}) + [I(\mathbf{Q};\mathbf{P}') - R]^+ \ge D(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q}) + [I(\mathbf{Q};\hat{\mathbf{P}}) - R]^+ \tag{45}$$

We shall show that in the alternative case, $I(\mathbf{Q};\mathbf{P}') < I(\mathbf{Q};\hat{\mathbf{P}})$,

$$D(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q}) + [I(\mathbf{Q};\mathbf{P}') - R]^+ \ge D(\mathbf{P}'||\mathbf{P}|\mathbf{Q}) + [I(\mathbf{Q};\mathbf{P}') - R]^+ \tag{46}$$

This will complete the proof since it will show that all terms in the minimization of (43) are lower bounded by terms in the minimization of (44). To demonstrate (46), recall that $\sum_k Q_k P'(j|k) = \omega_j = \sum_k Q_k \hat{P}(j|k)$. Thus $I(\mathbf{Q};\mathbf{P}') < I(\mathbf{Q};\hat{\mathbf{P}})$ implies that $\sum_{k,j} Q_k P'(j|k) \ln P'(j|k) < \sum_{k,j} Q_k \hat{P}(j|k) \ln \hat{P}(j|k)$. Also, since $\mathbf{P}' \in P(\hat{\mathbf{P}})$, $\sum_{k,j} Q_k P'(j|k) \ln P(j|k) \ge \sum_{k,j} Q_k \hat{P}(j|k) \ln P(j|k)$. Combining these two inequalities shows that $D(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q}) \ge D(\mathbf{P}'||\mathbf{P}|\mathbf{Q})$ and (46) follows.

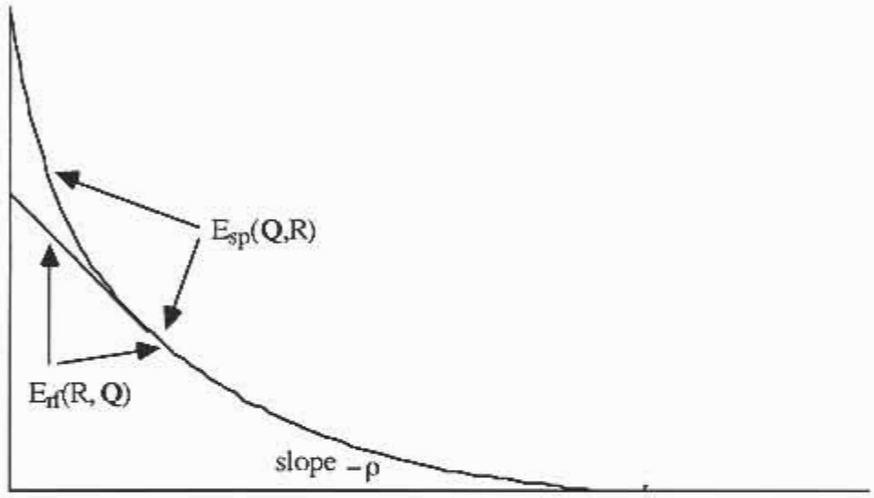
THEOREM 5: Over the ensemble of codes of fixed composition **Q** with block length N and rate R, the average error probability for each message m using max likelihood decoding satisfies

$$P_{e,m} \le \exp\{-N[E_{rl}(R,\mathbf{Q}) - 2\delta_N]\} \text{ where} \tag{47}$$

$$E_{rl}(R,\mathbf{Q}) = \max_{0 \le \rho \le 1} [F(\rho,\mathbf{Q}) - \rho R]; \quad F(\rho,\mathbf{Q}) = \min_{\hat{\mathbf{P}}} [D(\hat{\mathbf{P}}||\mathbf{P}|\mathbf{Q}) + \rho I(\mathbf{Q};\hat{\mathbf{P}})] \tag{48}$$

Proof: Note that $[I(\mathbf{Q};\widehat{\mathbf{P}})-R]^+ \geq \rho I(\mathbf{Q};\widehat{\mathbf{P}})-\rho R$ for all ρ, $0\leq\rho\leq 1$. Substituting this into (46) yields (47) and (48).

Recalling that $E_{sp}(R,\mathbf{Q}) = \max_{\rho\geq 0} [F(\rho,\mathbf{Q})-\rho R]$, we see that $E_{sp}(R,\mathbf{Q}) = E_{rf}(R,\mathbf{Q})$ for all R such that the maximizing ρ in $\max_{\rho\geq 0} [F(\rho,\mathbf{Q})-\rho R]$ is at most 1. Recalling that this maximizing ρ is the negative of the slope of $E_{sp}(R,\mathbf{Q})$ in R, we see that the relationship is that given in the figure below.



We have already seen (from Eq. (31)) that $E_r(R,\mathbf{Q})$ is strictly less than $E_{sp}(R,\mathbf{Q})$ for all rates less than $I(\mathbf{Q};\mathbf{P})$ and for all but the optimizing $\mathbf{Q}$. Thus, $E_r(R,\mathbf{Q}) < E_{rf}(R,\mathbf{Q})$ for all but the optimizing $\mathbf{Q}$. This is not the result of any weakness in the derivation of $E_r(R,\mathbf{Q})$, but rather is that the fixed composition codes (for non-optimum $\mathbf{Q}$) are better (as an ensemble) than the ensemble with independently chosen letters. The reason for this is that in the random coding bound with independently chosen letters, typical code words have a range of different compositions, most of them close to composition $\mathbf{Q}$. After optimizing over $\mathbf{Q}$, the error probability of code words of compositions close to $\mathbf{Q}$ is about the same as those of composition $\mathbf{Q}$ (because $\mathbf{Q}$ is at a stationary point). For non-optimal $\mathbf{Q}$, however, the code words of poorer composition than $\mathbf{Q}$ have markedly poorer error probability than those of composition $\mathbf{Q}$, leading to larger error probability over all.

6) ADAPTIVE DECODING FOR UNKNOWN CHANNELS

Suppose that we know that a channel is a discrete memoryless channel, but we do not know the channel transition matrix $\mathbf{P}$. We can still choose a random code of fixed composition $\mathbf{Q}$, but we can not implement a max likelihood decoder. As an alternative (albeit not very

practical), we can consider a decoder which, when presented with $\mathbf{y}$, calculates the empirical noise distribution between $\mathbf{x}_m$ and $\mathbf{y}$ for each code word $\mathbf{x}_m$ in the code. Denote this distribution as $\hat{\mathbf{P}}_{\mathbf{x}_m}$. Thus $\hat{\mathbf{P}}_{\mathbf{x}_m}$ is that noise composition $\hat{\mathbf{P}}$ such that $\mathbf{x}_m\mathbf{y} \in T(\mathbf{Q}\hat{\mathbf{P}})$. We then decode that m for which $\hat{\mathbf{P}}_{\mathbf{x}_m}(\mathbf{y}|\mathbf{x}_m)$ is maximal. What is very surprising is that the random coding exponent using this decoding rule is equal to $E_{rf}(R,\mathbf{Q})$. In other words, the fact that the decoder does not know what the channel is does not materially effect the error probability over that when the decoder can use that information. To derive this result, we again look at Pr(error | m,$\mathbf{x}_m\mathbf{y}$) for $\mathbf{x}_m\mathbf{y} \in T(\mathbf{Q}\hat{\mathbf{P}})$. We calculate this for the channel $\mathbf{P}$, but using the adaptive decoder above which does not know $\mathbf{P}$. An error occurs if there is some $\mathbf{x}_{m'} \in T_{\mathbf{y}}(\mathbf{P}')$ for some $\mathbf{P}'$ such that

$$\exp[N \sum_{k,j} Q_k P'(j|k) \ln P'(j|k)] \geq \exp[N \sum_{k,j} Q_k \hat{P}(j|k) \ln \hat{P}(j|k)] \quad \text{and} \tag{49a}$$

$$\sum_k Q_k P'(j|k) = \sum_k Q_k \hat{P}(j|k) \; ; \; 0 \leq j \leq J-1 \tag{49b}$$

Define $P^*(\hat{\mathbf{P}})$ as the set of noise compositions $\mathbf{P}'$ such that (39) is satisfied. We then have

$$\Pr(\text{error} \mid m, \mathbf{x}_m, \mathbf{y}) \leq \sum_{m' \neq m} \sum_{\mathbf{P}' \in P^*(\hat{\mathbf{P}})} \Pr\left[\mathbf{x}_{m'} \in T_{\mathbf{y}}(\mathbf{P}')\right] \tag{50}$$

Using the same arguments as in lemmas 7 and 8,

$$P_{e,m} \leq \exp\left\{-N \min_{\hat{\mathbf{P}}} \min_{\mathbf{P}' \in P^*(\hat{\mathbf{P}})} D(\hat{\mathbf{P}}|\mathbf{P}|\mathbf{Q}) + \left[I(\mathbf{Q};\mathbf{P}') - R - 2\delta_N \right]^+ \right\} \tag{51}$$

From (49), it follows that $I(\mathbf{Q};\mathbf{P}') \geq I(\mathbf{Q};\hat{\mathbf{P}})$ for all $\mathbf{P}' \in P^*(\hat{\mathbf{P}})$, and thus

$$P_{e,m} \leq \exp\left\{-N \min_{\hat{\mathbf{P}}} D(\hat{\mathbf{P}}|\mathbf{P}|\mathbf{Q}) + \left[I(\mathbf{Q};\hat{\mathbf{P}}) - R - 2\delta_N \right]^+ \right\} \tag{52}$$

From this it follows that $P_{e,m} \leq \exp\{-N[E_{rf}(R,\mathbf{Q})-2\delta_N\}$ as claimed.