Class 9 Exercises

CS250/EE387, Winter 2025

Today we'll show (from scratch, without information theory) that the capacity of the binary erasure channel (with erasure probability p) is $1-p$: that is, there exist codes of rate $1-p-\varepsilon$ that can communicate reliably over BEC_p, but any code of rate $1-p+\varepsilon$ cannot.

1. Suppose that G is a random matrix in $\mathbb{F}_2^{n\times k}$. Show that the probability that G has rank less than k is at most 2^{k-n}.

   Solution

   The probability that G is not full rank is the same as the probability that there is some x so that $Gx = 0$. For a fixed nonzero $x \in \mathbb{F}_2^k$, Gx is uniformly distributed in $\mathbb{F}_2^n$, and so $\mathbb{P}\{Gx = 0\} = 2^{-n}$. Now union bounding over all nonzero $x \in \mathbb{F}_2^k$, we see that $\mathbb{P}\{\operatorname{rank} G < k\} \le 2^{k-n}$.

2. Consider the linear code given by the encoding map $x \mapsto Gx$. Define a decoder $D : \{0,1,\perp\}^n \to \{0,1\}^k \cup$ `FAIL` for this code to be the decoding algorithm that, on input $y \in \{0,1,\perp\}^n$, returns x if Gx agrees with y on all of the un-erased symbols if such an x exists and is unique; otherwise it returns `FAIL`.

   Let $\varepsilon > 0$ and suppose that $k \le (1-p-\varepsilon)n$. Let $G \in \mathbb{F}_2^{n\times k}$ be a random binary matrix. In the following parts, you will show that, for all $x \in \mathbb{F}_2^k$, there is some constant γ so that

   $$\mathbb{E}_G\left[\mathbb{P}_{BEC_p}\{D(BEC_p(Gx)) \neq x \mid G\}\right] \le 2^{-\gamma n},$$

   where the randomness in the $\mathbb{E}$ is over the choice of the matrix G, and the randomness in the $\mathbb{P}$ is over the channel BEC_p.

   (a) Let $J \subset [n]$ be the set of erased positions. Explain why the decoder succeeds if and only if the rank of $G|_{\bar{J}}$ is k, where $G|_{\bar{J}}$ is the restriction of G to the rows *not* in J.

   (b) Explain why

   $$\mathbb{P}\{\operatorname{rank}(G|_{\bar{J}}) < k\} \le \min\{1, 2^{k-(n-|J|)}\},$$

   and conclude that, for any $x \in \mathbb{F}_2^k$,

   $$\mathbb{P}_G\{D(BEC_p(Gx)) \neq x|J\} \le \min\{1, 2^{k-(n-|J|)}\},$$

   where above the conditional probability means "the probability that the decoder fails, given that J is the set of erased positions."

   (c) We are later going to want to take the expectation over J (that is, over the set of positions erased by the BSC_p). When we do that in the conclusion above, we get:

   $$\mathbb{E}_J\mathbb{P}_G\{D(BEC_p(Gx)) \neq x|J\} \le \mathbb{E}_J \min\{1, 2^{k-(n-|J|)}\},$$

   so we want to bound the right hand side, which we'll do in this part.

   Show that

   $$\mathbb{E}_J \min\left\{1, 2^{k+|J|-n}\right\} \le \mathbb{P}_J\{|J| \ge n(p+\varepsilon/2)\} + 2^{-n\varepsilon/2}.$$

   (Remember that we are assuming that $k \le n(1-p-\varepsilon)$).

(d) If you have not seen Chernoff bounds much before, you may use the fact that

$$\mathbb{P}\{|J| \geq n(p+\varepsilon/2)\} \leq 2^{-Cn\varepsilon^2}$$

for some constant C. (If you have seen Chernoff bounds before, convince yourself that this follows from a Chernoff bound). Either way, nothing to write down for this part, just be aware of this fact.

(e) Put it all together to prove the thing that this part asks you to prove, that for any $x \in \mathbb{F}_2^k$,

$$\mathbb{E}_G\left\{\mathbb{P}_{BEC_p}\left\{D(BEC_p(Gx)) \neq x \,|\, G\right\}\right\} \leq 2^{-\gamma n}$$

for some constant γ. (Note that γ is allowed to depend on ε and on the choice of the constant C). Hint: If it's been a while since probability, remember that for two random variables X, Y and some event $\mathcal{E}$ that depends on X and Y,

$$\mathbb{E}_Y[\mathbb{P}_X\{\mathcal{E} \,|\, Y\}] = \mathbb{E}_X[\mathbb{P}_Y\{\mathcal{E} \,|\, X\}].$$

Solution

(a) Let J denote the set of indices erased by the channel. The decoder succeeds if and only if the rank of $G|_{\bar{J}}$ is k, where $G|_{\bar{J}}$ denotes the restriction of G to the non-erased rows, because if that matrix has full rank we can solve it to find x; if it doesn't have full rank, then it has a kernel of dimension at least 1, and hence there is an affine subspace of dimension at least 1 of x's that are consistent with $BEC_p(x)$, in which case the decoder returns `FAIL`.

(b) For a fixed J,

$$\mathbb{P}_G\{D(BEC_p(Gx)) \neq x|J\} = \mathbb{P}\{\text{rank}\,(G|_{\bar{J}}) < k\} \leq \min\left\{1, 2^{k-(n-|J|)}\right\},$$

where we used part (a) for the first equality and problem 1 for the inequality.

(c) We have

$$\begin{aligned}\mathbb{E}_J \min\left\{1, 2^{k+|J|-n}\right\} &\leq 1 \cdot \mathbb{P}\{|J| \geq n(p+\varepsilon/2)\} + 2^{k+n(p+\varepsilon/2)-n} \cdot 1 \\ &\leq \mathbb{P}\{|J| \geq n(p+\varepsilon/2)\} + 2^{-n\varepsilon/2},\end{aligned}$$

using the asumption that $p \leq (1-p-\varepsilon)n$ in the final line.

(d) I have seen Chernoff bounds before, and I can convince myself that

$$\mathbb{P}\{|J| \geq n(p+\varepsilon/2)\} \leq 2^{-Cn\varepsilon^2}$$

for some constant C.

(e) Putting it all together, we have

$$\mathbb{E}_G\mathbb{P}_{BEC}\{D(BEC_p(Gx)) \neq x \,|\, G\} = \mathbb{E}_J\mathbb{P}_G\{D(BEC_p(Gx) \neq x|BEC_p \text{ erased } J\} \leq 2^{-Cn\varepsilon^2} + 2^{-n\varepsilon/2}.$$

For an appropriate choice of γ (which depends on ε and the constant C), this is at most $2^{-\gamma n}$ for large enough n.

3. In this part, we will show that:

Claim 1. *For all $\varepsilon > 0$, there is a code of rate at least $1-p-\varepsilon$ so that the error probability on BEC_p is at most $2^{-\gamma n}$, for some γ (which depends on p and ε).*

(a) Explain why this is not obvious from part 2. That is, part 2 says that:

For all $\varepsilon > 0$, if G is a random generator matrix for a code of rate about $1 - p - \varepsilon$, then for all x,

$$\mathbb{E}_G[\mathbb{P}_{BEC_p}\{D(BEC_p(Gx)) \neq x \,|\, G\}] \leq 2^{-\gamma n}$$

for some constant γ.

My (incorrect) claim is that we are done! If the expected failure probability (that is, the expectation of $\mathbb{P}_{BEC_p}\{D(BEC_p(Gx)) \neq x\}$, which is what we are looking at) is less than $2^{-\gamma n}$, then there must exist a code that has failure probability less than $2^{-\gamma n}$. But that's exactly what Claim 1 says!

What's wrong with this reasoning?

Solution

The guarantee from part 2 implies that *for some fixed x,* there is a G that's good for x (in the sense that the decoder is likely to work when Gx is transmitted), but G might depend on x. What we want to show is that there's some G that's simultaneously good for *all* x's. To conclude something about the failure probability, we'd actually want part 2 to say something like:

For all $\varepsilon > 0$, if G is a random generator matrix for a code of rate about $1 - p - \varepsilon$, then

$$\mathbb{E}_G[\max_x \mathbb{P}_{BEC_p}\{D(BEC_p(Gx)) \neq x \,|\, G\}] \leq 2^{-\gamma n}$$

for some constant γ.

The difference is that the "for all x" on the outside of the expectation turned into a "$\max_x$" on the inside.

(b) We will show Claim 1 by "throwing out" some bad codewords of our random code, and showing that this new (slightly smaller) code will satisfy the statement that we want. Towards that end, imagine choosing $x \in \mathbb{F}_2^k$ uniformly at random. Explain why

$$\mathbb{E}_G\mathbb{E}_x\mathbb{P}_{BEC_p}\{D(BEC_p(Gx) \neq x) \,|\, G, x\} \leq 2^{-\gamma n},$$

and conclude that there exists some generator matrix G^* so that

$$\mathbb{E}_x\mathbb{P}_{BEC_p}\{D(BEC_p(G^*x) \neq x) \,|\, x\} \leq 2^{-\gamma n}.$$

Solution

The first equation follows by taking the expectation over x and using part 2, and then switching $\mathbb{E}_G$ and $\mathbb{E}_x$. The second equation is the probabilistic method: because the expected value of that quantity is small, there must be some G^* that has a value at least that small.

(c) Show that there is some set $\mathcal{X} \subseteq \mathbb{F}_2^k$ so that $|\mathcal{X}| \geq 2^{k-1}$ and that for all $x \in \mathcal{X}$,

$$\mathbb{P}_{BEC_p}\{D(BEC_p(G^*x)) \neq x\} \leq 2 \cdot 2^{-\gamma n}.$$

(Hint: Markov's inequality says that, for any non-negative random variable Z, $\mathbb{P}[Z \geq 2\mathbb{E}[Z]] \leq \frac{1}{2}$.)

Solution

As per the hint, we use Markov's inequality on the previous part, to conclude that

$$\mathbb{P}_x\left\{\mathbb{P}_{BEC_p}\{D(BEC_p(G^*x)) \neq x\} \geq 2 \cdot 2^{-\gamma n}\right\} \leq \frac{1}{2}.$$

Thus, at least half of the x's in $\mathbb{F}_2^k$ have the property that

$$\mathbb{P}_{BEC_p}\{D(BEC_p(G^*x)) \neq x\} \leq 2 \cdot 2^{-\gamma n}.$$

Let $\mathcal{X}$ be that collection of x's.

(d) Explain why the above establishes Claim 1, and thus proves that the capacity of the BEC_p is at least $1-p$.

Solution

Let $\mathcal{X}$ and G^* be the things that we just showed existed. Then consider the code $\mathcal{C} = \{G^*x : x \in \mathcal{X}\}$. We've just seen that for all $x \in \mathcal{X}$,

$$\mathbb{P}_{BEC_p}\{D(BEC_p(G^*x)) \neq x\} \geq 2 \cdot 2^{-\gamma n}.$$

This is the definition of failure probability, and we've proved Claim 1.
From the definition of capacity, Claim 1 implies that the capacity is *at least* $1-p$.

Aside: Should we be worried about the fact that the G^* that we get might have rank less than k? If we are drawing G's uniformly at random, then some of them might have low rank...but then the rate of our code might actually be worse than $1-p-\varepsilon$! In fact, we don't need to worry about this. The reason is that we just proved that the codewords G^*x (for all $x \in \mathcal{X}$) are not "confusable," in the sense that we'll probably be able to tell them apart. In particular, they are all distinct. So the size of this code is indeed $|\mathcal{X}| = 2^{k-1}$, and the rate really is at least $\frac{k-1}{n} \approx 1-p-\varepsilon$.

4. **(Bonus, if you have extra time)** So far we've seen that the capacity of the BEC_p is at least $1-p$. Now we'll show that it's *exactly* $1-p$.

Claim 2. *For any code with rate at least $1-p+\varepsilon$, the error probability on BEC_p must be at least $1/2$.*

Prove Claim 2. Notice that we must prove the result for any code (not necessarily linear) and for any decoding map (not necessarily the one proposed above).

Hint: try the following steps. Let $\mathcal{C}$ be a binary code with encoding map $E : \{0,1\}^k \to \{0,1\}^n$, so that $k/n \geq 1-p+\varepsilon$. Let $D : \{0,1,\perp\} \to \{0,1\}^k$ be any decoding map for $\mathcal{C}$.

For ease of notation, let $\hat{x} = D(BEC_p(E(x)))$ be the guess at x that the decoder recovers.

(a) Let J again denote the set of indices erased by the BEC. Explain why

$$\mathbb{P}\{|J| < n(p-\varepsilon/2)\} \leq 2^{-Cn\varepsilon^2}$$

for some constant C.

Solution

It's a Chernoff bound, just like before!

(b) Choose an x at random. Explain why

$$\mathbb{E}_x\mathbb{P}_{BEC_p}\{\hat{x} \neq x\} \geq (1-2^{-Cn\varepsilon^2})\mathbb{E}_J\left[\mathbb{P}_x\{\hat{x} \neq x \mid |J| \leq n(p-\varepsilon/2)\}\right]$$

Solution

First, we switch the order of the expectations:

$$\mathbb{E}_x\mathbb{P}_J\{\hat{x} \neq x\} = \mathbb{E}_J\mathbb{P}_x\{\hat{x} \neq x\}$$

Now, we consider two cases. The first is when $|J| \geq n(p-\varepsilon/2)$ (call this event "B") and the second is when it's smaller (call this "$\neg B$"). For any quantity A that's always at most 1, we have

$$\mathbb{E}[A] = \mathbb{E}[A|B]\mathbb{P}[B] + \mathbb{E}[A|\neg B]\mathbb{P}[\neg B] \geq \mathbb{E}[A|B]\mathbb{P}[B],$$

so we plug in the probability $\mathbb{P}[B] \geq 1 - 2^{-C\varepsilon^2 n}$ from the previous part, and we plug in $\mathbb{P}_x[\hat{x} \neq x]$ for A.

(c) Fix any set J of size at least $n(p-\varepsilon/2)$. Show that, if the BEC deletes the set J, then

$$\mathbb{P}_x\{\hat{x} = x\} \leq \frac{1}{2^k} \cdot 2^{n(1-p+\varepsilon/2)}.$$

(Notice that the probability here is over x. Since we are fixing J, the behavior of the BEC is fixed).

(Hint: What does the matrix $G|_{\bar{J}}$ look like? Does it have a kernel?)

Solution

The matrix $G|_{\bar{J}}$ is short and fat: it has $k = (1-p+\varepsilon)n$ columns and only $|\bar{J}| \leq n(1-p+\varepsilon/2)$ rows. That means that it has a kernel of dimension at least $k - n(1-p+\varepsilon/2)$. This means that given $G|_{\bar{J}}x$, there are at least $2^{k-n(1-p+\varepsilon/2)}$ possible $\hat{x}$'s consistent with it. If we choose x at random, the probability that it is the one that the decoder chooses to return is thus at most 1 over that, which is $2^{n(1-p+\varepsilon/2)-k}$, as desired.

Another way to see this: We have

$$\begin{aligned}
\mathbb{P}_x\{\hat{x} = x\} &= \frac{1}{2^k} \sum_{x \in \{0,1\}^k} \sum_{y \in \{0,1\}^J} \mathbf{1}_{E(x)|_{\bar{J}}=y} \mathbf{1}_{D(y)=x} \\
&= \frac{1}{2^k} \sum_{y \in \{0,1\}^J} \left(\sum_{x \in \{0,1\}^k} \mathbf{1}_{E(x)|_{\bar{J}}=y} \mathbf{1}_{D(y)=x} \right) \\
&\leq \frac{1}{2^k} \sum_{y \in \{0,1\}^J} (1) \\
&\leq 2^{-k} 2^{n(1-p+\varepsilon/2)}.
\end{aligned}$$

Above, we have replaced that sum with 1 because for each y, there is at most one x so that $D(y) = x$. In the last line we have used the fact that $|J| \geq n(p-\varepsilon/2)$.

(d) Prove Claim 2.

Solution

By the above, we have, for any J with $|J| \geq n(p-\varepsilon/2)$,

$$\mathbb{P}_x\{\hat{x} = x\} \leq 2^{-k+n(1-p+\varepsilon/2)} \leq 2^{-n\varepsilon/2}$$

using our assumption on k. Finally, by part (b) we have

$$\begin{aligned}
\mathbb{E}_x \mathbb{P}_{BEC_p}\{\hat{x} \neq x\} &\geq (1 - 2^{-Cn\varepsilon^2}) \mathbb{E}_J\{\mathbb{P}_x\{\hat{x} \neq x \,|\, |J| \geq n(p-\varepsilon/2)\}\} \\
&\geq (1 - 2^{-Cn\varepsilon^2})(1 - 2^{-n\varepsilon/2})
\end{aligned}$$

which is larger than $1/2$ for large enough n.