EE 231A
Information Theory
Instructor: Rick Wesel

Handout #9, Problem Set 4 Solutions
Wednesday, July 27, 2017

110 pts
Reading: Chapter 7 of *Elements of Information Theory*

Lectures 7: Definition and Computation of Channel Capacity

1. (9 pts) *An additive noise channel.*

$$Y = X + Z \qquad X \in \{0, 1\}, \quad Z \in \{0, a\} \tag{1}$$

We have to distinguish various cases depending on the values of a.

$a = 0$ In this case, $Y = X$, and $\max I(X;Y) = \max H(X) = 1$. Hence the capacity is 1 bit per transmission.

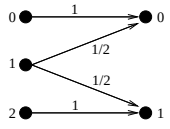
$a \neq 0, \pm 1$ In this case, Y has four possible values $0, 1, a$ and $1 + a$. Knowing Y, we know the X which was sent, and hence $H(X|Y) = 0$. Hence $\max I(X;Y) = \max H(X) = 1$, achieved for an uniform distribution on the input X.

$a = 1$ In this case Y has three possible output values, $0, 1$ and 2, and the channel is identical to the binary erasure channel discussed in class, with $a = 1/2$. As derived in class, the capacity of this channel is $1 - a = 1/2$ bit per transmission.

$a = -1$ This is similar to the case when $a = 1$ and the capacity here is also $1/2$ bit per transmission.

2. (8 pts) *Inverse Erasure Channel.*

Find the capacity of the inverse erasure channel shown below.



With input X and output Y we have

$$\begin{aligned} I(X;Y) &= H(Y) - H(Y|X) && (2)\\ &\leq H(Y) && (3)\\ &\leq 1\,, && (4) \end{aligned}$$

and $I(X;Y) = 1$ is achievable with $P(X = 0) = P(X = 2) = 0.5$. Thus $C = 1$.

3. (10 pts) *Cyclic Symmetry*

In class we defined a channel to have cyclic symmetry if the mutual information is invariant to cyclic shifts of the input distribution. This problem gives a matrix form of this definition.

Consider the channel transition matrix P where the entry in the xth row and the yth column denotes the conditional probability $p(y|x)$.

Matrix Conditions for Cyclic Symmetry The channel described by P has cyclic symmetry if the following two conditions are satisfied:

- All the rows of P are permutations of each other.
- The set C of all columns of P can be separated into a certain collection of mutually exclusive, collectively exhaustive subsets S_i. (i.e. The equations below are satisfied:)

$$\bigcup_i S_i = C \tag{5}$$

$$\text{if } S_i \neq S_j \text{ then } S_i \bigcap S_j = \emptyset\,. \tag{6}$$

(Note that the elements of both C and S_i are columns.) Furthermore, each subset S_i may be completely constructed from any one element (i.e. any one column) of S_i as follows: S_i contains exactly one instance of each cyclic shift of that element, and nothing else.

(a) (1 pt) Give P for the binary erasure channel.

$$P_{\text{BEC}} = \begin{bmatrix} 1-\alpha & \alpha & 0 \\ 0 & \alpha & 1-\alpha \end{bmatrix} \tag{7}$$

(b) (2 pts) For the binary erasure channel, decompose the columns of P into the subsets S_i described above.

$$S_1 = \left\{ \begin{bmatrix} \alpha \\ \alpha \end{bmatrix} \right\}, \qquad S_2 = \left\{ \begin{bmatrix} 1-\alpha \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1-\alpha \end{bmatrix} \right\} \tag{8}$$

(c) (4 pts) Prove that the **Matrix conditions for Cyclic Symmetry** described above are indeed sufficient for a channel to have cyclic symmetry. While the first two parts were specific to the binary erasure channel, this proof needs to work for *any* discrete memoryless channel.

We will use the decomposition $I(X;Y) = H(Y) - H(Y|X)$. The proof uses the two conditions to show that $H(Y|X)$ and $H(Y)$ are invariant to cyclic shifts of the input distribution.

- *First we use the first condition to show that $H(Y|X)$ is invariant to cyclic shifts of the input distribution.*

$$H(Y|X) = \sum_x p(x)H(Y|X = x) \tag{9}$$

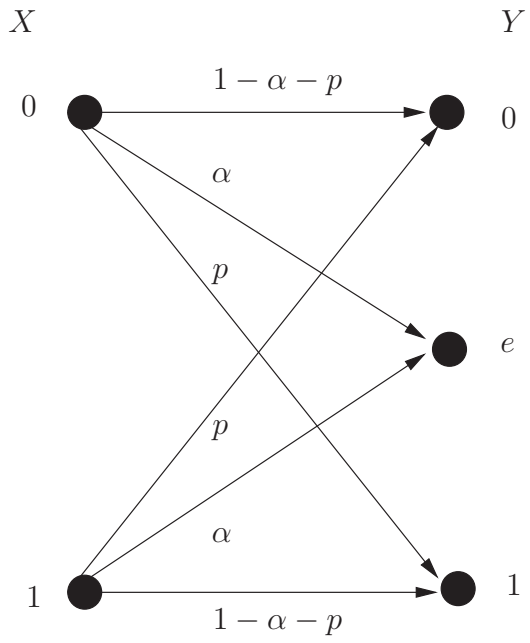
  The first condition, that all the rows of P are permutations of each other, forces $H(Y|X = x)$ to be the same for all x. This means that $H(Y|X)$ is the same regardless of the input distribution. In particular, it is invariant under cyclic shifts of the input distribution.

- *Now we use the second condition to show that $H(Y)$ is invariant to cyclic shifts in the input distribution. The inner product of the yth column of P with the input distribution computes $p(Y = y)$. from the input distribution. Each of the subsets S_i contain a set of cyclic shifts of columns such that a cyclic shift of the input distibution won't change the set of probabilities $p(Y = y)$ induced by any set S_i. Thus the set of probabilities in the output distribution is invariant to cylic shifts in the input distribution. Thus $H(Y)$ is invariant to shifts in the input distribution*

Since $H(Y|X)$ and $H(Y)$ are invariant to cyclic shifts of the input distribution, $I(X;Y)$ is invariant to cyclic shifts in the input distribution, and the channel has cyclic symmetry.

4. (15 pts) *Errors, Erasures, and Symmetry.*

Consider a binary channel that has *both* erasures and symmetric bit errors. This channel is illustrated below.



(a) (3 pts) Write down the transition matrix for this channel.
Solution:

$$\begin{bmatrix} 1-\alpha-p & \alpha & p \\ p & \alpha & 1-\alpha-p \end{bmatrix} \tag{10}$$

(b) (3 pts) Does this channel satisfy the matrix conditions for cyclic symmetry? Explain your answer.
Solution: Yes. Both of the matrix conditions for cyclic symmetry are met. The two rows are permutations of each other. The set of columns can be separated into two subsets that are collectively exhaustive, mutually exclusive, and such that each subset is exactly all cyclic shifts of any one element in the subset.. The two subsets are as follows:

$$S_1 = \left\{ \begin{bmatrix} 1-\alpha-p \\ p \end{bmatrix}, \begin{bmatrix} p \\ 1-\alpha-p \end{bmatrix} \right\}, S_2 = \left\{ \begin{bmatrix} \alpha \\ \alpha \end{bmatrix} \right\}, \tag{11}$$

(c) (3 pts) Does this channel satisfy the matrix conditions for weak symmetry? Explain your answer.
Solution: No. For weak symmetry we need all of the column sums to be equal. They are not in general equal because, in general, $2\alpha \neq 1-\alpha$. I accepted as correct the one answer that said that the channel *was* weakly symmetric as long as $\alpha = 1/3$, since for that one value of α, the column sums are indeed equal.

(d) (3 pts) Find the capacity of this errors and erasures channel. Feel free to use the notation $h_3(p_1, p_2, p_3)$ for the entropy of a discrete random variable that takes on three values with probabilities p_1, p_2, p_3.
Solution: There are many correct expressions and many ways to get them. However, the easiest approach is to use the fact shown above that the channel has cyclic symmetry and evaluate the mutual information for a uniform input. This leads to the following:

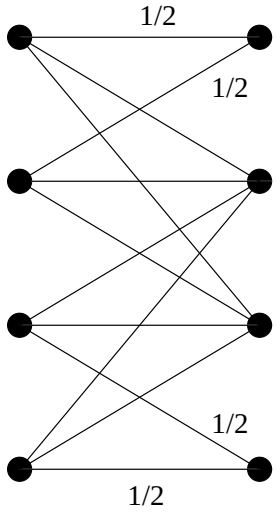
$$C = H(Y) - H(Y|X) \quad \text{(for X uniform)} \tag{12}$$

$$= h_3\left(\frac{1-\alpha}{2}, \alpha, \frac{1-\alpha}{2}\right) - h_3\left(1-\alpha-p, \alpha, p\right) \tag{13}$$

(e) (3 pts) Show how your capacity expression simplifies to other channel capacities that we have studied when α or p are set to zero.
Solution: Setting $\alpha = 0$ makes this a BSC. Substituting $\alpha = 0$ into (**??**) yields $C = 1 - h(p)$ as expected. Setting $p = 0$ makes this an erasure channel. Substituting $p = 0$ into (**??**) yields $C = 1 - \alpha$ as expected.

5. (9 pts) *Symmetric Channel?*

Consider the channel illustrated below where all branches have probability 1/2.



(a) (4 pts) State the conditions for a channel to be weakly symmetric. Write down the matrix of conditional probabilities for this channel. Does this channel meet the conditions to be weakly symmetric?

A channel is weakly symmetric if the rows of the matrix of conditional probabilities are permutations of each other and if the columns of that matrix sum to a constant. The matrix of conditional probabilities is

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} . \tag{14}$$

This matrix satisfies both conditions, and thus the channel is weakly symmetric.

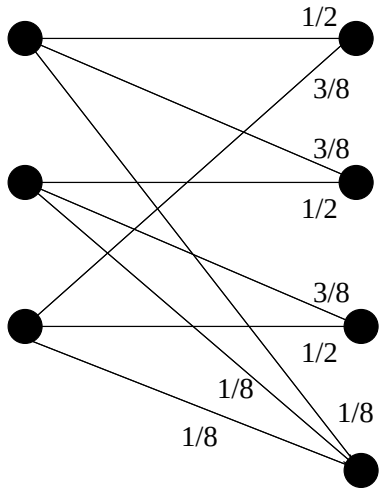
(b) (3 pts) Is the mutual information of this channel invariant to shifts in the input probability distribution? If yes, give a brief justification. If no, give a counterexample.

No, the mutual information is not invariant to shifts in the input probability distribution. For example, $I(X;Y) = 0$ *for input probabilities* $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \end{bmatrix}$ *while* $I(X;Y) = \frac{1}{2}$ *for input probabilities* $\begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$.

(c) (3 pts) What is the capacity of this channel? *Since the channel is weakly symmetric,* $C = 2 - H(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = 0.5$

6. (9 pts) *Symmetric Channel? (Part II)*

Consider the channel illustrated below where branches have probabilities as labeled.



(a) (3 pts) State the conditions for a channel to be weakly symmetric. Write down the matrix of conditional probabilities for this channel. Does this channel meet the conditions to be weakly symmetric? If so compute the capacity using weak symmetry. *A channel is weakly symmetric if the rows of the matrix of conditional probabilities are permutations of each other and if the columns of that matrix sum to a constant. The matrix of conditional probabilities is*

$$\begin{bmatrix} \frac{1}{2} & \frac{3}{8} & 0 & \frac{1}{8} \\ 0 & \frac{1}{2} & \frac{3}{8} & \frac{1}{8} \\ \frac{3}{8} & 0 & \frac{1}{2} & \frac{1}{8} \end{bmatrix} . \tag{15}$$

Since the column sums are not all the same. this matrix does not satisfy both conditions. The channel is not weakly symmetric.

(b) (3 pts) Is the mutual information of this channel invariant to cyclic shifts in the input probability distribution? If yes, give a brief justification and compute the capacity. If no, give a counterexample.

Yes the channel is invariant to cyclic shifts. We can decompose it into the appropriate sets S_i as follows:

$$S_1 = \left\{ \begin{bmatrix} 1/8 \\ 1/8 \\ 1/8 \end{bmatrix} \right\}, \qquad S_2 = \left\{ \begin{bmatrix} 1/2 \\ 0 \\ 3/8 \end{bmatrix}, \begin{bmatrix} 3/8 \\ 1/2 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 3/8 \\ 1/2 \end{bmatrix} \right\}. \tag{16}$$

The figure below illustrates the three-fold symmetry.Because of the cyclic symmetry, the capacity may be computed as the mutual information achieved by a uniform input distribution. Hence

$$\begin{aligned} C &= I_{\text{uniform}}(X;Y) && (17)\\ &= H(Y) - H(Y|X) && (18)\\ &= H\left(\frac{7}{24}, \frac{7}{24}, \frac{7}{24}, \frac{1}{8}\right) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right). && (19) \end{aligned}$$

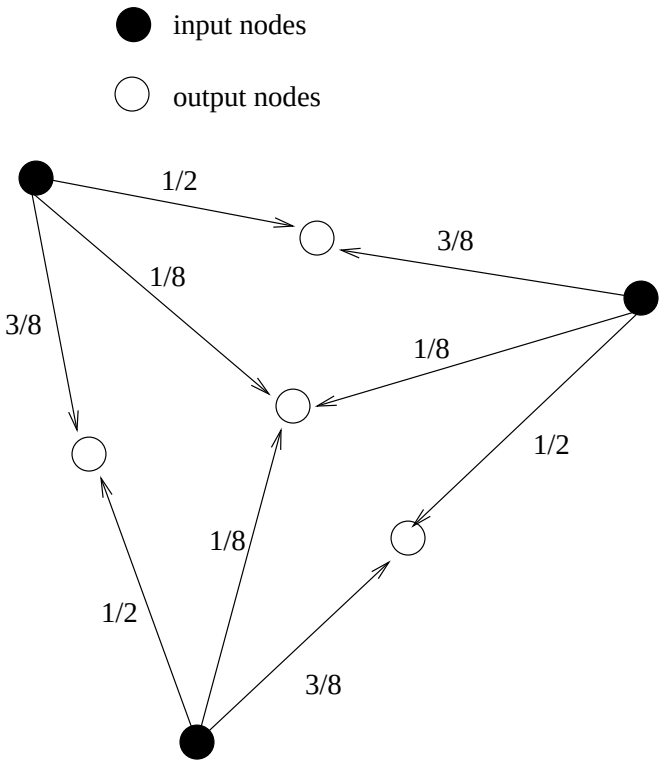


Figure 1: Since the channel has three-fold symmetry corresponding to the three inputs, cyclic shifts of the input distribution don't change the mutual information.

(c) (3 pts) Compute the capacity using the technique of finding an upper bound and then achieving it. Show that this capacity is the same as what you computed earlier. (The grouping axiom will be useful here.)

$$
\begin{aligned}
I(X;Y) &= H(Y) - H(Y|X) && (20)\\
&= H(Y) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right) && (21)\\
&\leq H(Y,E) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right) && (22)\\
&= H(E) + H(Y|E) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right) && (23)\\
&\leq H\left(\frac{1}{8}\right) + \frac{7}{8}H\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right) && (24)\\
&= H\left(\frac{7}{24}, \frac{7}{24}, \frac{7}{24}, \frac{1}{8}\right) - H\left(\frac{1}{2}, \frac{3}{8}, \frac{1}{8}\right) \text{ (grouping axiom)}, && (25)
\end{aligned}
$$

which is achieved by a uniform distribution as shown in part b.

Lectures 8: Proof of Achievability of Channel Capacity

7. (15 pts) *Joint Typicality.*

   (a) From the definition of joint typicality, this limit goes to 1.

   (b) Again, from the basic properties of joint typicality:

$$Pr\{(x_1^n, y_2^n) \in A_\epsilon^{(n)}\} \leq 2^{-n(I(X;Y)-3\epsilon)}\,. \tag{26}$$

   (c) For a BSC with $p = 0.031130$, $I(X;Y) = 1 - H(p) = 0.8$. With $n = 100$ and $\epsilon = 0.1$,

$$2^{-n(I(X;Y)-3\epsilon)} = 2^{-50}. \tag{27}$$

   (d) As long as $I(X;Y) - 3\epsilon > 0$

$$Pr\{(x_1^n, y_2^n) \in A_\epsilon^{(n)}\} \to 0 \tag{28}$$

   as $n \to \infty$.

   (e) $Pr\{(x_1^n, y_2^n) \in A_\epsilon^{(n)}\}$ is the probability that y^n is jointly typical with a particular x^n other than the transmitted codeword. When that x^n is another valid codeword, this causes a decoding error.

8. (10 pts) *Triple Typicality.*

(a)
$$P(A_\epsilon^{(n)}) \longrightarrow 1 \tag{29}$$

By the weak law of large numbers, the probability of satisfying any one of the seven conditions above converges to 1. Thus the probability of violating any one particular condition converges to zero. By the union bound, the probability that any condition is violated also converges to zero. Thus the probability that all conditions are satisfied converges to one.

(b) From the third condition, $P(z^n) \leq 2^{-n(H(Z)-\epsilon)}$. From the fourth condition, $P(x^n, y^n) \leq 2^{-n(H(X,Y)-\epsilon)}$. From the last condition, $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y,Z)+\epsilon)}$

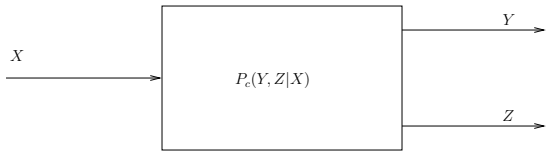
$$\sum_{x^n,y^n,z^n \in A_\epsilon^{(n)}} P(x^n, y^n)P(z^n) \leq |A_\epsilon^{(n)}| \times 2^{-n(H(X,Y)-\epsilon)} \times 2^{-n(H(Z)-\epsilon)} \tag{30}$$

$$\leq 2^{n(H(X,Y,Z)+\epsilon)} \times 2^{-n(H(X,Y)-\epsilon)} \times 2^{-n(H(Z)-\epsilon)} \tag{31}$$

$$\leq 2^{-n(I(X,Y;Z)-3\epsilon)} \tag{32}$$

9. (10 pts) *Joint Typicality on the Three-Way Channel.*

Consider a discrete memoryless channel with one input X, but *two* outputs, Y, and Z, that both contain information about X and are in general dependent. This channel is illustrated below.



(a) Find (and demonstrate) an upper bound on the probability that X^n, Y^n, Z^n is a member of the triple typical set if X^n and (Y^n, Z^n) are chosen independently, X^n according to $P(X)$ and (Y^n, Z^n) according to $P(Y, Z)$. Your upper bound will involve a term ϵ that converges to zero as $n \longrightarrow \infty$

Recall that the triple typical set is defined as follows:

The triple sequence (x^n, y^n, z^n) is a member of the triple typical set $A_\epsilon^{(n)}$ for the joint distribution $p(x, y, z)$ if the following inequalities are satisfied:

$$\left| -\frac{1}{n} \log p(x^n) - H(X) \right| < \epsilon \tag{33}$$

$$\left| -\frac{1}{n} \log p(y^n) - H(Y) \right| < \epsilon \tag{34}$$

$$\left| -\frac{1}{n} \log p(z^n) - H(Z) \right| < \epsilon \tag{35}$$

$$\left| -\frac{1}{n} \log p(x^n, y^n) - H(X, Y) \right| < \epsilon \tag{36}$$

$$\left| -\frac{1}{n} \log p(x^n, z^n) - H(X, Z) \right| < \epsilon \tag{37}$$

$$\left| -\frac{1}{n} \log p(y^n, z^n) - H(Y, Z) \right| < \epsilon \tag{38}$$

$$\left| -\frac{1}{n} \log p(x^n, y^n, z^n) - H(X, Y, Z) \right| < \epsilon \tag{39}$$

The marginal distributions above are those implied by the joint distribution $p(x, y, z)$.

From the first condition , $P(x^n) \leq 2^{-n(H(X)-\epsilon)}$. From the sixth condition, $P(y^n, z^n) \leq 2^{-n(H(Y,Z)-\epsilon)}$. From the last condition, $|A_\epsilon^{(n)}| \leq 2^{n(H(X,Y,Z)+\epsilon)}$

$$\sum_{x^n, y^n, z^n \in A_\epsilon^{(n)}} P(x^n)P(y^n, z^n) \leq |A_\epsilon^{(n)}| \times 2^{-n(H(X)-\epsilon)} \times 2^{-n(H(Y,Z)-\epsilon)} \tag{40}$$

$$\leq 2^{n(H(X,Y,Z)+\epsilon)} \times 2^{-n(H(X)-\epsilon)} \times 2^{-n(H(Y,Z)-\epsilon)} \tag{41}$$

$$\leq 2^{-n(I(X,:Y,Z)-3\epsilon)} \tag{42}$$

(b) Prove via a random coding argument that with input distribution $P_i(X)$ we can achieve the rate $I(X;Y,Z)$ where the joint distribution used to compute the mutual information is $P(X,Y,Z) = P_i(X)P_c(Y,Z|X)$. In your proof of achievability, you need only show that the average probability of error converges to zero.

Randomly select the code $\mathcal{C}$ by choosing each symbol in each codeword i.i.d. $\sim P(X)$. The decoded message $\hat{w}$ is selected using typical set decoding based on the received values of Y^n, Z^n.

$$\text{Average error probability} = E_{P(\mathcal{C})}\left[\frac{1}{2^{nR}}\sum_{i=1}^{2^{nR}} P(\hat{w} \neq i | w = i, \mathcal{C} = q)\right] \tag{43}$$

$$= \sum_{q\in\mathcal{C}} P(q)\frac{1}{2^{nR}}\sum_{i=1}^{2^{nR}} P(\hat{w} \neq i | w = i, \mathcal{C} = q) \tag{44}$$

$$= \frac{1}{2^{nR}}\sum_{i=1}^{2^{nR}}\sum_{q\in\mathcal{C}} P(q)P(\hat{w} \neq i | w = i, \mathcal{C} = q) \tag{45}$$

$$= \sum_{q\in\mathcal{C}} P(q)P(\hat{w} \neq 1 | w = 1, \mathcal{C} = q) \tag{46}$$

$$= P(E_1^c \cup E_2 \cup \cdots \cup E_{2^{nR}}) \tag{47}$$

$$\text{where } E_i = \{(X^n(w_i), Y^n, Z^n) \in A_\epsilon^{(n)}\} \tag{48}$$

$$\leq P(E_1^c) + \sum_{i=2}^{2^{nR}} P(E_i) \tag{49}$$

$$\leq \epsilon + \sum_{i=2}^{2^{nR}} 2^{-n(I(X,:Y,Z)-3\epsilon)} \tag{50}$$

$$\leq \epsilon + 2^{-n(I(X,:Y,Z)-R-3\epsilon)} \tag{51}$$

which converges to zero as $n \longrightarrow \infty$ since $\epsilon \longrightarrow 0$ as $n \longrightarrow \infty$.

10. (10 pts) *All codes are good.*
Prove Markov's inequality:

$$P(X \geq \delta) \leq \frac{EX}{\delta}\,, \tag{52}$$

and use it to show that the channel coding theorem we proved in class implies the following:

For a fixed target probability of error, the probability that a randomly selected code will exceed that target goes to zero as the blocklenth goes to infinity.

Solution:
Proof of Markov's inequality:
For X a positive random variable and $\delta > 0$,

$$EX = \int x f(x) dx \tag{53}$$

$$\geq \int \delta I(x \geq \delta) f(x) dx \tag{54}$$

$$= \delta P(X \geq \delta) \tag{55}$$

Now fix a target error probability δ. As long as the code rate is selected below capacity, we know that the expected probability of codeword error goes to zero as the blocklength goes to infinity. Hence we have:

$$P(X \geq \delta) \leq \frac{EX}{\delta} \tag{56}$$

$$\longrightarrow 0 \tag{57}$$