2.1. Prove Fano’s inequality.

2.3. Prove the properties of jointly typical sequences with \( \delta(\epsilon) \) constants explicitly specified.

2.4. Inequalities. Label each of the following statements with \( = \), \( \leq \), or \( \geq \). Justify each answer.

(a) \( H(X|Z) \) vs. \( H(X|Y) + H(Y|Z) \).
(b) \( I(X^2;Y^2) \) vs. \( I(X_1;Y_1) + I(X_2;Y_2) \), if \( p(y^2|x^2) = p(y_1|x_1)p(y_2|x_2) \).
(c) \( I(X^2;Y^2) \) vs. \( I(X_1;Y_1) + I(X_2;Y_2) \), if \( p(x^2) = p(x_1)p(x_2) \).

2.14. Variations on the joint typicality lemma. Let \( (X,Y,Z) \sim p(x,y,z) \) and \( 0 < \epsilon' < \epsilon \). Prove the following statements.

(a) Let \( (X^n,Y^n) \sim \prod_{i=1}^{n} p_{X,Y}(x_i,y_i) \) and \( \hat{Z}^n \mid \{ X^n = x^n, Y^n = y^n \} \sim \prod_{i=1}^{n} p_{Z|X}(\hat{z}_i|x_i) \), conditionally independent of \( Y^n \) given \( X^n \). Then
\[
P\left\{ (X^n,Y^n,\hat{Z}^n) \in \mathcal{T}_{\epsilon'}(X,Y,Z) \right\} \leq 2^{-n I(Y;Z|X)}.
\]

(b) Let \( (x^n,y^n) \in \mathcal{T}_{\epsilon'}(X,Y) \) and \( \hat{Z}^n \sim \text{Unif}(\mathcal{T}_{\epsilon'}(Z|x^n)) \). Then
\[
P\left\{ (x^n,y^n,\hat{Z}^n) \in \mathcal{T}_{\epsilon'}(X,Y,Z) \right\} \leq 2^{-n I(Y;Z|X)}.
\]

(c) Let \( x^n \in \mathcal{T}_{\epsilon'}(X) \), \( \hat{y}^n \) be an arbitrary sequence, and \( \hat{z}^n \sim p(\hat{z}^n|x^n) \), where
\[
p(\hat{z}^n|x^n) = \begin{cases} \prod_{i=1}^{n} p_{Z|X}(\hat{z}_i|x_i) & \text{if } \hat{z}^n \in \mathcal{T}_{\epsilon'}(Z|x^n), \\ 0 & \text{otherwise.} \end{cases}
\]

Then
\[
P\left\{ (x^n,\hat{y}^n,\hat{z}^n) \in \mathcal{T}_{\epsilon'}(X,Y,Z) \right\} \leq 2^{-n I(Y;Z|X)-\delta(\epsilon')},
\]

(d) Let \( (\hat{x}^n,\hat{y}^n,\hat{z}^n) \sim \prod_{i=1}^{n} p_{X}(\hat{x}_i)p_{Y}(\hat{y}_i)p_{Z|X,Y}(\hat{z}_i|\hat{x}_i,\hat{y}_i) \). Then
\[
P\left\{ (\hat{x}^n,\hat{y}^n,\hat{z}^n) \in \mathcal{T}_{\epsilon'}(X,Y,Z) \right\} \leq 2^{-n I(X;Y)}.
\]

2.15. Jointly typical triples. Given \( (X,Y,Z) \sim p(x,y,z) \), let
\[
\mathcal{A}_n = \{(x^n,y^n,z^n) : (x^n,y^n) \in \mathcal{T}_{\epsilon'}(X,Y), (y^n,z^n) \in \mathcal{T}_{\epsilon'}(Y,Z), (x^n,z^n) \in \mathcal{T}_{\epsilon'}(X,Z)\}.
\]

(a) Show that \( |\mathcal{A}_n| \leq 2^{n(H(X,Y)+H(Y,Z)+H(X,Z)+\delta(\epsilon))/2} \).
(Hint: First show that \( |\mathcal{A}_n| \leq 2^{n(H(X,Y)+H(Y,Z)+\delta(\epsilon))/2} \).)

(b) Does a corresponding lower bound hold in general? (Hint: Consider \( X = Y = Z \).)

3.1. Memoryless property. Show that under the given definition of a \( (2^nR,n) \) code, the memoryless property
\[
p(y_i|x^i,y^{i-1},m) = p_{Y|X}(y_i|x_i), \quad i \in [1:n],
\]
reduces to
\[
p(y^n|x^n,m) = \prod_{i=1}^{n} p_{Y|X}(y_i|x_i).
\]
3.2. **Z channel.** The Z channel has binary input and output alphabets, and conditional pmf \( p(0|0) = 1, \ p(1|1) = p(0|1) = 1/2 \). Find the capacity \( C \).

3.5. **Maximum likelihood decoding.** The achievability proof of the channel coding theorem in Section 3.1.1 uses joint typicality decoding. This technique greatly simplifies the proof, especially for multiuser channels. However, given a codebook, the joint typicality decoding is not optimal in terms of minimizing the probability of decoding error (it is in fact surprising that such a suboptimal decoding rule can still achieve capacity).

Since the messages are equally likely, maximum likelihood decoding (MLD)

\[
\hat{m} = \arg \max_m p(y^n|m) = \arg \max_m \prod_{i=1}^n p_{Y|X}(y_i|x_i(m))
\]

is the optimal decoding rule (when there is a tie, choose an arbitrary index that maximizes the likelihood). Achievability proofs using MLD are more complex but provide tighter bounds on the optimal error exponent (reliability function); see, for example, Gallager (1968).

In this problem we use MLD to establish achievability of the capacity for a BSC(\( p \)), \( p < 1/2 \). Define the Hamming distance \( d(x^n, y^n) \) between two binary sequences \( x^n \) and \( y^n \) as the number of positions where they differ, i.e., \( d(x^n, y^n) = |\{i : x_i \neq y_i\}| \).

(a) Show that the MLD rule reduces to the minimum Hamming distance decoding rule—declare \( \hat{m} \) is sent if \( d(x^n(\hat{m}), y^n) < d(x^n(m), y^n) \) for all \( m \neq \hat{m} \).

(b) Now fix \( X \sim \text{Bern}(1/2) \). Using random coding and minimum distance decoding, show that for every \( \epsilon > 0 \), the probability of error averaged over codebooks is upper bounded as

\[
P_e^{(n)} = P\{M \neq \hat{M} \mid M = 1\} \leq P\{d(X^n(1), Y^n) > n(p + \epsilon) \mid M = 1\} + (2^{nR} - 1)P\{d(X^n(2), Y^n) \leq n(p + \epsilon) \mid M = 1\}.
\]

(c) Show that the first term tends to zero as \( n \to \infty \). It can be shown using the Chernoff–Hoeffding bound (Hoeffding 1963) that

\[
P\{d(X^n(1), Y^n) > n(p + \epsilon) \mid M = 1\} \leq 2^{-n(1-\epsilon)}.
\]

Using these results, show that any \( R < C = 1 - H(p) \) is achievable.

3.7. **Nonuniform message.** Recall that a \( (2^{nR}, n) \) code for the DMC \( p(y|x) \) consists of an encoder \( x^n = \phi_n(m) \) and a decoder \( \hat{m} = \psi_n(y^n) \). Suppose that there exists a sequence of \( (2^{nR}, n) \) codes such that \( P_e^{(n)} = P\{M \neq \hat{M} \} \) tends to zero as \( n \to \infty \), where \( M \) is uniformly distributed over \( [1 : 2^{nR}] \). (In other words, the rate \( R \) is achievable.) Now suppose that we wish to communicate a message \( M' \) that is arbitrarily (not uniformly) distributed over \( [1 : 2^{nR}] \).

(a) Show that there exists a sequence of \( (2^{nR}, n) \) codes with encoder–decoder pairs \( (\phi'_n, \psi'_n) \) such that

\[
\lim_{n \to \infty} P\{M' \neq \hat{M}' \} = 0.
\]

(Hint: Consider a random ensemble of codes \( \Phi'_n = \phi_n \circ \sigma \) and \( \Psi'_n = \sigma^{-1} \circ \psi_n \), where \( \sigma \) is a random permutation. Show the probability of error, averaged over \( M' \) and \( \sigma \), is equal to \( P_e^{(n)} \) and conclude that there exists a good permutation \( \sigma \) for each \( M' \).)

(b) Does this result imply that the capacity for the maximal probability of error is equal to that for the average probability of error?