10) (*Pairwise Independence*) $X_1, X_2, \ldots, X_{n-1}$ are i.i.d. Bernoulli(1/2) random variables. We will first prove that for any $k \leq n-1$, the probability that $\sum_{i=1}^{k} X_i$ is odd is $1/2$. We will prove this by induction. Clearly this is true for $k = 1$. Assume that it is true for $k-1$. Let $S_k = \sum_{i=1}^{k} X_i$. Then

$$\begin{aligned} P(S_k \text{ odd}) &= P(S_{k-1} \text{ odd})P(X_k = 0) + P(S_{k-1} \text{ even})P(X_k = 1) && (257)\\ &= \frac{1}{2}\frac{1}{2} + \frac{1}{2}\frac{1}{2} && (258)\\ &= \frac{1}{2}. && (259)\end{aligned}$$

Hence for all $k \leq n-1$, the probability that S_k is odd is equal to the probability that it is even. Hence,

$$P(X_n = 1) = P(X_n = 0) = \frac{1}{2}. \quad (260)$$

a) It is clear that when i and j are both less than n, X_i and X_j are independent. The only possible problem is when $j = n$. Taking $i = 1$ without loss of generality,

$$\begin{aligned} P(X_1 = 1, X_n = 1) &= P(X_1 = 1, \sum_{i=2}^{n-1} X_i \text{ even}) && (261)\\ &= P(X_1 = 1)P(\sum_{i=2}^{n-1} X_i \text{ even}) && (262)\\ &= \frac{1}{2}\frac{1}{2} && (263)\\ &= P(X_1 = 1)P(X_n = 1) && (264)\end{aligned}$$

and similarly for other possible values of the pair (X_1, X_n). Hence X_1 and X_n are independent.

b) Since X_i and X_j are independent and uniformly distributed on $\{0, 1\}$,

$$H(X_i, X_j) = H(X_i) + H(X_j) = 1 + 1 = 2 \text{ bits}. \quad (265)$$

c) By the chain rule and the independence of $X_1, X_2, \ldots, X_{n_1}$, we have

$$\begin{aligned} H(X_1, X_2, \ldots, X_n) &= H(X_1, X_2, \ldots, X_{n-1}) + H(X_n | X_{n-1}, \ldots, X_1) && (266)\\ &= \sum_{i=1}^{n-1} H(X_i) + 0 && (267)\\ &= n - 1, && (268)\end{aligned}$$

since X_n is a function of the previous X_i's. The total entropy is not n, which is what would be obtained if the X_i's were all independent. This example illustrates that pairwise independence does not imply complete independence.

14) The key point is that functions of a random variable have lower entropy. Since $(Y_1, Y_2, \ldots, Y_n)$ is a function of $(X_1, X_2, \ldots, X_n)$ (each Y_i is a function of the corresponding X_i), we have (from Problem 2.4)

$$H(Y_1, Y_2, \ldots, Y_n) \leq H(X_1, X_2, \ldots, X_n) \tag{275}$$

Dividing by n, and taking the limit as $n \to \infty$, we have

$$\lim_{n\to\infty} \frac{H(Y_1, Y_2, \ldots, Y_n)}{n} \leq \lim_{n\to\infty} \frac{H(X_1, X_2, \ldots, X_n)}{n} \tag{276}$$

or

$$\mathcal{H}(\mathcal{Y}) \leq \mathcal{H}(\mathcal{X}) \tag{277}$$

(b) Again, we shall first verify that Z is stationary:

Let $m, l \geq 1$. We have

$$P(Z_{1+l} = z_1, \cdots Z_{m+l} = z_m)$$

$$= P(\psi(X_{1+l}, X_{2+l}) = z_1, \cdots, \psi(X_{m+l}, X_{m+1+l}) = z_m)$$

X stationary

$$= P(\psi(X_1, X_2) = z_1, \cdots, \psi(X_m, X_{m+1}) = z_m)$$

$$= P(Z_1 = z_1, \cdots, Z_m = z_m),$$

showing that $\mathcal{Z}$ is indeed stationary.

Now claim that $H(\mathcal{Z}) \le H(\mathcal{X})$ #:

pf For each $n \ge 1$, $(Z_1, \cdots, Z_n)$ is a function of $(X_1, \cdots, X_{n+1})$, and thus

$$H(Z_1, \cdots, Z_n) \le H(X_1, \cdots, X_{n+1}),$$

or equivalently,

$$\frac{1}{n} H(Z_1, \cdots, Z_n) \le \frac{1}{n} H(X_1, \cdots, X_{n+1}) \quad \cdots (2\text{-}2).$$

As $n \to \infty$, the left hand side of (2-2) goes to

$$\lim_{n\to\infty} \frac{1}{n} H(Z_1, \cdots, Z_n) = H(\mathcal{Z}) \quad \cdots (2\text{-}3).$$

In addition, as $n \to \infty$, the right hand side of (2-2) goes to

$$\lim_{n\to\infty} \frac{1}{n} H(X_1, \cdots, X_{n+1}) = \lim_{n\to\infty} \frac{n+1}{n} \frac{H(X_1, \cdots, X_{n+1})}{n+1}$$

both limits exist

$$= \left(\lim_{n\to\infty} \frac{n+1}{n}\right)\left(\lim_{n\to\infty} \frac{H(X_1, \cdots, X_{n+1})}{n+1}\right)$$

$$= 1 \cdot H(\mathcal{X}) = H(\mathcal{X}). \quad \cdots (2\text{-}4).$$

By (2-3) and (2-4), we can let $n \to \infty$ on both sides of (2-2) and get

$$H(\mathcal{Z}) = \lim_{n\to\infty} \frac{1}{n} H(Z_1, \cdots, Z_n)$$

$$\leq \lim_{n\to\infty} \frac{1}{n} H(X_1, \cdots, X_{n+1})$$

$$= H(\mathcal{X}),$$

which proves the claim.

27) *Test for unique decodability.*

The proof of the Sardinas-Patterson test has two parts. In the first part, we will show that if there is a code string that has two different interpretations, then the code will fail the test. The simplest case is when the concatenation of two codewords yields another codeword. In this case, S_2 will contain a codeword, and hence the test will fail.

In general, the code is not uniquely decodeable, iff there exists a string that admits two different parsings into codewords, e.g.

$$x_1x_2x_3x_4x_5x_6x_7x_8 = x_1x_2, x_3x_4x_5, x_6x_7x_8 = x_1x_2x_3x_4, x_5x_6x_7x_8. \tag{414}$$

In this case, S_2 will contain the string x_3x_4, S_3 will contain x_5, S_4 will contain $x_6x_7x_8$, which is a codeword. It is easy to see that this procedure will work for any string that has two different parsings into codewords; a formal proof is slightly more difficult and using induction.

In the second part, we will show that if there is a codeword in one of the sets $S_i, i \geq 2$, then there exists a string with two different possible interpretations, thus showing that the code is not uniquely decodeable. To do this, we essentially reverse the construction of the sets. We will not go into the details - the reader is referred to the original paper.

a) Let S_1 be the original set of codewords. We construct S_{i+1} from S_i as follows: A string y is in S_{i+1} iff there is a codeword x in S_1, such that xy is in S_i or if there exists a $z \in S_i$ such that zy is in S_1 (i.e., is a codeword). Then the code is uniquely decodable iff none of the S_i, $i \geq 2$ contains a codeword. Thus the set $S = \cup_{i \geq 2} S_i$.

b) A simple upper bound can be obtained from the fact that all strings in the sets S_i have length less than l_{max}, and therefore the maximum number of elements in S is less than $2^{l_{max}}$.

c) i) $\{0, 10, 11\}$. This code is instantaneous and hence uniquely decodable.

ii) $\{0, 01, 11\}$. This code is a suffix code (see problem 11). It is therefore uniquely decodable. The sets in the Sardinas-Patterson test are $S_1 = \{0, 01, 11\}$, $S_2 = \{1\} = S_3 = S_4 = \ldots$.

iii) $\{0, 01, 10\}$. This code is not uniquely decodable. The sets in the test are $S_1 = \{0, 01, 10\}$, $S_2 = \{1\}$, $S_3 = \{0\}$, ... Since 0 is codeword, this code fails the test. It is easy to see otherwise that the code is not UD - the string 010 has two valid parsings.

iv) $\{0, 01\}$. This code is a suffix code and is therefore UD. THe test produces sets $S_1 = \{0, 01\}$, $S_2 = \{1\}$, $S_3 = \phi$.

v) $\{00, 01, 10, 11\}$. This code is instantaneous and therefore UD.

vi) $\{110, 11, 10\}$. This code is uniquely decodable, by the Sardinas-Patterson test, since $S_1 = \{110, 11, 10\}$, $S_2 = \{0\}$, $S_3 = \phi$.

vii) $\{110, 11, 100, 00, 10\}$. This code is UD, because by the Sardinas Patterson test, $S_1 = \{110, 11, 100, 00, 10\}$, $S_2 = \{0\}$, $S_3 = \{0\}$, etc.

d) We can produce infinite strings which can be decoded in two ways only for examples where the Sardinas Patterson test produces a repeating set. For example, in part (ii), the string 011111... could be parsed either as 0,11,11,... or as 01,11,11,.... Similarly for (viii), the string 10000... could be parsed as 100,00,00,... or as 10,00,00,.... For the instantaneous codes, it is not possible to construct such a string, since we can decode as soon as we see a codeword string, and there is no way that we would need to wait to decode.

4) *Channel capacity.*

$$Y = X + Z (\text{mod } 11) \tag{594}$$

where

$$Z = \begin{cases} 1 & \text{with probability} 1/3 \\ 2 & \text{with probability} 1/3 \\ 3 & \text{with probability} 1/3 \end{cases} \tag{595}$$

In this case,

$$H(Y|X) = H(Z|X) = H(Z) = \log 3, \tag{596}$$

independent of the distribution of X, and hence the capacity of the channel is

$$\begin{aligned} C &= \max_{p(x)} I(X;Y) && (597)\\ &= \max_{p(x)} H(Y) - H(Y|X) && (598)\\ &= \max_{p(x)} H(Y) - \log 3 && (599)\\ &= \log 11 - \log 3, && (600) \end{aligned}$$

which is attained when Y has a uniform distribution, which occurs (by symmetry) when X has a uniform distribution.

a) The capacity of the channel is $\log \frac{11}{3}$ bits/transmission.

b) The capacity is achieved by an uniform distribution on the inputs. $p(X = i) = \frac{1}{11}$ for $i = 0, 1, \ldots, 10$.

11) *Time-varying channels*.
We can use the same chain of inequalities as in the proof of the converse to the channel coding theorem. Hence

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - H(Y^n|X^n) && (615)\\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|Y_1,\ldots,Y_{i-1},X^n) && (616)\\ &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i), && (617) \end{aligned}$$

since by the definition of the channel, Y_i depends only on X_i and is conditionally independent of everything

else. Continuing the series of inequalities, we have

$$\begin{aligned} I(X^n;Y^n) &= H(Y^n) - \sum_{i=1}^{n} H(Y_i|X_i) && (618)\\ &\le \sum_{i=1}^{n} H(Y_i) - \sum_{i=1}^{n} H(Y_i|X_i) && (619)\\ &\le \sum_{i=1}^{n} (1 - h(p_i)), && (620) \end{aligned}$$

with equality if $X_1, X_2, \ldots, X_n$ is chosen i.i.d. $\sim$ Bern(1/2). Hence

$$\max_{p(\mathbf{x})} I(X_1, X_2, \ldots, X_n; Y_1, Y_2, \ldots, Y_n) = \sum_{i=1}^{n} (1 - h(p_i)). \quad (621)$$