1. **The Weak Law of Large Numbers**

Let $X_1, X_2, \ldots$ be a sequence of independent identically distributed random variables with common mean $\mu$ and associated transform $M_X$. We assume that $M_X(s)$ is finite when $s \in (-d, d)$, where $d$ is some positive number. Let

$$\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}.$$ 

(a) Show that the transform associated with $\bar{X}_n$ satisfies

$$M_{\bar{X}_n}(s) = M_X(s/n).$$

(b) Suppose that the transform $M_X(s)$ has a first order Taylor series expansion around $s = 0$, of the form

$$M_X(s) = a + bs + o(s),$$

where $o(s)$ is a function that satisfies $\lim_{s \to 0} o(s)/s = 0$. Find $a$ and $b$ in terms of $\mu$.

(c) Show that

$$\lim_{n \to \infty} M_{\bar{X}_n}(s) = e^{\mu s}, \quad \text{for all } s \in (-d, d).$$

*Hint:* If $\{a_n\}_{n \in \mathbb{N}}$ is a real sequence with $\lim_{n \to \infty} a_n = a \in \mathbb{R}$, then $\lim_{n \to \infty} \left(1 + \frac{a_n}{n}\right)^n = e^a$.

(d) Deduce that $\bar{X}_n \overset{d}{\to} \mu$.

*Note:* From this together with Problem 6 we can conclude that $\bar{X}_n \overset{P}{\to} \mu$.

**Solution:**

(a) Let $S_n = X_1 + \cdots + X_n$. Then because the sequence of random variables is i.i.d. we have that

$$M_{S_n}(s) = M_X(s)^n.$$
In addition

\[ M_{\bar{X}_n}(s) = M_{\frac{S_n}{n}}(s) \]

\[ = \mathbb{E}\left[ \exp\left(\frac{S_n}{n}\right) \right] \]

\[ = M_{S_n}(s/n) \]

\[ = M_X(s/n)^n. \]

(b) \( a = 1 \) and \( b = \mu. \)

(c)

\[ \lim_{n \to \infty} M_{\bar{X}_n}(s) = \lim_{n \to \infty} \left(1 + \frac{\mu s + \frac{o(1/n)}{1/n}}{n}\right)^n \]

\[ = e^{\mu s}. \]

(d) Let \( M = \mu \) be a constant random variable, and observe that \( M_M(s) = e^{s\mu}. \)

Since \( M_{\bar{X}_n} \to M_M \) as \( n \to \infty, \) from the inversion property of the MGF we can deduce that \( \bar{X}_n \xrightarrow{d} M = \mu \) as \( n \to \infty. \)

2. Huffman Questions

Consider a set of \( n \) objects. Let \( X_i = 1 \) or 0 accordingly as the \( i \)-th object is good or defective. Let \( X_1, X_2, \ldots, X_n \) be independent with \( \mathbb{P}(X_i = 1) = p_i \); and \( p_1 > p_2 > \cdots > p_n > 1/2. \) We are asked to determine the set of all defective objects. Any yes-no question you can think of is admissible.

(a) Propose an algorithm based on Huffman coding in order to identify all defective objects.

(b) If the longest sequence of questions is required by nature’s answers to our questions which are based on Huffman coding, then what (in words) is the last question we should ask? And what two sets are we distinguishing with this question?

Note: This problem is related to the ‘Entropy and Information Content’ section of Lab 4.

Solution:

(a) Let \( x \in \{0,1\}^n \) be a possible configuration of whether each object is good or defective. Because of the independence assumption, we can calculate the joint probabilities as follows

\[ \mathbb{P}(X = x) = \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} p_i^{x_i}(1 - p_i)^{1-x_i}. \]

Now according to those joint probabilities we use Huffman coding to encode all possible configurations \( x \in \{0,1\}^n. \)
The naive strategy is to try to determine directly the true configuration $x_t \in \{0,1\}^n$ by identifying each bit of $x_t$, which results in $n$ yes-no questions.

Instead our strategy is to try to determine the Huffman code $C(x_t) \in \{0,1\}^+$ that corresponds to the true configuration $x_t$. We are going to do so by identifying each bit of $C(x_t) \in \{0,1\}^+$.

Using this strategy the expected number of questions that we are going to ask is going to be between $H(X_1,\ldots,X_n)$ and $H(X_1,\ldots,X_n) + 1$.

Because of independence $H(X_1,\ldots,X_n) = H(X_1) + \cdots + H(X_n)$, and this quantity can be way smaller than $n$.

(b) If the longest sequence of questions is required, then the last question would try to distinguish whether the true configuration is the one with lowest probability or the one with the second lowest probability. So according to the information $p_1 > p_2 > \cdots > p_n > 1/2$, the last question would try to distinguish if the true configuration is $(0,\ldots,0,0)$ or $(0,\ldots,0,1)$, and the actual question could be "Is the $n$-th object defective?"

3. Channel Capacity of the Binary Symmetric Channel

Random Code

Each word $w \in \{0,1\}^k$, independently is encoded by a codeword $X_w = (X_w(1),\ldots,X_w(n))$ which consists of $n$ i.i.d. Bernoulli(1/2) random variables. Define the encoding function $f_n : \{0,1\}^k \to \{0,1\}^n$ as

$$f_n(w) = (X_w(1),\ldots,X_w(n)).$$

Noise

During transmission, a codeword $X_w = (X_w(1),\ldots,X_w(n))$ is corrupted by noise $N = (N(1),\ldots,N(n))$ which is assumed to consist of $n$ i.i.d. Bernoulli($p$) random variables. Then the received message $Y_w = (Y_w(1),\ldots,Y_w(n))$ can be represented as $Y_w(i) = X_w(i) \oplus N(i)$, for $i = 1,\ldots,n$, where $\oplus$ denotes the XOR operation.

Decoding

We decode the received message $Y_w$, using the following decoding function $g_n : \{0,1\}^n \to \{0,1\}^k$

$$g_n(Y_w) = \begin{cases} u, & \text{if there is a unique } u \in \{0,1\}^k \text{ s.t. } Y_w \in \text{DecodeBox}(X_u) \\ \text{fail}, & \text{otherwise,} \end{cases}$$

where for any $\epsilon > 0$ we define

$$\text{DecodeBox}(x) = \left\{ y \in \{0,1\}^n : \left| \sum_{i=1}^n y(i) \oplus x(i) - pn \right| \leq n\epsilon \right\}.$$

(a) Using the Chernoff bound argue that

$$\mathbb{P}(Y_w \notin \text{DecodeBox}(X_w)) \leq 2e^{-2n\epsilon^2}.$$
(b) Argue that \(|\text{DecodeBox}(0)| = |\text{DecodeBox}(x)|\), for all \(x \in \{0, 1\}^n\).

(c) Argue that each component \(Y_w(i)\) is marginally a Bernoulli\((1/2)\) random variable.

(d) For \(w \neq u\), show that

\[ P(Y_w \in \text{DecodeBox}(X_u)) = \frac{|\text{DecodeBox}(0)|}{2^n}. \]

(e) Fix a word \(w \in \{0, 1\}^k\). Argue that

\[ P(\exists u \neq w : Y_w \in \text{DecodeBox}(X_u)) \leq 2^k \frac{|\text{DecodeBox}(0)|}{2^n}. \]

(f) Use the A.E.P. from Homework 5 Problem 5 (c) to deduce that

\[ |\text{DecodeBox}(0)| \leq 2^n (H(p) + \epsilon'), \]

where \(\epsilon'\) is some linear function of \(\epsilon\).

(g) How large can the channel capacity \(C = k/n\) be in order to ensure that asymptotically both error probabilities go to zero?

\textbf{Solution:}

(a)

\[ P(Y_w \notin \text{DecodeBox}(X_w)) = P\left( \left| \sum_{i=1}^{n} Y_w(i) \oplus X_w(i) - pm \right| > ne \right) \]

\[ = P\left( \left| \sum_{i=1}^{n} N(i) - pm \right| > ne \right) \]

\[ \leq 2e^{-2ne^2}. \]

(b) We are going to give a bijection \(f_x : \text{DecodeBox}(0) \to \text{DecodeBox}(x)\). Consider the function

\[ f_x(y) = y \oplus x, \]

where the XOR operation is taking place component-wise.

- \(f_x\) is 1-1, since if \(y \oplus x = y' \oplus x\), then \(y = y'\).
- \(f_x\) is onto \(\text{DecodeBox}(x)\), since if \(y \in \text{DecodeBox}(x)\) then \(y \oplus x \in \text{DecodeBox}(0)\) and \(f_x(y \oplus x) = y\).

Since \(f_x : \text{DecodeBox}(0) \to \text{DecodeBox}(x)\) is a bijection, we can conclude that \(|\text{DecodeBox}(0)| = |\text{DecodeBox}(x)|\).

(c)

\[ P(Y_w(i) = 1) = P(X_w(i) = 1)P(N(i) = 0) + P(X_w(i) = 0)P(N(i) = 1) = 1/2. \]

Therefore, \(Y_w(i) \sim \text{Bernoulli}(1/2)\).
(d) By construction $X_u$ and $X_w$ are independent, so $X_u$ and $Y_w$ are independent as well. In addition $Y_w$ is marginally, uniformly distributed over $\{0,1\}^n$. Putting those two together with part (b) we obtain that

$$\mathbb{P}(Y_w \in \text{DecodeBox}(X_u)) = \sum_{x \in \{0,1\}^n} \mathbb{P}(Y_w \in \text{DecodeBox}(x) \mid X_u = x)\mathbb{P}(X_u = x)$$

$$= \sum_{x \in \{0,1\}^n} \mathbb{P}(Y_w \in \text{DecodeBox}(x))\mathbb{P}(X_u = x)$$

$$= \sum_{x \in \{0,1\}^n} \frac{\text{DecodeBox}(x)}{2^n} \mathbb{P}(X_u = x)$$

$$= \sum_{x \in \{0,1\}^n} \frac{\text{DecodeBox}(0)}{2^n} \mathbb{P}(X_u = x)$$

$$= \frac{\text{DecodeBox}(0)}{2^n} \sum_{x \in \{0,1\}^n} \mathbb{P}(X_u = x)$$

$$= \frac{\text{DecodeBox}(0)}{2^n}.$$  

(e) From the union bound and the previous part, we have that

$$\mathbb{P}(\exists u \neq w : Y_w \in \text{DecodeBox}(X_u)) \leq (2^k - 1)\frac{\text{DecodeBox}(0)}{2^n} \leq 2^k \frac{\text{DecodeBox}(0)}{2^n}.$$  

(f) Let $\epsilon' > 0$. The typical set for a sequence of $n$ i.i.d. Bernoulli($p$) random variables is defined as

$$A_{\epsilon'}^{(n)} := \left\{ y \in \{0,1\}^n : \left| \frac{1}{n} \log p(y_1, \ldots, y_n) - H(p) \right| \leq \epsilon' \right\}.$$  

We can reformulate this expression as

$$A_{\epsilon'}^{(n)} = \left\{ y \in \{0,1\}^n : \left| \log \left( p \sum_{i=1}^{n} y(i) (1-p)^{n-\sum_{i=1}^{n} y(i)} \right) - np \log p - n(1-p) \log(1-p) \right| \leq n \epsilon' \right\}$$

$$= \left\{ y \in \{0,1\}^n : \left| \left( \sum_{i=1}^{n} y(i) - np \right) \log p - \left( n(1-p) - n + \sum_{i=1}^{n} y(i) \right) \log(1-p) \right| \leq n \epsilon' \right\}$$

$$= \left\{ y \in \{0,1\}^n : \left| \left( \sum_{i=1}^{n} y(i) - np \right) \log \frac{p}{1-p} \right| \leq n \epsilon' \right\}.$$  

Therefore, if we set $\epsilon' = \left| \log \frac{p}{1-p} \right| \epsilon$ we have that $\text{DecodeBox}(0) = A_{\epsilon'}^{(n)}$, and Homework 5 problem 5 (c) yields

$$|\text{DecodeBox}(0)| = |A_{\epsilon'}^{(n)}| \leq 2^{n(H(p)+\epsilon')}.$$  

(g) From Part (a) we have that

$$\mathbb{P}(Y_w \not\in \text{DecodeBox}(X_w)) \leq 2 \exp(-2n\epsilon^2) \to 0, \text{ as } n \to \infty.$$
Combining Parts (e) and (f) we have that
\[ P(\exists u \neq w : Y_w \in \text{DecodeBox}(X_u)) \leq 2^{-n((1-(H(p)+\epsilon'))-C)}. \]

For this probability to go to zero as \( n \) goes to infinity we need to ensure that
\[ C < 1 - (H(p) + \epsilon'). \]

4. Number of Parameters

(a) Let \( Y_0, Y_1, \ldots, Y_n \) be binary random variables. How many parameters are required to parametrize the joint distribution \( P(Y_0 = y_0, Y_1 = y_1, \ldots, Y_n = y_n) \)?

(b) Let \( Z_0, Z_1, \ldots, Z_n \) be binary, independent random variables. How many parameters are required to parametrize the joint distribution \( P(Z_0 = z_0, Z_1 = z_1, \ldots, Z_n = z_n) \)?

(c) Let \( X_0, X_1, \ldots, X_n, \ldots \) be a Markov chain with state space \( S = \{0, 1\} \), initial distribution \( \pi_0 \) and transition probability matrix \( P \). How many parameters are required to parametrize the joint distribution \( P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) \), where \( X_0, X_1, \ldots, X_n \) are the first \( n + 1 \) random variables of the Markov chain given above?

(d) Say we want to construct a countably infinite sequence \( Z_0, Z_1, \ldots, Z_n, \ldots \) of independent random variables, which is not a Markov chain. Which defining property of Markov chains must this sequence violate? Give a concrete example of such sequence of independent random variables.

Solution:

(a) The joint distribution is an arbitrary function from \( \{0, 1\}^{n+1} \) to \([0,1]\), with the only requirement that its values sum up to one. Therefore \(2^{n+1} - 1\) parameters are required to parametrize the joint.

(b) The distribution of each \( Z_i \) just needs 1 parameter, hence the joint requires \( n + 1 \) parameters due to the independence assumption.

(c) Let \( \pi_0 \) be the initial distribution of the Markov chain and \( P \) be transition matrix. Then the joint can be decomposed as
\[ P(X_0 = x_0, X_1 = x_1, \ldots, X_n = x_n) = \pi_0(x_0) \cdot P(x_0, x_1) \cdots P(x_{n-1}, x_n), \]
which means that the joint only requires 1 parameter for the initial distribution \( \pi_0 \), and 2 parameters for the transition matrix \( P \), so overall we need 3 parameters in order to characterize the joint distribution of a binary Markov chain.

(d) Because the random variables that we want to construct are independent the Markov property is going to be satisfied for sure. It is the time homogeneous property that is going to be violated in any valid example. For instance let
\[ Z_n \sim \begin{cases} \text{Bernoulli}(\frac{1}{2}), & \text{if } n \text{ is even} \\ \text{Bernoulli}(\frac{3}{4}), & \text{if } n \text{ is odd.} \end{cases} \]
Then the transition probabilities depend on \( n \) and in particular
\[
Z_{n+1} \mid Z_n \sim \begin{cases} 
\text{Bernoulli}(\frac{3}{4}), & \text{if } n \text{ is even} \\
\text{Bernoulli}(\frac{1}{4}), & \text{if } n \text{ is odd}.
\end{cases}
\]

5. Backwards Markov Property
Let \( (X_n)_{n \in \mathbb{N}} \) be a Markov chain with state space \( S \). Show that for every \( m, k \in \mathbb{N} \), with \( m \geq 1 \), we have
\[
P(X_k = i_0 \mid X_{k+1} = i_1, \ldots, X_{k+m} = i_m) = P(X_k = i_0 \mid X_{k+1} = i_1),
\]
for all states \( i_0, i_1, \ldots, i_m \in S \).

Solution: By definition of conditional probability we can write
\[
P(X_k = i_0 \mid X_{k+1} = i_1, \ldots, X_{k+m} = i_m) = \frac{P(X_k = i_0, X_{k+1} = i_1, \ldots, X_{k+m} = i_m)}{P(X_{k+1} = i_1, \ldots, X_{k+m} = i_m)}.
\]
Now using the Markov property the numerator can be written as
\[
P(X_k = i_0, X_{k+1} = i_1) \prod_{j=2}^{m} P(X_{k+j} = i_j \mid X_{k+j-1} = i_{j-1}),
\]
and the denominator can be written as
\[
P(X_{k+1} = i_1) \prod_{j=2}^{m} P(X_{k+j} = i_j \mid X_{k+j-1} = i_{j-1}).
\]
So the products cancel out and we obtain
\[
P(X_k = i_0 \mid X_{k+1} = i_1, \ldots, X_{k+m} = i_m) = \frac{P(X_k = i_0, X_{k+1} = i_1)}{P(X_{k+1} = i_1)} = P(X_k = i_0 \mid X_{k+1} = i_1).
\]

6. [Bonus] The CLT Implies the WLLN
The bonus question is just for fun. You are not required to submit the bonus question, but do give it a try and write down your progress.

(a) Let \( \{X_n\}_{n \in \mathbb{N}} \) be a sequence of random variables. Show that if \( X_n \stackrel{d}{\to} c \), where \( c \) is a constant, then \( X_n \stackrel{P}{\to} c \).

(b) Let \( \{X_n\}_{n \in \mathbb{N}} \) be a sequence of i.i.d. random variables, with mean \( \mu \) and finite variance \( \sigma^2 \). Show that the CLT implies the WLLN.

Solution:
(a) Since \( X_n \xrightarrow{d} c \), we can deduce that for any \( \epsilon > 0 \), we have

\[
\lim_{n \to \infty} F_{X_n}(c - \epsilon) = 0,
\]

\[
\lim_{n \to \infty} F_{X_n}\left(c - \frac{\epsilon}{2}\right) = 1.
\]

Using this fact we have that

\[
\lim_{n \to \infty} \mathbb{P}(|X_n - c| \geq \epsilon) = \lim_{n \to \infty} \left[\mathbb{P}(X_n \leq c - \epsilon) + \mathbb{P}(X_n \geq c + \epsilon)\right]
\]

\[
= \lim_{n \to \infty} \mathbb{P}(X_n \leq c - \epsilon) + \lim_{n \to \infty} \mathbb{P}(X_n \geq c + \epsilon)
\]

\[
= \lim_{n \to \infty} F_{X_n}(c - \epsilon) + \lim_{n \to \infty} \mathbb{P}(X_n \geq c + \epsilon)
\]

\[
\leq 0 + \lim_{n \to \infty} \mathbb{P}\left(X_n > c + \frac{\epsilon}{2}\right)
\]

\[
= 1 - \lim_{n \to \infty} F_{X_n}\left(c + \frac{\epsilon}{2}\right)
\]

\[
= 0.
\]

Therefore \( \lim_{n \to \infty} \mathbb{P}(|X_n - c| \geq \epsilon) = 0 \), for all \( \epsilon > 0 \) which means that \( X_n \xrightarrow{p} c \).

(b) From the CLT we know that

\[
\sqrt{\frac{n}{\sigma}} \left(\frac{1}{n} \sum_{i=1}^{n} X_i - \mu\right) \xrightarrow{d} Z \sim \mathcal{N}(0, 1).
\]

In addition \( \frac{\sigma}{\sqrt{n}} \to 0 \), so

\[
\frac{1}{n} \sum_{i=1}^{n} X_i - \mu \xrightarrow{d} 0
\]

or stated another way

\[
\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{d} \mu.
\]

Finally using Part (a) we can conclude that

\[
\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow{p} \mu.
\]