1. **MMSE Estimation, Data Processing and Innovations**

The random variables $X,Y,Z$ on a common probability space $(\Omega, \mathcal{F}, P)$ are said to form a Markov chain in that order, denoted by $X \rightarrow Y \rightarrow Z$, if the conditional distribution of $Z$ depends only on $Y$ and is conditionally independent of $X$. In other words, $X,Y,Z$ form a Markov chain $X \rightarrow Y \rightarrow Z$ if for any two events $\mathcal{A}, \mathcal{B} \in \mathcal{F}$

$$P(X \in \mathcal{A}, Z \in \mathcal{B}|Y) = P(X \in \mathcal{A}|Y)P(Z \in \mathcal{B}|Y).$$

Assuming that $X,Y,Z$ are jointly continuous, this condition can be equivalently described by

$$f_{XYZ}(x,y,z) = f_X(x)f_{Y|X}(y|x)f_{Z|Y}(z|y).$$

Moreover, for any two random variables $S,T$, define the average conditional variance of $S$ given $T$ as the mean square error of the conditional mean estimator of $S$ given $T$:

$$E[\text{Var}(S|T)] = \text{MMSE}(S|T) = E\left[(S - E[S|T])^2\right],$$

where $\text{Var}(S|T) = E[(S - E[S|T])^2|T]$. Here, we assume that all the involved expectations exist (i.e., the random variables $S,T$ are second order).

(a) Suppose that $X \rightarrow Y \rightarrow Z$. Show that

$$E[\text{Var}(X|Z)] - E[\text{Var}(X|Y)] = E\left[(E[X|Z] - E[X|Y])^2\right].$$

Conclude that $E[\text{var}(X|Y)] \leq E[\text{var}(X|Z)]$, i.e., the MMSE along Markov chains increases.

**Note:** The inequality in this part is an analogue of the data processing inequality in information theory: if $X \rightarrow Y \rightarrow Z$, then $I(X;Y) \geq I(X;Z)$, where $I(S;T)$ is the mutual information of two random variables $S,T$.

**Hint:** To solve this part, you need to use the Markov property of $X \rightarrow Y \rightarrow Z$ and to think in terms of the Orthogonality Principle.

**Solution**

There are alternative ways to argue on this problem. All these approaches are based on combining the Markov property with the orthogonality principle and relevant results.

Directly using the definitions, we have:
\[ E[\text{Var}(X \mid Z)] - E[\text{Var}(X \mid Y)] = E[(X - E[X \mid Z])^2] - E[(X - E[X \mid Y])^2] = \]
\[ E[-2X E[X \mid Z] + (E[X \mid Z])^2 + 2X E[X \mid Y] - (E[X \mid Y])^2] = \]
\[ E[(E[X \mid Z])^2 + (E[X \mid Y])^2 - 2E[X \mid Y]E[X \mid Z] + 2E[X \mid Y]E[X \mid Z] - 2(E[X \mid Y])^2 - 2X E[X \mid Z] + 2X E[X \mid Y] = \]
\[ E[(E[X \mid Z] - E[X \mid Y])^2] - 2E[(E[X \mid Y] - X)(E[X \mid Y] - E[X \mid Z])]. \]

To finish, we need to show that \(2E[(E[X \mid Y] - X)(E[X \mid Y] - E[X \mid Z])] = 0\). Note that
\[
E[(E[X \mid Y] - X)(E[X \mid Y] - E[X \mid Z])]
= E[(E[X \mid Y] - X)E[X \mid Y] + E[E[X \mid Z] (X - E[X \mid Y])] = 
= E[E[X \mid Z] (X - E[X \mid Y]) ] \quad \text{(Orthogonality principle)}
= E[E[X \mid Z] (X - E[X \mid Y, Z])] \quad \text{(Markov property : } X \perp Z \text{ given } Y) \]
= 0 \quad \text{(projection onto nested classes)}

A different way to see this is the following:
\[
E[E[X \mid Z] (X - E[X \mid Y])] = E[E[X \mid Z] X] - E[E[X \mid Z] E[X \mid Y]] = 
E[(E[X \mid Z])^2] - E[E[X \mid Z] E[E[X \mid Y] \mid Z]] = 
E[(E[X \mid Z])^2] - E[E[X \mid Z] E[E[X \mid Y, Z] \mid Z]] = \quad \text{(Markov property)}
E[(E[X \mid Z])^2] - E[(E[X \mid Z])^2] = 0. \quad \text{(smoothing property of conditional expectation)}
\]

Alternatively, one may use Proposition 3.4 in the textbook to directly obtain:
\[
E[(X - E[X \mid Z])^2] = E[(X - E[X \mid Y, Z])^2] + E[(E[X \mid Y, Z] - E[X \mid Z])^2]
= E[(X - E[X \mid Y])^2] + E[(E[X \mid Y] - E[X \mid Z])^2]. \quad \text{(Markov property)}
\]

[Note: Parts (b) and (c) are bonus questions].

In the remaining of this problem, all random variables will be finite-variance, zero mean, circularly-symmetric complex Gaussian random variables. (Note: We only assume complex Gaussian random variables to state the following results in full generality. No complex calculations are required for the solution of this part. Moreover, this part aims to show that the geometric intuition, when the involved random
varialbes are complex-valued, does not change.). We use calligraphic letters, e.g., \( \mathcal{X} = \{X_i\} \), to denote finite sets of such variables. On sets of Gaussian random variables, we will assume that their statistics are jointly Gaussian. The set of all complex linear combinations of a given finite set \( \mathcal{X} \) of finite-variance, zero mean, circularly-symmetric jointly Gaussian complex random variables is a complex vector space \( \mathcal{G} \). Every element of \( \mathcal{G} \) is a finite-variance, zero mean, circularly-symmetric complex Gaussian random variable and every subset of \( \mathcal{G} \) is jointly Gaussian. The zero vector of \( \mathcal{G} \) is the unique zero variable \( 0 \). The dimension of \( \mathcal{G} \) is at most the cardinality, \( |\mathcal{X}| \), of \( \mathcal{X} \). If an inner product is defined on \( \mathcal{G} \) as the correlation \( \langle X, Y \rangle = E[XY^\ast] \) (\( \ast \) denotes complex conjugation), then \( \mathcal{G} \) becomes a Hilbert space. The squared norm of \( X \in \mathcal{G} \) is therefore its variance, i.e., \( \|X\|^2 = E[|X|^2] \), where \(|\cdot|\) denotes the complex modulus. The geometry of \( \mathcal{G} \) is totally determined by the matrix of inner products between the elements of \( \mathcal{X} \), i.e., by the autocorrelation or Gram matrix of \( \mathcal{X} \), \( R_{XY} = [E[X_iX_j^\ast]] \). Two random variables \( X, Y \in \mathcal{G} \) are orthogonal if \( \langle X, Y \rangle = E[XY^\ast] = 0 \). For jointly Gaussian random variables, orthogonality means independence. If \( \langle X, Y \rangle = 0 \), the Pythagorean Theorem holds:

\[
\|X + Y\|^2 = \|X\|^2 + \|Y\|^2.
\]

Given any subset \( \mathcal{Z} \subset \mathcal{G} \), the closure \( \bar{\mathcal{Z}} \) of \( \mathcal{Z} \), or the subspace generated by \( \mathcal{Z} \), is the set of all linear combinations of elements of \( \mathcal{Z} \). Also, the orthogonal complement \( \mathcal{Z}^\perp \) of \( \mathcal{Z} \) is defined as:

\[
\mathcal{Z}^\perp = \{X \in \mathcal{G} : \langle X, Z \rangle = 0, \ \forall Z \in \mathcal{Z}\}.
\]

\( \mathcal{Z}^\perp \) is also a subspace of \( \mathcal{G} \). Moreover, \( 0 \) is the only common element of \( \bar{\mathcal{Z}} \) and \( \mathcal{Z}^\perp \).

The key geometric property of the Hilbert space \( \mathcal{G} \) is the projection theorem: if \( \mathcal{V} \) and \( \mathcal{V}^\perp \) are orthogonal subspaces of \( \mathcal{G} \), then there exists a unique \( X_{\mathcal{V}} \) in \( \mathcal{V} \) and a unique \( X_{\mathcal{V}^\perp} \) in \( \mathcal{V}^\perp \) such that \( X = X_{\mathcal{V}} + X_{\mathcal{V}^\perp} \) and \( X_{\mathcal{V}}, X_{\mathcal{V}^\perp} \) are called the projections of \( X \) onto \( \mathcal{V} \) and \( \mathcal{V}^\perp \), respectively. Also, the Pythagorean Theorem implies that \( \|X\|^2 = \|X_{\mathcal{V}}\|^2 + \|X_{\mathcal{V}^\perp}\|^2 \).

Let \( \mathcal{X} \) be a subset of \( \mathcal{G} \) and \( \bar{\mathcal{X}} \) be the subspace generated by the elements of \( \mathcal{X} \). An orthogonal basis of \( \bar{\mathcal{X}} \) can be found by a recursive procedure, known as Gram-Schmidt orthogonalization: Let \( X_1, X_2, \ldots \) denote the elements of \( \mathcal{X} \) and let \( \bar{\mathcal{X}}^{i-1} \) be the subspace generated by \( \{X_1, X_2, \ldots, X_{i-1}\} \). Set \( \bar{\mathcal{X}}^0 = \emptyset \). For the \( i \)th iteration, we use the Projection Theorem to write:

\[
X_i = (X_i)_{\perp \bar{\mathcal{X}}^{i-1}} + (X_i)_{\parallel \bar{\mathcal{X}}^{i-1}}.
\]

We now have that \( (X_i)_{\perp \bar{\mathcal{X}}^{i-1}} = 0 \) if and only if \( \bar{\mathcal{X}}^{i-1} = \bar{\mathcal{X}}^{i} \) and therefore \( X_i \) may be excluded from the generating set \( \mathcal{X} \) without changing \( \bar{\mathcal{X}} \). If \( (X_i)_{\perp \bar{\mathcal{X}}^{i-1}} \neq 0 \), we can replace \( X_i \) by the variable \( E_i = (X_i)_{\perp \bar{\mathcal{X}}^{i-1}} \) in the generating set of \( \mathcal{X} \) without changing \( \bar{\mathcal{X}} \). Thus, the space generated by \( \bar{\mathcal{X}}^{i-1} \cup E_i \) is still \( \bar{\mathcal{X}}^i \), but \( E_i \perp \bar{\mathcal{X}}^{i-1} \). \( E_i \) is
called an *innovation variable*. Summarizing, given a generating set $\mathcal{X} = \{X_1, X_2, \ldots\}$ of some subspace $\overline{\mathcal{X}}$ of $\mathcal{G}$, we can reduce $\mathcal{X}$ to a linearly independent set $\mathcal{X}'$ generating the same subspace $\overline{\mathcal{X}}$. Assuming therefore and without loss of generality that $\mathcal{X}$ is a linearly independent set, we can find an orthogonal basis $\mathcal{E} = \{E_1, E_2, \ldots\}$ of $\overline{\mathcal{X}}$ such that $E_i = (X_i)_{\perp \overline{\mathcal{X}}_{i-1}} = X_i - (X_i)_{\overline{\mathcal{X}}_{i-1}}$, which corresponds to the *innovation set*.

Using the above introductory material, try to answer the following:

(b) Let $\mathcal{X}, \mathcal{Y}, \mathcal{Z}$ be jointly Gaussian sets of random variables. The MMSE estimate of $\mathcal{X}$ based on $\mathcal{Y}$ and $\mathcal{Z}$ is $\mathcal{X}_{\mid \mathcal{YZ}}$, which corresponds to the projection of $\mathcal{X}$ onto the subspace generated by the variables in $\mathcal{Y}$ and $\mathcal{Z}$ collectively. Show that

$$X_{\mid \mathcal{YZ}} = X_{\mid \mathcal{Y}} + (X_{\perp \mathcal{Y}})_{\mid \mathcal{Z}}.$$ 

**Solution**

$X_{\mid \mathcal{YZ}}$ is the projection of $\mathcal{X}$ onto the subspace generated by the variables in $\mathcal{Y}$ and $\mathcal{Z}$. Denote this subspace as $\overline{\mathcal{Y}} \overline{\mathcal{Z}}$ (this corresponds to the sum of the subspaces $\overline{\mathcal{Y}}$ and $\overline{\mathcal{Z}}$.) This subspace can be decomposed into the sum of two orthogonal subspaces:

$$\overline{\mathcal{Y}} + \overline{\mathcal{Z}} = \overline{\mathcal{Y}} + (\overline{\mathcal{Y}} \cap \overline{\mathcal{Z}}).$$

Therefore, $X_{\mid \mathcal{YZ}} = X_{\mid \mathcal{Y}} + (X_{\perp \mathcal{Y}})_{\mid \mathcal{Z}}$.

(c) Generalize the expression in the previous part to a general expression for the case where we wish to estimate $\mathcal{X}$ from a sequence of observations $\mathcal{Y} = \{Y_1, Y_2, \ldots\}$, when $\mathcal{X}, \mathcal{Y}$ are jointly Gaussian.

**Solution**

By generalizing part (b) we have:

$$X_{\mid \mathcal{Y}} = X_{\mid \overline{\mathcal{Y}}_1} + (X_{\perp \mathcal{Y}})_{\mid \overline{\mathcal{Y}}_2} + \ldots + (X_{\perp \mathcal{Y}^{i-1}})_{\mid \overline{\mathcal{Y}}_i} + \ldots,$$

where $\overline{\mathcal{Y}}_i$ is the subspace spanned by every new single observation (new information) and $\overline{\mathcal{Y}}^{i-1}$ is the subspace spanned by all previous observations up to time $i - 1$. 
2. DTMCs: A First Example

Let \( \{X_n\} \) be a Discrete-Time Markov Chain (DTMC) taking values in \( \{1, 2\} \) with probability transition matrix

\[
P = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix},
\]

where \( P_{ij} = P(X_{n+1} = j | X_n = i) \). Let \( \{Y_n\} \) be a different random process defined as:

\[
Y_n = \begin{cases} 
X_n, & \text{with probability 0.9} \\
X_n - 1, & \text{with probability 0.1}
\end{cases}
\]

Find \( \lim_{n \to \infty} P(X_n = 1 | Y_n = 1) \).

Solution

By Bayes rule:

\[
P(X_n = 1 | Y_n = 1) = \frac{P(Y_n = 1 | X_n = 1) P(X_n = 1)}{P(Y_n = 1)}.
\]

Also, by the definition of \( Y_n \),

\[
P(Y_n = 1 | X_n = 1) = P(Y_n = X_n) = 0.9.
\]

We now solve for the stationary distribution of the DTMC by solving for \( \pi = (\pi_1, \pi_2) \) the equation \( \pi P = \pi \) subject to \( \pi_1 + \pi_2 = 1 \). The solution is:

\[
\pi_1 = 1/3, \quad \pi_2 = 2/3.
\]

In this case, one can check that the stationary distribution is also limiting and therefore

\[
\lim_{n \to \infty} P(X_n = 1) = \pi_1 = 1/3, \quad \lim_{n \to \infty} P(X_n = 2) = \pi_2 = 2/3.
\]

Also, by total probability

\[
\lim_{n \to \infty} P(Y_n = 1) = \lim_{n \to \infty} P(Y_n = 1 | X_n = 1) P(X_n = 1) + P(Y_n = 1 | X_n = 2) P(X_n = 2)
\]

\[
= \frac{1}{3} \cdot 0.9 + \frac{2}{3} \cdot 1.
\]

Therefore,

\[
\lim_{n \to \infty} P(X_n = 1 | Y_n = 1) = P(Y_n = 1 | X_n = 1) \lim_{n \to \infty} P(X_n = 1) = \frac{0.9 \cdot 1}{\frac{1}{3} \cdot 0.9 + \frac{2}{3} \cdot 1} = \frac{9}{11}.
\]
3. Markov Chains and Martingales

Let $X$ be a random process taking values in a discrete, finite set $S$. Prove that the following statements are equivalent:

(a) $X$ is a Markov chain.
(b) For any function $f$,

$$M_t = f(X_t) - f(X_0) - \sum_{s=0}^{t-1} (E[f(X_{s+1})|X_s] - f(X_s))$$

is a martingale with respect to $X$.

Solution

Let $X$ be a finite state Markov chain. Then,

$$|M_t| \leq |f(X_t)| + |f(X_0)| + \sum_{s=0}^{t-1} |(E[f(X_{s+1})|X_s] - f(X_s))| < \infty.$$  

Additionally,

$$M_{t+1} - M_t = f(X_{t+1}) - f(X_t) - E[f(X_{t+1})|X_t] + f(X_t) = f(X_{t+1}) - E[f(X_{t+1})|X_t].$$

Therefore,

$$E[M_{t+1} - M_t|X_0, X_1, \ldots, X_t] = E[f(X_{t+1})|X_t] - E[f(X_{t+1})|X_t] = 0,$$

which shows that $M_t$ is a martingale with respect to $X$.

For the reverse implication: Let $f(X_t) = \mathbb{I}_{X_t=x}$. Then,

$$M_{t+1} - M_t = \mathbb{I}_{X_{t+1}=x} - E[\mathbb{I}_{X_{t+1}=x}|X_t].$$

By assumption, $E[M_{t+1} - M_t|X_0, X_1, \ldots, X_t] = 0$ since $M$ is a martingale with respect to $X$. By conditioning, we obtain:

$$E[\mathbb{I}_{X_{t+1}=x}|X_0, X_1, \ldots, X_t] = E[E[\mathbb{I}_{X_{t+1}=x}|X_t]|X_0, X_1, \ldots, X_t]$$

or

$$P(X_{t+1}|X_0, X_1, \ldots, X_t) = P(X_{t+1}|X_t),$$

since $x$ is arbitrary, i.e., $X$ is a Markov process.
4. DTMCs again

A fair coin is flipped till three consecutive heads are seen. What is the expected number of coin flips? Start your proof by first drawing the associated DTMC.

**Hint:** Define a DTMC $X$ as follows: $X_k$ takes one of the four values $\{0, 1, 2, 3\}$. $X_k$ keeps track of the latest consecutive streak of heads seen up to time $k$. For example, when $X_k = 2, X_{k+1} = 3$ with probability $1/2$ (if the $(k + 1)$th coin flip results in heads) and $X_{k+1} = 0$ with probability $1/2$ (if the $(k + 1)$th coin flip results in tails). Compute the corresponding hitting time in this DTMC.

**Solution**

The DTMC is:

![Figure 1: DTMC](image)

Let $T_{i3}$ be the expected time to hit state 3 starting from state $i \in \{0, 1, 2, 3\}$. By Fig. 1, we have:

$$T_{03} = \frac{1}{2} (1 + T_{03}) + \frac{1}{2} (1 + T_{13}),$$

which results in

$$T_{03} = 2 + T_{13}. \quad (1)$$

Similarly,

$$T_{13} = \frac{1}{2} (1 + T_{03}) + \frac{1}{2} (1 + T_{23}),$$

i.e.,

$$T_{13} = 1 + \frac{1}{2} T_{03} + \frac{1}{2} T_{23} \quad (2)$$

and

$$T_{23} = \frac{1}{2} (1 + T_{03}) + \frac{1}{2},$$

i.e.,

$$T_{23} = 1 + \frac{1}{2} T_{03}. \quad (3)$$

Combining (1) and (2), we obtain:

$$T_{03} = 6 + T_{23}$$
and then using (3), we finally have:

\[ T_{03} = 14. \]

5. Martingales and Stopping Times

Given a discrete time stochastic process \( X = \{X_n : n \geq 0\} \), a random time is a discrete random variable on the same probability space as \( X \) taking values in \( \mathbb{N} = \{0, 1, 2, \ldots\} \), such that \( X_T \) denotes the state at random time \( T \); if \( T = n \), then \( X_T = X_n \). If the event \( \{T = n\} \) is determined by (at most) the information \( \{X_0, X_1, \ldots, X_n\} \) known up to time \( n \), then \( T \) is a stopping time.

Let \( \{M_n : n \geq 0\} \) be a martingale with respect to \( \{Z_n : n \geq 0\} \). An important property of martingales is that \( \mathbb{E}[M_n] = \mathbb{E}[M_0] \), \( \forall n \geq 0 \). Assume now that \( T \) is a bounded random variable, i.e., \( P(T \leq M) = 1 \) for some \( M < \infty \), which is a stopping time with respect to \( \{Z_n : n \geq 0\} \). Show that \( \mathbb{E}[M_T] = \mathbb{E}[M_0] \).

**Hint:** Express \( M_T \) using the martingale difference sequence \( \{D_1, D_2, \ldots\} \), where \( D_i = M_i - M_{i-1} \). More specifically, start by expressing \( M_T \) as follows:

\[ M_T = M_0 + \sum_{i=1}^{M} D_i I(T \geq i). \]

**Solution**

By taking the expectation to both sides of the provided expression in the hint we have:

\[
\mathbb{E}[M_T] = \mathbb{E}[M_0] + \sum_{i=1}^{M} \mathbb{E}[D_i I(T \geq i)]
\]

\[
= \mathbb{E}[M_0] + \sum_{i=1}^{M} \mathbb{E}[\mathbb{E}[D_i I(T \geq i)|Z_0, Z_1, \ldots, Z_{i-1}]]
\]

\[
\leq \mathbb{E}[M_0] + \sum_{i=1}^{M} \mathbb{E}[I(T \geq i)\mathbb{E}[D_i|Z_0, Z_1, \ldots, Z_{i-1}]]
\]

\[ \overset{(a)}{=} \mathbb{E}[M_0] \]

\[ \overset{(b)}{=} \mathbb{E}[M_0] \]

- In (a) we use the fact that since \( T \) is a stopping time, the event \( \{T \geq i\} \) is determined by the knowledge of \( Z_0, \ldots, Z_{i-1} \).
In (b) we use the definition of a martingale sequence with respect to some other sequence to get \( E[D_i | Z_0, Z_1, \ldots, Z_{i-1}] = 0 \). This observation has been also used in the proofs of Azuma-Hoeffding and McDiarmid’s inequality taught in class.

6. Erdős-Rényi graphs and Concentration

Consider again the ensemble \( G(n, p) \) of Erdős-Rényi graphs in Problem 3 of HW#1.

(a) Let \((A_1, \ldots, A_m)\) be a partition of the edge set of the complete graph with \( n \) vertices, usually denoted by \( K_n \). Define the graph function \( f \) such that

\[
|f(G) - f(G')| \leq 1 \quad \text{whenever the symmetric difference } \mathcal{E}(G) \triangle \mathcal{E}(G') = (\mathcal{E}(G) \setminus \mathcal{E}(G')) \cup (\mathcal{E}(G') \setminus \mathcal{E}(G)) \text{ of the edge sets of } G \text{ and } G' \text{ is contained in a single set } A_k \text{ for some } k \in \{1, 2, \ldots, m\}.
\]

Then for a graph \( G_{n,p} \) drawn from \( G_{n,p} \) the random variable \( Z = f(G_{n,p}) \) satisfies:

\[
P(Z - E[Z] \geq t) \leq e^{-\frac{2t^2}{m}}, \quad \forall t \geq 0.
\]

Solution

Follows from McDiarmid’s inequality for \( c_i = 1, \forall i \).

(b) Two edges are said to be adjacent if they contain a common endpoint. Three edges form a triangle if, together, they contain only three distinct vertices. For a graph \( G = (V, E) \), the subset \( \mathcal{T} \) of \( E \times E \times E \) containing all possible triangles has size \( N = \binom{n}{3} \). Let \( \{X_e \sim \text{Ber}(p)\} \) be a set of i.i.d. Bernoulli variables representing the existence or not of an edge in an Erdős-Rényi graph. The number of triangles in a graph \( G_{n,p} \) drawn from \( G_{n,p} \) is

\[
f(G_{n,p}) = \sum_{\{e_1, e_2, e_3\} \in \mathcal{T}} X_{e_1}X_{e_2}X_{e_3}.
\]

i. What is the expected number of triangles in \( G_{n,p} \)?

ii. Give an upper bound for \( P(|f(G_{n,p}) - E[f(G_{n,p})]| \geq t) \). Based on this bound, assume that we wish to show

\[
P(|f(G_{n,p}) - E[f(G_{n,p})]| \geq \frac{1}{2} E[f(G_{n,p})]) \to 0.
\]

For what values of \( p \) depending on \( n \) is this possible?
Solution

(i) The expected number of triangles is \(Np^3\).

(ii) From the two-sided version of McDiarmid’s inequality, we have:

\[
P(|f(G_{n,p}) - E[f(G_{n,p})]| \geq t) \leq 2 \exp \left( -\frac{2t^2}{m(n-2)^2} \right),
\]

where \(m = \binom{n}{2}\) is the number of pairs of distinct vertices or equivalently the number of possible edges.

Explanation: A change in a single \(x_e\) (\(x_e\) is the realization of \(X_e\)) has a significant impact on \(f(G_{n,p})\). If \(x_e = 1\) for all \(e \in E\) and just one \(x_e\) is changed to 0, then \(n - 2\) triangles disappear. This shows that the \(c_i's\) in McDiarmid’s inequality are all equal to \(n - 2\). For

\[
P(|f(G_{n,p}) - E[f(G_{n,p})]| \geq \frac{1}{2} E[f(G_{n,p})])
\]

we have that \(N\) grows as \(n^3\). Therefore, \(t\) grows as \(n^3p^3\) (part (i)). Also, \(m\) grows as \(n^2\). Thus, the exponent becomes \(O(n^2p^6)\). We need \(n^{1/3}p \to \infty\) to make the above probability go to zero.

7. Erdős-Rényi graphs and Doob Martingales

We have given in the class notes the construction resulting to the so-called Doob martingale. Let’s revisit it here. Consider a set of random variables \(X = \{X_1, X_2, \ldots, X_n\}\), each one taking values in some set \(S\) and consider a function \(f : S^n \to \mathbb{R}\). Define the random variables

\[
Z_i = E[f(X_1, \ldots, X_n)|X_1, \ldots, X_i].
\]

Then \(\{Z_i\}\) is always a martingale, regardless of the properties of \(X\) and it is called the Doob martingale for \(f\).

Let now \(f\) be a function on graphs with \(n\) vertices. Let \(m = \binom{n}{2}\) be the number of edges in the corresponding complete graph, denoted by \(K_n\), and assume that \(e_1, e_2, \ldots, e_m\) is an arbitrary labeling (and ordering) of these edges. For a graph \(G \in \mathcal{G}_{n,p}\) (\(\mathcal{G}_{n,p}\) is the Erdős-Rényi ensemble), let

\[
Z_i = E[f(G)|X_1, X_2, \ldots, X_i], \quad i \in \{0, 1, \ldots, m\},
\]
where \( X_i = \mathbb{1}\{e_i \in G\} \) and \( i = 0 \) corresponds to no conditioning. In other words, \( Z_i \) corresponds to the expectation of \( f \) over all graphs \( G \in \mathcal{G}_{n,p} \), which agree with \( G \) on \( \{e_1, e_2, \ldots, e_i\} \). Clearly, \( \{Z_i\} \) is a martingale and it is called \textit{edge exposure martingale} because the edges of \( G \) are exposed sequentially in the conditioning. Similarly, let \( m = n - 1 \) and \( v_1, v_2, \ldots, v_n \) be an arbitrary labeling (and ordering) of the vertices of \( K_n \). Define

\[
Z'_i = E[f(G)|G|v_1,\ldots,v_{i+1}], \quad i \in \{0, 1, \ldots, m\},
\]

where conditioning on \( G|v_1,\ldots,v_{i+1} \) is interpreted as conditioning on graphs which agree with \( G \) on \( \{v_1, \ldots, v_{i+1}\} \). In other words, in step \( i \) we reveal all the edges between the vertex \( v_{i+1} \) and its predecessors \( \{v_1, \ldots, v_i\} \). Again, \( \{Z'_i\} \) is a martingale and it is called \textit{vertex exposure martingale}. Note that in both the above martingales \( Z_0 = E[f(G)] \) (and similarly for \( Z'_0 \)) and \( Z_m = f(G) \) (and similarly for \( Z'_m \)).

**Definition:** A proper graph coloring of a graph \( \Gamma = (V, E) \) using \( k \) colors is a mapping \( \phi : V \to \{1, 2, \ldots, k\} \) that maps neighboring vertices (i.e., vertices joined by an edge) to distinct colors:

\[
\phi(u) \neq \phi(v), \quad \forall (u, v) \in E.
\]

The smallest number \( k \) of colors resulting in a proper coloring is called \textit{chromatic number} of \( \Gamma \) and is denoted by \( \chi(\Gamma) \).

Show that for \( G \sim \mathcal{G}_{n,p} \),

\[
P(|\chi(G) - E[\chi(G)]| > t\sqrt{n}) \leq 2e^{-\frac{t^2}{2}}.
\]

by following the steps:

(a) (Bonus part) Consider the vertex exposure martingale. Argue that \( |Z'_i - Z'_{i-1}| \leq 1 \) for all \( i \).

(b) Using the previous part, show the desired result.

**Solution**

(b) Follows directly from the Azuma-Hoeffding inequality, by using part (a).

8. Poisson Processes

Assume that tourists arrive at the port of a small Greek island in the Aegean Sea as a Poisson process with rate \( \lambda \).
(a) The only ferry departs after a deterministic time $T$. Let $W$ be the total waiting time for all tourists. Compute $E[W]$.

**Solution**

Let $T_1, T_2, \ldots$ be the arrival times of the tourists in $[0, T]$. The combined waiting time is $W = (T - T_1) + (T - T_2) + \cdots$. Let $N(t)$ be the number of arrivals in $[0, t]$. Conditioning with respect to $N(T)$, we have:

$$E[W] = \sum_{k=0}^{\infty} E[W | N(T) = k] P(N(T) = k)$$

$$= \sum_{k=0}^{\infty} k \frac{T}{2} P(N(T) = k) = \frac{T}{2} \sum_{k=0}^{\infty} k P(N(T) = k) = \frac{\lambda T^2}{2},$$

where we use the fact that conditioning on $N(T) = k$ makes the arrival times i.i.d. Unif$[0, T]$ random variables.

(b) Suppose now that two ferries depart, one at $T$ and one at $S < T$. What is now $E[W]$?

**Solution**

We have two independent Poisson processes in $[0, S]$ and $[S, T]$ resulting in $N(T) = N(S) + N(T - S)$. Therefore,

$$E[W] = \lambda S^2 + \lambda \frac{(T - S)^2}{2}.$$

(c) What is $E[W]$, if $T \sim \text{Exp}(\mu)$, independent of the arrivals of the tourists?

**Solution**

In this case

$$E[W] = \int_{0}^{\infty} E[W | T = t] f_T(t) dt$$

$$= \int_{0}^{\infty} \frac{\lambda t^2}{2} f_T(t) dt = \frac{\lambda}{2} E[T^2] = \frac{\lambda}{\mu^2}.$$
9. Poisson Processes Again

Let \( N(t) \) be a Poisson process with rate \( \lambda \).

(a) Obtain \( P(N(3) = 5) \).

Solution
\[
N(3) \sim \text{Poi}(\lambda(3 - 0)) = \text{Poi}(3\lambda).
\]
Thus, \( P(N(3) = 5) = e^{-3\lambda(3\lambda)^5} \).

(b) Obtain \( P(N(7) - N(4) = 5) \) and \( E[N(7) - N(4)] \).

Solution
\[
N(7) - N(4) \sim \text{Poi}(3\lambda).
\]
Thus, \( P(N(7) - N(4) = 5) = e^{-3\lambda(3\lambda)^5} \).

Also,
\[
E[N(7) - N(4)] = 3\lambda.
\]

(c) Obtain \( P(N(7) - N(4) = 5 | N(6) - N(4) = 2) \).

Solution
Split the interval \([4, 7]\) into \([4, 6]\) and \([6, 7]\). These two intervals are disjoint and therefore \( N_7 - N_6 \sim \text{Poi}(\lambda) \) is independent of \( N_6 - N_4 \sim \text{Poi}(2\lambda) \). Thus,
\[
P(N(7) - N(4) = 5 | N(6) - N(4) = 2) = P(N(7) - N(6) = 3) = e^{-\lambda\lambda^3}.
\]

(d) Obtain \( P(N(6) - N(4) = 2 | N(7) - N(4) = 5) \).

Solution
\[
P(N(6) - N(4) = 2 | N(7) - N(4) = 5) = \frac{P(N(6) - N(4) = 2, N(7) - N(4) = 5)}{P(N(7) - N(4) = 5)}
= \frac{P(N(6) - N(4) = 2, N(7) - N(6) = 3)}{P(N(7) - N(4) = 5)}
= \frac{P(N(6) - N(4) = 2)P(N(7) - N(6) = 3)}{P(N(7) - N(4) = 5)}
= \frac{(e^{-2\lambda(2\lambda)^2})(e^{-\lambda\lambda^3})}{e^{-3\lambda(3\lambda)^5}}.
\]

10. Brownian Motion

Let’s revisit the definition of standard Brownian motion:

**Definition**: A real-valued stochastic process \( \{W(t) : t \geq 0\} \) is called a (linear) **Brownian motion** with start in \( x \in \mathbb{R} \) if:

(a) \( W(0) = x \).

(b) The process has *independent increments*, i.e., for all \( 0 \leq t_1 \leq t_2 \leq \cdots \leq t_n \), the increments \( W(t_n) - W(t_{n-1}), \ldots, W(t_2) - W(t_1) \) are independent random variables.

(c) For all \( t \geq 0 \) and \( h > 0 \), the increments are distributed as follows: \( W(t + h) - W(t) \sim \mathcal{N}(0, h) \).

(d) Almost surely, the function \( t \to W(t) \) is continuous.
We say that \( \{W(t) : t \geq 0\} \) is a **standard Brownian motion** if \( x = 0 \).

Suppose that \( \{W(t) : t \geq 0\} \) is a standard Brownian motion. Define the process

\[
X(t) = \begin{cases} 
0, & t = 0 \\
\frac{tW(1/t)}{t}, & t > 0
\end{cases}.
\]

Show that \( X(t) \) is also a standard Brownian motion.

**Hint**: You need to use the fact that a Brownian motion is a Gaussian process and the corresponding finite dimensional distributions are totally characterized by the mean and autocovariance functions. See the class notes. You then need to verify that \( X(t) \) satisfies all the properties of the provided definition.

**Solution**

By the lecture notes, \( W(t_1), ..., W(t_n) \) is a Gaussian vector characterized by

\[
E[W(t_i)] = 0 \quad \text{and} \quad \text{Cov}(W(t_i), W(t_j)) = \min(t_i, t_j) = t_i,
\]

if \( 0 \leq t_i \leq t_j \). Clearly, \( X(t) \) is also a Gaussian process with \( E[X(t_i)] = 0 \) and

\[
\text{Cov}(X(t+s), X(t)) = (t+s)s\text{Cov} \left( W \left( \frac{1}{t+s} \right), W \left( \frac{1}{t} \right) \right) = t(t+s) \frac{1}{t+s} = t.
\]

Therefore, all the finite dimensional vectors \( (X(t_1), ..., X(t_n)) \) of \( X(t) \) are distributed as in a Brownian motion. The paths \( t \to X(t) \) are clearly continuous. Also, \( X(t) \to 0 \) as \( t \to 0 \) by the continuity of the sample paths.