Homework #4 Solutions

1. **Convergence to a random variable.** Consider a coin with random bias $P \sim F_P(p)$. Flip the coin $n$ times independently to generate $X_1, X_2, \ldots, X_n$, where $X_i = 1$ if the $i$-th outcome is heads and $X_i = 0$ otherwise. Let $S_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ be the sample average. Show that $S_n$ converges to $P$ in mean square.

**Solution (15 points)**

\[
E((S_n - P)^2) = E_P(E((S_n - P)^2|P))
= E_P(Var(Sn|P))
= E_P\left(\frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i|P\right)\right)
= E_P\left(\frac{1}{n^2}(nP(1-P))\right)
= \frac{1}{n}(E(P) - E(P^2)).
\]

Therefore $\lim_{n \to \infty} E((S_n - P)^2) = 0$ and $S_n$ converges to $P$ in mean square.

2. **Convergence in probability implies convergence in distribution.** Prove that convergence in probability implies convergence in distribution following these steps:

a. Show that for any random variable $A$, any real number $a$ and any $\epsilon > 0$,
\[
P(A \leq a - \epsilon) - P(|A_n - A| > \epsilon) \leq P(A_n \leq a) \leq P(A \leq a + \epsilon) + P(|A_n - A| > \epsilon)
\]

b. Assume that $A_n \to A$ in probability. Use the expression derived in (a) to prove that $A_n \to A$ in distribution. Hint: Remember that $A_n \to A$ in distribution if $\lim_{n \to \infty} F_{A_n}(a) = F_A(a)$ for points $a$ at which $F_A$ is continuous, which implies
\[
\lim_{\epsilon \to 0} F_A(a \pm \epsilon) = F_A(a).
\]

**Solution**

a. If $A_n > a$ and $|A_n - A| \leq \epsilon$ then $A > a - \epsilon$, so that
\[
\{A_n > a\} \cap \{|A_n - A| \leq \epsilon\} \subseteq \{A > a - \epsilon\}.
\]

For any sets $S_1$ and $S_2$, $S_1 \subseteq S_2$ implies $S_2^c \subseteq S_1^c$, so the above equation becomes
\[
\{A \leq a - \epsilon\} \subseteq \{A_n \leq a\} \cup \{|A_n - A| > \epsilon\},
\]
after applying the De Morgan’s laws. Finally, applying the union bound we obtain
\[ P(A \leq a - \epsilon) \leq P(A_n \leq a) + P(|A_n - A| > \epsilon). \]
This proves the left inequality. For the right inequality, note that if \( A > a + \epsilon \) and \(|A_n - A| \leq \epsilon\) then \( A_n > a\). Following the same reasoning as before, this implies
\[ \{A_n \leq a\} \subseteq \{A \leq a + \epsilon\} \cup \{|A_n - A| > \epsilon\}, \]
which again by the union bound allows us to conclude
\[ P(A_n \leq a) \leq P(A \leq a + \epsilon) + P(|A_n - A| > \epsilon). \]

b. Taking the limit as \( n \to \infty \) in the expression obtained in (a) and using the fact that since \( A_n \to A \) in probability \( \lim_{n \to \infty} P(|A_n - A| > \epsilon) = 0 \) yields
\[ P(A \leq a - \epsilon) \leq \lim_{n \to \infty} P(A_n \leq a) \leq P(A \leq a + \epsilon). \]

Now, we take the limit as \( \epsilon \) tends to zero and assume that \( F_A \) is continuous at \( a \) to obtain
\[ F_A(a) = P(A \leq a) \leq \lim_{n \to \infty} F_{A_n}(a) \leq P(A \leq a) = F_A(a). \]
This implies that for any \( a \) such that \( F_A \) is continuous at \( a \) \( \lim_{n \to \infty} F_{A_n}(a) = F_A(a) \), which establishes convergence in distribution.

3. Convergence experiments. The purpose of this problem is to demonstrate different types of convergence and laws of large numbers.
a. Use MATLAB to generate 200 samples \( X_1, \ldots, X_{200} \) of i.i.d.
zero mean, unit variance Gaussian random variables. Plot the sample average sequence \( S_n = \frac{1}{n} \sum_{i=1}^{n} X_i \) as a function of \( n \). Hint: use \texttt{randn} and \texttt{cumsum}.
b. Generate 5000 sequences each consisting of 200 samples of i.i.d.
zero mean, unit variance Gaussian random variables:
\[ X_{i,1}, X_{i,2}, \ldots, X_{i,200}, \quad i = 1, 2, \ldots, 5000 \]
Compute the sample average sequences \( S_{i,1}, S_{i,2}, \ldots, S_{i,200} \) for \( i = 1, \ldots, 5000 \).
You can generate a 5000\( \times \)200 matrix of i.i.d. Gaussians using the \texttt{randn} command.
c. Strong law of large numbers. To explore convergence with probability 1, let \( N_m \) be the number of \( S_{i,n} \) sequences with \(|S_{i,n} - E(X)| > 0.1\) for some \( n > m \). Use \( E_m = N_m/5000 \) as an estimate of the probability \( P_m = P(|S_n - E(X)| > 0.1\) for some \( n > m \). \)
Plot \( E_m \) vs. \( m \) for 1 \( \leq \) \( m \) \( \leq \) 200.
d. Mean square convergence. To explore convergence in mean square, compute
\[ M_n = \frac{1}{5000} \sum_{i=1}^{5000} (S_{i,n} - E(X))^2, \]
which are estimates of the mean square. Plot \( M_n \) vs. \( n \) for 1 \( \leq \) \( n \) \( \leq \) 200.
e. **Weak law of large numbers.** To explore convergence in probability, let $N_n$ be the number of $S_{i,n}$ sequences for which $|S_{i,n} - E(X)| > 0.1$. Use $E_n = N_n/5000$ as an estimate of the probability $P_n = P\{|S_n - E(X)| > 0.1\}$. Plot $E_n$ and $P_n$ vs. $n$ for $1 \leq n \leq 200$.

**SOLUTION (15 points)**

The MATLAB code is below. The output is shown in Figure [1](#).

```matlab
clear all;
cf;

% Part (a)
% Generate 200 samples (X_1 to X_200) of i.i.d. zero
% mean, unit variance Gaussian random variables.
% Hint: Use randn and cumsum.

n = 1:200;

% WRITE MATLAB CODE HERE
X = randn(1, 200);
S = cumsum(X);

% Do the divide by n part.
S = S./n;

subplot(4,1,1);
plot(n, S);
xlabel('n');
ylabel('Sn');
title('2(a) Sample average sequence');

% Part (b)

% Now generate 5000 such sequences.
% Hint: use randn and cumsum (be careful!) again

% WRITE MATLAB CODE HERE
X = randn(5000, 200);
S = cumsum(X, 2);
S = S./repmat(n, 5000, 1);

% Part (c) Strong Law of Large Numbers (this loop will run for a minute)

E_m = zeros(200,1);
for m = 0:199,

% N_m should be the number of rows in S that have an entry whose
% absolute value is > 0.1 in columns m+1 through 200.

```
% WRITE MATLAB CODE HERE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
N_m = sum(sum((abs(S(:, m+1:200)) > 0.1), 2) > 0);
E_m(m+1) = N_m/5000;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
end;
subplot( 4, 1, 2 );
plot( E_m );
xlabel( 'n' );
ylabel( 'E_m' );
title( '2(c) Strong Law of Large Numbers' );

% Part (d) Convergence in Mean Square

% S_squared is the square of the S matrix. EX = 0.
S_squared = S.^2;

% WRITE MATLAB CODE HERE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
M = 1/5000*sum( S_squared );
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
subplot( 4, 1, 3 );
plot( n, M );
xlabel( 'n' );
ylabel( 'Mn' );
title( '2(d) Mean square convergence' );

% Part (e) Weak Law of Large Numbers

% Count the number of times |S_i,n| > 0.1 in each column.

% WRITE MATLAB CODE HERE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
N = sum( abs( S ) > 0.1 );
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
E = N/5000;

% Plot Pn and En vs. n. Since EX = 0, Pn is the probability that |Sn| > 0.1. Sn has zero mean and a standard deviation sigma_Sn. Hint: sigma_Sn is a function of n.
% Therefore P(|Sn| > 0.1) = 2*Q(0.1/sigma_Sn).
% erf( x ) = 2/sqrt( pi )* integral from x to inf of exp( -t^2 ) dt
% So Q( x ) = 1/2*erfc( x/sqrt( 2 ) );

% Find sigma_Sn and Pn. Hint: Pn and En look quite similar.

% WRITE MATLAB CODE HERE
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
sigma_Sn = 1./sqrt(n);
P = erfc( ( 0.1/sigma_Sn )/sqrt(2) );
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```matlab
subplot( 4, 1, 4 );
plot( n, P, 'r--' );
hold on;
plot( n, E );
axis( [ 0 200 0 1 ] );
xlabel( 'n' );
ylabel( 'E_n (solid), P_n (dashed)' );
title( '2(e) Convergence in probability' );

% Produce hardcopy
orient tall
print hw6q7
```

Figure 1: Output of convergence experiments
4. *Minimum waiting time.* Let \( X_1, X_2, \ldots \) be i.i.d. exponentially distributed random variables with parameter \( \lambda \), i.e. \( f_{X_i}(x) = \lambda e^{-\lambda x}, \) for \( x \geq 0 \).

(a) Does \( Y_n = \min\{X_1, X_2, \ldots, X_n\} \) converge in probability as \( n \) approaches infinity?
(b) If it converges what is the limit?
(c) What about \( Z_n = nY_n \), does it converge in probability? In distribution?

**Solution:**

(a) For any set of values \( X_i \)'s, the sequence \( Y_n \) is monotonically decreasing in \( n \). Since the random variables are non negative, it is reasonable to guess that \( Y_n \) converges to 0. Now \( Y_n \) will converge in probability to 0 if and only if for any \( \epsilon > 0 \), \( \lim_{n \to \infty} P\{|Y_n| > \epsilon\} = 0 \).

\[
P\{|Y_n| > \epsilon\} = P\{Y_n > \epsilon\} = P\{X_1 > \epsilon, X_2 > \epsilon, \ldots, X_n > \epsilon\} = P\{X_1 > \epsilon\}P\{X_2 > \epsilon\} \cdots P\{X_n > \epsilon\} = (1 - F_X(\epsilon))(1 - F_X(\epsilon)) \cdots (1 - F_X(\epsilon)) = (1 - (1 - e^{-\lambda \epsilon}))^n = e^{-\lambda\epsilon n}.
\]

As \( n \) goes to infinity (for any finite \( \epsilon > 0 \)) this converges to zero. Therefore \( Y_n \) converges to 0 in probability.

(b) The limit to which \( Y_n \) converges in probability is 0.

(c) Does \( Z_n = nY_n \) converges to 0 in probability? No, it does not. In fact,

\[
P\{|Z_n| > \epsilon\} = P\{nY_n > \epsilon\} = P\{Y_n > \frac{\epsilon}{n}\} = e^{-\lambda\frac{\epsilon}{n}} = e^{-\lambda \epsilon / n}
\]

which does not depend on \( n \). So \( Z_n \) does not converge to 0 in probability. Note that the distribution of \( Z_n \) is exponential with parameter \( \lambda \), the same as the distribution of \( X_i \).

\[
F_{Z_n}(z) = P\{Z_n < z\} = 1 - e^{-\lambda z}.
\]

In conclusion, if \( X_i \)'s are i.i.d. \( \sim \exp(\lambda) \), then

\[
Y_n = \min_{1 \leq i \leq n}\{X_i\} \sim \exp(n\lambda)
\]

and

\[
Z_n = nY_n \sim \exp(\lambda).
\]

Thus

\[
P\{Y_n > \epsilon\} = e^{-\lambda\epsilon n} \to 0, \text{ so } Y_n \to 0 \text{ in probability},
\]

but

\[
P\{Z_n > \epsilon\} = e^{-\lambda \epsilon / n} \not\to 0, \text{ so } Z_n \not\to 0 \text{ in probability}.
\]
5. **Nonlinear estimator.** Consider a channel with the observation \( Y = XZ \), where the signal \( X \) and the noise \( Z \) are uncorrelated Gaussian random variables. Let \( E[X] = 1, E[Z] = 2, \sigma_X^2 = 5 \), and \( \sigma_Z^2 = 8 \).

(a) Using the fact that \( E(W^3) = \mu^3 + 3\mu\sigma^2 \) and \( E(W^4) = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4 \) for \( W \sim \mathcal{N}(\mu, \sigma^2) \), find the mean and covariance matrix of \( X \).

(b) Find the MMSE linear estimate of \( X \) given \( Y \) and the corresponding MSE.

(c) Find the MMSE linear estimate of \( X \) given \( Y^2 \) and the corresponding MSE.

(d) Find the MMSE linear estimate of \( X \) given \( Y \) and \( Y^2 \) and the corresponding MSE.

(e) Compare your answers in parts (b) through (d). Is the MMSE estimate of \( X \) given \( Y \) (namely, \( E(X|Y) \)) linear?

**Solution:**

(a) Since \( X \) and \( Z \) are uncorrelated Gaussian random variables, they are independent. We have

\[
E(X^2) = \sigma_X^2 + E^2(X) = 5 + 1 = 6, \\
E(X^3) = 1 + 3 \times 1 \times 5 = 16, \\
E(X^4) = 1 + 6 \times 1 \times 5 + 3 \times 25 = 106.
\]

\[
E(Z^2) = \sigma_Z^2 + E^2(Z) = 8 + 4 = 12, \\
E(Z^3) = 8 + 3 \times 2 \times 8 = 56, \\
E(Z^4) = 2^4 + 6 \times 4 \times 8 + 3 \times 64 = 400.
\]

Since \( X \) and \( Z \) are independent, we have

\[
E(Y) = E(XZ) = E(X)E(Z) = 2, \\
E(Y^2) = E(X^2Z^2) = E(X^2)E(Z^2) = 6 \times 12 = 72, \\
E(Y^3) = E(X^3Z^3) = E(X^3)E(Z^3) = 16 \times 56 = 896, \\
E(Y^4) = E(X^4)E(Z^4) = 106 \times 400 = 42400.
\]

Therefore, the mean of \( [X \ Y \ Y^2]^T \) is \( [1 \ 2 \ 72]^T \).

\[
\text{Var}(Y) = E(Y^2) - E^2(Y) = 72 - 4 = 68, \\
\text{Var}(Y^2) = E(Y^4) - E^2(Y^2) = 37216, \\
\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X^2)E(Z) - E(X)E(Y) = 10, \\
\text{Cov}(X, Y^2) = E(XY^2) - E(X)E(Y^2) = E(X^3)E(Z^2) - E(X)E(Y^2) = 120, \\
\text{Cov}(Y, Y^2) = E(Y^2) - E(Y)E(Y^2) = E(X^3)E(Z^3) - E(Y)E(Y^2) = 752.
\]
Therefore, the covariance matrix of \([X \ Y \ Y^2]^T\) is
\[
\begin{bmatrix}
5 & 10 & 120 \\
10 & 68 & 752 \\
120 & 752 & 37216
\end{bmatrix}.
\]

(b) The MMSE linear estimate of \(X\) given \(Y\) is
\[
\hat{X} = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)} (Y - \text{E}(Y)) + \text{E}(X) = \frac{10}{68} (Y - 2) + 1 = \frac{5}{34} Y + \frac{24}{34},
\]
and its MSE is given by
\[
\text{MSE} = \text{Var}(X) - \frac{\text{Cov}^2(X, Y)}{\text{Var}(Y)} = 5 - \frac{100}{68} = 3.5294.
\]

(c) The MMSE linear estimate of \(X\) given \(Y^2\) is
\[
\hat{X} = \frac{\text{Cov}(X, Y^2)}{\text{Var}(Y^2)} (Y^2 - \text{E}(Y^2)) + \text{E}(X) = \frac{120}{37216} (Y^2 - 72) + 1 = \frac{15}{4652} Y^2 + \frac{893}{1163},
\]
and its MSE is given by
\[
\text{MSE} = \text{Var}(X) - \frac{\text{Cov}^2(X, Y^2)}{\text{Var}(Y^2)} = 5 - \frac{14400}{37216} = 4.6131.
\]

(d) We first normalize the random variables by subtracting off their means to get
\[
X' = X - \text{E}(X) = X - 2,
\]
\[
Y' = Y - \text{E}(Y) = Y - 2,
\]
\[
Y'^2 = Y^2 - \text{E}(Y^2) = Y^2 - 72.
\]

Using the covariance matrix in part a, we have
\[
\Sigma_{[Y \ Y^2]X} = \begin{bmatrix} 10 & 120 \end{bmatrix}^T,
\]
\[
\Sigma_{[Y \ Y^2]^T} = \begin{bmatrix} 68 & 752 \\ 752 & 37216 \end{bmatrix}.
\]

Therefore,
\[
\hat{X}' = \Sigma_{[Y \ Y^2]^T}X \Sigma_{[Y \ Y^2]^T}^{-1} \begin{bmatrix} Y' \\ Y'^2 \end{bmatrix} = 0.1435Y' + 0.0003Y'^2,
\]
and hence
\[
\hat{X} = \hat{X}' + \text{E}(X) = 0.1435(Y - 2) + 0.0003(Y^2 - 72) + 1 = 0.1435Y + 0.0003Y^2 + 0.6896.
\]
The corresponding MSE is given by
\[
\text{MSE} = \text{Var}(X) - \Sigma_{[Y \ Y^2]^T}X \Sigma_{[Y \ Y^2]^T}^{-1} \Sigma_{[Y \ Y^2]^T}X = 3.526.
\]

(e) MSE linear estimate of \(X\) given \(Y\) and \(Y^2\) results in the minimum MSE among the three. Therefore, MSE linear estimate of \(X\) given \(Y\) does not have the minimum MSE and MMSE estimate of \(X\) given \(Y\) is not linear.
6. Orthogonality. Let \( \hat{X} \) be the minimum MSE estimate of \( X \) given \( Y \).

(a) Show that for any function \( g(y) \), \( E((X - \hat{X})g(Y)) = 0 \), i.e., the error \( (X - \hat{X}) \) and \( g(Y) \) are orthogonal.

(b) Show that

\[
\text{Var}(X) = E(\text{Var}(X|Y)) + \text{Var}(\hat{X})
\]

Provide a geometric interpretation for this result.

Solution:

(a) We use iterated expectation and the fact that \( E(g(Y)|Y) = g(Y) \):

\[
E \left( (X - \hat{X})g(Y) \right) = E \left[ E \left( (X - \hat{X})g(Y)|Y \right) \right]
\]

\[
= E \left[ E \left( (X - E(X|Y))g(Y)|Y \right) \right]
\]

\[
= E \left( g(Y)E \left( (X - E(X|Y))|Y \right) \right)
\]

\[
= E \left( g(Y)(E(X|Y) - E(X|Y)) \right)
\]

\[
= 0.
\]

(b) First we write

\[
E(\text{Var}(X|Y)) = E(X^2) - E((E(X|Y))^2),
\]

and

\[
\text{Var}(E(X|Y)) = E((E(X|Y))^2) - (E(E(X|Y)))^2
\]

\[
= E((E(X|Y))^2) - (E(X))^2.
\]

Adding the two terms completes the proof.

Interpretation: If we view \( X, E(X|Y) \), and \( X - E(X|Y) \) as vectors with “norm” \( \sqrt{\text{Var}(X)} \), \( \sqrt{\text{Var}(E(X|Y))} \), and \( \sqrt{E(\text{Var}(X|Y))} \), respectively, then this result provides a “Pythagorean theorem”, where the signal, the error, and the estimate are the sides of a right triangle (estimate and error being orthogonal).