Solutions to Practice Midterm Examination (Winter 2017)

1. **MMSE estimation (40 pts).** Let \( X \sim N(\mu, P) \) be the input to the additive Gaussian noise channel with two outputs

\[
Y_1 = X + Z_1, \\
Y_2 = X + Z_2,
\]

where \( Z_1 \sim N(0, N_1) \) and \( Z_2 \sim N(0, N_2) \) are independent of each other and of \( X \).

(a) Find the estimate of \( X \) in the form

\[
\hat{X} = a_1 Y_1 + a_2 Y_2
\]

that minimizes the mean square error \( \mathbb{E}[(X - \hat{X})^2] \).

(b) Find the MSE of the estimate found in part (a).

(c) Find the estimate of \( X \) in the form

\[
\hat{X} = a_1 Y_1^2 + a_2 Y_2^2 + a_{12} Y_1 Y_2 + b_1 Y_1 + b_2 Y_2 + c
\]

that minimizes the MSE.

(d) Find the MSE of the estimate found in part (c).

**Solution:**

(a) Note that \((X, Y_1, Y_2)\) is a Gaussian random vector with mean \([m \ m \ m]^T\) and covariance matrix

\[
\begin{bmatrix}
P & P & P \\
P & P + N_1 & P \\
P & P & P + N_2
\end{bmatrix}.
\]

Let the optimal estimate be given by \( \hat{X} = a^* Y \), where \( a = [a_1^* \ a_2^*]^T \) and \( Y = [Y_1 \ Y_2]^T \). By the orthogonality principle, we have

\[
\mathbb{E}[Y(X - \hat{X})] = 0,
\]
or equivalently,
\[ E[XY] = E[YY^T a^*], \]
which implies that
\[ a^* = (E[YY^T])^{-1} E[XY]. \]
Substituting
\[ E[YY^T] = \begin{bmatrix} \mu^2 + P + N_1 & \mu^2 + P \\ \mu^2 + P & \mu^2 + P + N_2 \end{bmatrix} \]
and
\[ E[XY] = (\mu^2 + P) \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \]
we have
\[ a^* = \frac{(\mu^2 + P)}{(\mu^2 + P)(N_1 + N_2) + N_1 N_2} \begin{bmatrix} N_2 \\ N_1 \end{bmatrix}. \]
Thus, the optimal estimate is
\[ \hat{X} = \frac{(\mu^2 + P)N_2}{(\mu^2 + P)(N_1 + N_2) + N_1 N_2} Y_1 + \frac{(\mu^2 + P)N_1}{(\mu^2 + P)(N_1 + N_2) + N_1 N_2} Y_2. \]

(b) The MSE of the estimate found in the previous part is
\[ E[(X - \hat{X})^2] = E[ (X - a_1^*(X + Z_1) - a_2^*(X + Z_2))^2 ] \\
= (1 - a_1^* - a_2^*)^2 E[X^2] + a_1^2 E[Z_1^2] + a_2^2 E[Z_2^2] \\
= \frac{N_1 N_2 (\mu^2 + P)}{N_1 N_2 + (\mu^2 + P)(N_1 + N_2)}. \]

(c) Since \((X, Y_1, Y_2)\) is Gaussian, the MMSE estimate of \(X\) given \((Y_1, Y_2)\) is linear in \((Y_1, Y_2)\). Thus, the best quadratic estimate of \(X\) given \((Y_1, Y_2)\) is also, in fact, linear, and given by
\[ E[X \mid Y] = E[X] + \Sigma_{XY} \Sigma_Y^{-1} (Y - E[Y]). \]
Substituting the values \(E[X] = \mu, E[Y] = [\mu \ \mu^T], \Sigma_{XY} = [P \ P], \) and
\[ \Sigma_Y = \begin{bmatrix} P + N_1 & P \\ P & P + N_2 \end{bmatrix}, \]
we have
\[ \hat{X} = \frac{PN_2}{(N_1 + N_2)P + N_1 N_2} Y_1 + \frac{PN_1}{(N_1 + N_2)P + N_1 N_2} Y_2 + \frac{N_1 N_2 \mu}{(N_1 + N_2)P + N_1 N_2}. \]
(d) The MSE of the estimate found in the previous part is

\[ \mathbb{E}[\text{Var}[X \mid Y]] = P - \begin{bmatrix} P & P \\ P & P + N_1 \\ P & P + N_2 \end{bmatrix}^{-1} \begin{bmatrix} P \\ P \\ P \end{bmatrix} = \frac{PN_1N_2}{(N_1 + N_2)P + N_1N_2}. \]

2. Random number of coin flips (30 pts). Suppose that we flip a coin with bias \( p \) independently \( N \) times, where \( N \sim \text{Geom}(p) \) is random. Let \( X \) be the number of heads, that is, \( X \mid \{N = n\} \sim \text{Binom}(n, p) \).

(a) Find \( \mathbb{E}[X] \).

(b) Find \( \text{Var}(X) \).

(c) Find \( P\{X = 0\} \).

Solution:

(a) First, note that \( X \mid \{N = n\} \sim \text{Binom}(n, p) \). Hence, \( \mathbb{E}[X \mid N] = Np \), and \( \text{Var}(X \mid N) = Np(1 - p) \). To find the expectation, we use the law of iterated expectation:

\[ \mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid N]] = \mathbb{E}[Np] = \frac{1}{p} \cdot p = 1. \]

(b) To find the variance, we use the law of conditional variance:

\[ \text{Var}(X) = \mathbb{E}[\text{Var}(X \mid N)] + \text{Var}(\mathbb{E}[X \mid N]) = \mathbb{E}[Np(1 - p)] + \text{Var}(Np) = \frac{1}{p} \cdot p(1 - p) + p^2 \text{Var}(N) = (1 - p) + p^2 \cdot \frac{1 - p}{p^2} = 2(1 - p). \]

(c) To find \( P\{X = 0\} \), we use the law of iterated expectation again. Since \( X \mid \{N = n\} \sim \text{Binom}(n, p) \), it immediately follows that \( P\{X = 0 \mid N \} = \)
\[(1 - p)^N. \text{ Therefore,}
\]
\[
P\{X = 0\} = E[P\{X = 0 \mid N\}]
= E[(1 - p)^N]
= \sum_{n=1}^{\infty} (1 - p)^n p(1 - p)^{n-1}
= \frac{p}{1 - p} \sum_{n=1}^{\infty} (1 - p)^{2n}
= \frac{p}{1 - p} \frac{(1 - p)^2}{1 - (1 - p)^2}
= \frac{1 - p}{2 - p}.
\]

3. Convergence (10 pts). Suppose that \(X_1, X_2, \ldots\) are i.i.d. \text{Unif}[0,1] random variables. Find the limit
\[
\lim_{n \to \infty} \left( \prod_{i=1}^{n} X_i \right)^{1/n}
\]
in probability.

\textbf{Solution:} Let
\[
S_n := \left( \prod_{i=1}^{n} X_i \right)^{1/n}.
\]
We have
\[
\ln S_n = \frac{1}{n} \sum_{i=1}^{n} \ln X_i
= \frac{1}{n} \sum_{i=1}^{n} Y_i,
\]
where \(Y_i = \ln X_i\). Now, since \(X_1, X_2, \ldots\) are i.i.d., \(Y_1, Y_2, \ldots\) are i.i.d. as well, and by the weak law of large numbers, \(\ln S_n\) converges to
\[
E[Y_1] = E[\ln X_1]
= \int_{0}^{1} \ln x \, dx
= x \ln x - x \bigg|_{0}^{1}
= -1
\]
in probability. In fact, it is easy to see that $-Y_1, -Y_2, \ldots$ are i.i.d. $\text{Exp}(1)$ random variables. Now for every $\epsilon > 0$

$$\lim_{n \to \infty} P \{|\ln S_n - (-1)| < \epsilon\} = 1.$$  

This implies that for every $\delta > 0$

$$\lim_{n \to \infty} P \{|S_n - e^{-1}| < \delta\} = 1.$$  

Therefore, $S_n \to e^{-1}$ in probability.
Solutions to Practice Midterm Examination (Winter 2018)

1. **Inequalities (18 points).** Let $X$ be a random variable, with $E[X] = 0$ and $\text{Var}[X] = 1$. For each of the following pairs of quantities $A$ vs. $B$, indicate their relationship with $\leq$, $=$, or $\geq$. That is: $A \leq B$, $A = B$, or $A \geq B$. If the correct answer is $A = B$, no credit will be given for the other options. Justify your answers.

#### Justify your answers.

(a) (i) $E[e^{2X^4}]$ vs. $e^{2E[X^4]}$
(ii) $E\left[\frac{1}{X^4}\right]$ vs. $\frac{1}{E[X^4]}$
(iii) $P(X^4 \leq 4)$ vs. $P(X^4 \geq 4)$

Now let $X$ and $Y$ be random variables. Assume that $E[X] = E[Y] = 0$ and $\text{Var}[X] = \text{Var}[Y] = 1$. Follow the same instructions as in part (a).

(b) (i) $\text{Var}(X)$ vs. $E[\text{Var}(X|Y)]$
(ii) $E[(X - E[X|Y])(Y - E[X|Y])]$ vs. 0
(iii) $\text{Var}(X + Y) - 2\text{Cov}(X, Y)$ vs. 1

**Solution:**

(a) (i) $g(x) = e^x$ is convex. By Jensen’s inequality, $E[g(Y)] \geq g(E[Y])$. (We assume that $E[Y]$ and $E[g(Y)]$ are finite.) Setting $Y = 2X^4$, we get

$$E[e^{2X^4}] \geq e^{2E[X^4]} = e^{2E[X^4]}.$$

(ii) By Cauchy-Schwartz Inequality,

$$(E[XY])^2 \leq E[X^2]E[Y^2].$$

Substituting $X^2$ for $X$ and $1/X^2$ for $Y$ yields

$$\left(E\left[X^2 \cdot \frac{1}{X^2}\right]\right)^2 \leq E[X^4]E\left[\frac{1}{X^4}\right].$$

The left-hand side is equal to 1, so we conclude that

$$E\left[\frac{1}{X^4}\right] \geq \frac{1}{E[X^4]}.$$

Alternatively, note that $g(x) = 1/x$ is convex for $x > 0$, and apply Jensen’s Inequality to $Y = X^4$. 


By Chebyshev inequality, \( P(|X - E[X]| \geq aaX) \leq 1/a^2 \) for \( a > 1 \).
This implies \( P((X - E[X])^4 \geq (aaX)^4) \leq 1/a^2 \).
Here \( E[X] = 0 \) and \( \sigma_X = \sqrt{\text{Var}(X)} = 1 \), so \( P(X^4 \geq a^4) \leq 1/a^2 \).
Letting \( a = \sqrt{2} \), we get \( P(X^4 \geq 4) \leq 1/2 \).
Since \( P(X^4 \geq 4) + P(X^4 < 4) = 1 \), we conclude that
\[
P(X^4 < 4) \geq P(X^4 \geq 4).
\]

(b) (i) The law of conditional variance states
\[
\text{Var}(X) = E[\text{Var}(X|Y)] + \text{Var}(E[X|Y]).
\]
Since \( \text{Var}(E[X|Y]) \geq 0 \), it follows that
\[
\text{Var}(X) \geq E[\text{Var}(X|Y)].
\]

(ii) \( E[X|Y] \) is the MMSE estimate for \( X \) given \( Y \).
The estimation error \( X - E[X|Y] \) is orthogonal to any function \( g(Y) \).
Since \( Y - E[X|Y] \) is a function of \( Y \),
\[
E[(X - E[X|Y])(Y - E[X|Y])] = 0.
\]

(iii)
\[
\text{Var}(X + Y) = \text{Cov}(X + Y, X + Y)
= \text{Var}(X) + 2\text{Cov}(X, Y) + \text{Var}(Y).
\]
Therefore,
\[
\text{Var}(X + Y) - 2\text{Cov}(X, Y) = \text{Var}(X) + \text{Var}(Y) = 2.
\]
So,
\[
\text{Var}(X + Y) - 2\text{Cov}(X, Y) \geq 1.
\]

2. Big bank (22 points). You are an active customer at a big bank. The bank schedule calls for a different number of tellers at various times. The number of tellers when you visit each morning is \( N \), where \( N \sim \text{Geom}(p) \), i.e., \( P(N = n) = (1 - p)^{n-1}p \), for a fixed \( 0 < p < 1 \). (Since banking is a business, naturally the probability of there being \( n \) tellers decreases as a function of \( n \).)
You are always the first in line at the bank each morning, while the tellers are busy preparing to serve customers. The preparation times of the tellers, who are indexed by \( i = 1, 2, \ldots \), are independent exponentially distributed random variables \( X_i \sim \text{Exp}(\lambda) \), for all \( i \geq 1 \), and also independent of \( N \).

Let \( Y \) be your waiting time until a teller is ready to serve you.

(a) (7 points) Determine the pdf \( f_{Y|N}(y|n) \) of your waiting time given that the number of tellers is \( N = n \).

(b) (7 points) Determine the mean \( E[Y] \) of your waiting time.

(c) (8 points) Determine the variance \( \text{Var}(Y) \) of your waiting time.

Solution:

(a) Note that \( Y|\{N = n\} = \min(X_1, \ldots, X_n)|\{N = n\} \). Therefore

\[
P(Y > y|N = n) = P(X_1 > y, \ldots, X_n > y|N = n).
\]

Since \( X_1, \ldots, X_n \) are independent, and independent of \( N \),

\[
P(Y > y|N = n) = \prod_{i=1}^{n} P(X_i > y|N = n) = \prod_{i=1}^{n} P(X_i > y).
\]

Now \( X_i \sim \text{Exp}(\lambda) \), so \( P(X_i > y) = e^{-\lambda y} \), for \( i = 1, \ldots, n \). Therefore,

\[
P(Y > y|\{N = n\}) = e^{-n\lambda y}.
\]

This implies that \( Y|\{N = n\} \sim \text{Exp}(n\lambda) \). Therefore, the conditional pdf is given by

\[
f_{Y|N}(y|n) = \begin{cases} n\lambda e^{-n\lambda y} & y \geq 0 \\ 0 & \text{otherwise.} \end{cases}
\]

(b) From part (a), we know that \( Y|\{N = n\} \sim \text{Exp}(n\lambda) \), so

\[
E[Y|N = n] = \frac{1}{n\lambda}.
\]

Therefore,

\[
E[Y|N] = \frac{1}{N\lambda}.
\]
By the law of iterated expectation

\[ E[Y] = E[E[Y|N]] = E \left[ \frac{1}{N\lambda} \right] = \frac{1}{\lambda} E \left[ \frac{1}{N} \right]. \]

Since \( N \sim \text{Geom}(p) \),

\[ E \left[ \frac{1}{N} \right] = \sum_{n=1}^{\infty} \frac{1}{n} (1-p)^{n-1}p = \frac{p}{1-p} \sum_{n=1}^{\infty} \frac{1}{n} (1-p)^n. \]

Applying the fact that

\[ \sum_{n=1}^{\infty} \frac{a^n}{n} = -\log(1-a) \]

with \( a = 1-p \), we get

\[ E \left[ \frac{1}{N} \right] = \frac{p \log p}{1-p}. \]

So,

\[ E[Y] = \frac{p \log p}{\lambda(1-p)}. \]

(c) From part (a), we know that \( Y|\{N = n\} \sim \text{Exp}(n\lambda) \), so

\[ \text{Var}[Y|N = n] = \frac{1}{(n\lambda)^2}. \]

Therefore,

\[ \text{Var}(Y|N) = \frac{1}{N^2\lambda^2}. \]

By the law of conditional variance,

\[ \text{Var}(Y) = E[\text{Var}(Y|N)] + \text{Var}(E[Y|N]). \]

Therefore

\[ \text{Var}(Y) = E \left[ \frac{1}{N^2\lambda^2} \right] + \text{Var} \left( \frac{1}{N\lambda} \right) \]
or

\[
\text{Var}(Y) = \frac{1}{\lambda^2} \left( E \left[ \frac{1}{N^2} \right] + \text{Var} \left( \frac{1}{N} \right) \right) = \frac{1}{\lambda^2} \left( E \left[ \frac{1}{N^2} \right] + E \left[ \frac{1}{N^2} \right] - E \left[ \frac{1}{N} \right]^2 \right) = \frac{1}{\lambda^2} \left( 2E \left[ \frac{1}{N^2} \right] - \frac{p^2(\log p)^2}{(1-p)^2} \right) \]

(1)

(2)

(3)

If we define

\[
S(a) = \sum_{n=1}^{\infty} \frac{a^n}{n^2}
\]

then

\[
E \left[ \frac{1}{N^2} \right] = \sum_{n=1}^{\infty} \frac{1}{n^2} (1-p)^{n-1} p = \frac{p}{1-p} \sum_{n=1}^{\infty} \frac{1}{n^2} (1-p)^n = \frac{p}{1-p} S(1-p)
\]

So,

\[
\text{Var}(Y) = \frac{1}{\lambda^2} \left( \frac{2p}{1-p} S(1-p) - \frac{p^2(\log p)^2}{(1-p)^2} \right).
\]

3. **Linear Prediction (20 points).** Let \( \mathbf{X} = [X_2, X_2, X_3]^\top \) be a random vector, with mean zero and covariance matrix given by

\[
\Sigma_{\mathbf{X}} = \begin{bmatrix}
1 & \beta & \beta^2 \\
\beta & 1 & \beta \\
\beta^2 & \beta & 1
\end{bmatrix}
\]

where \(|\beta| < 1\).

(a) (6 points) Find the minimum mean-square error (MMSE) linear estimate of \( X_3 \) given \( X_1 \), and determine the corresponding mean-square error (MSE).

(b) (6 points) Find the minimum mean-square error (MMSE) linear estimate of \( X_3 \) given \( X_2 \), and determine the corresponding mean-square error (MSE).
(c) (6 points) Find the minimum mean-square error (MMSE) linear estimate of $X_3$ given $X_1$ and $X_2$, and determine the corresponding mean-square error (MSE).

(d) (2 points) What do your results tell you about the relative usefulness of $X_1$ and $X_2$ in linear prediction of $X_3$?

**Solution:**

(a) The linear MMSE estimate of $X$ given $Y$ is

$$
\hat{X} = \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}(Y - E[Y]) + E[X].
$$

So, the LMMSE of $X_3$ given $X_1$ is

$$
\hat{X}_3 = \beta^2 X_1.
$$

The MSE of the linear MMSE estimate of $X$ given $Y$ is

$$
\text{Var}(X) - \left(\frac{\text{Cov}(X, Y))^2}{\text{Var}(Y)}\right).
$$

So, the MSE in this case is

$$
1 - \beta^4.
$$

(b) The LMMSE estimate of $X_3$ given $X_2$ is

$$
\hat{X}_3 = \beta X_2.
$$

The MSE is

$$
1 - \beta^2.
$$

(c) The LMMSE estimate of $X$ given $Y = [Y_1, \ldots, Y_n]$ is

$$
\hat{X} = \mathbf{h}^\top (Y - E[Y]) + E[X]
$$

where

$$
\mathbf{h}^\top = \mathbf{\Sigma}_Y^{-1}\mathbf{\Sigma}_{YX}.
$$

Here $Y = [X_1, X_2]^\top$ and $X = X_3$, so

$$
\mathbf{\Sigma}_Y = \begin{bmatrix} 1 & \beta \\ \beta & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{\Sigma}_{YX_3} = \begin{bmatrix} \beta^2 \\ \beta \end{bmatrix}.
$$
Since
\[
\Sigma^{-1}_Y = \frac{1}{1-\beta^2} \begin{bmatrix} 1 & -\beta \\ -\beta & 1 \end{bmatrix}
\]
we see that
\[
h^T = \Sigma^{-1}_Y \Sigma Y_{X_3} = \begin{bmatrix} 0 \\ \beta \end{bmatrix}.
\]
So
\[
\hat{X}_3 = [0, \beta] \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \beta X_2,
\]
which is the same as the LMMSE of \( X_3 \) given \( X_2 \).

The MSE of the LMMSE of \( X \) given \( Y \) is, in general,
\[
\text{Var}(X) - \Sigma^T_{YX} \Sigma^{-1}_Y \Sigma_{XX}.
\]
Here, it is simply the same as the MSE is part (b), namely
\[
1 - \beta^2.
\]

(d) The MSE for the LMMSE estimate based upon \( X_1 \) is larger than that for the LMMSE estimate based upon \( X_2 \). When the estimate is based upon both \( X_1 \) and \( X_2 \), the results show that \( X_1 \) can be ignored.