1. Geometric MMSE

Let $N$ be a geometric random variable with parameter $1 - p$, and $(X_i)_{i \in \mathbb{N}}$ be i.i.d. exponential random variables with parameter $\lambda$. Let $T = X_1 + \cdots + X_N$. Compute the LLSE and MMSE of $N$ given $T$.

**Solution:**

First, we calculate $P(N = n \mid T = t)$, for $t > 0$ and $n \in \mathbb{Z}_+$.

\[
P(N = n \mid T = t) = \frac{P(N = n)f_{T \mid N}(t \mid n)}{\sum_{k=1}^{\infty} P(N = k)f_{T \mid N}(t \mid k)} = \frac{(1 - p)p^{n-1}\lambda^t(n-1)e^{-\lambda t}/(n-1)!}{\sum_{k=1}^{\infty}(1 - p)p^{k-1}\lambda^k(n-1)e^{-\lambda t}/(k-1)!} = \frac{\lambda(\lambda pt)^{n-1}/(n-1)!}{\lambda \sum_{k=1}^{\infty}(\lambda pt)^{k-1}/(k-1)!} = \frac{(\lambda pt)^{n-1}}{e^{\lambda pt}(n-1)!}, \quad n \in \mathbb{Z}_+.
\]

Next, we calculate $E[N \mid T = t]$.

\[
E[N \mid T = t] = \sum_{n=1}^{\infty} n \frac{(\lambda pt)^{n-1}}{e^{\lambda pt}(n-1)!} = \sum_{n=1}^{\infty} \frac{(\lambda pt)^{n-1}}{e^{\lambda pt}(n-1)!} + \sum_{n=1}^{\infty} (n-1) \frac{(\lambda pt)^{n-1}}{e^{\lambda pt}(n-1)!} = 1 + \frac{\lambda pt}{e^{\lambda pt}} \sum_{n=2}^{\infty} \frac{(\lambda pt)^{n-2}}{(n-2)!} = 1 + \frac{\lambda pt}{e^{\lambda pt}} - 1 = 1 + \lambda pt.
\]

Hence, the MMSE is $E[N \mid T] = 1 + \lambda pt$. The MMSE is linear, so it is also the LLSE.

In terms of a Poisson process, $T$ represents the first arrival of a marked Poisson process with rate $\lambda$, where arrivals are marked independently with probability $1 - p$. The marked Poisson process has rate $\lambda(1 - p)$. The unmarked points form a Poisson process of rate $\lambda p$. In time $T$, the expected number of unmarked points is $\lambda pT$, so the conditional expectation of the number of points at time $T$, $N$, is $1 + \lambda pT$. 


2. Property of MMSE

Let \(X,Y_1,\ldots,Y_n\) be square integrable random variables. Argue that

\[
E[(X - E[X | Y_1,\ldots,Y_n])^2] \leq E\left[\left(X - \sum_{i=1}^n E[X | Y_i]\right)^2\right].
\]

**Solution:**

Notice that \(\sum_{i=1}^n E[X | Y_i]\) is a function of \(Y_1,\ldots,Y_n\). The argument is true, since \(E[X | Y_1,\ldots,Y_n]\) is the MMSE estimate of \(X\) among all functions of \(Y_1,\ldots,Y_n\).

\[
E\left[\left(X - \sum_{i=1}^n E[X | Y_i]\right)^2\right]
= E\left[\left(X - E[X | Y_1,\ldots,Y_n] + E[X | Y_1,\ldots,Y_n] - \sum_{i=1}^n E[X | Y_i]\right)^2\right]
= E[(X - E[X | Y_1,\ldots,Y_n])^2] + E\left[\left(E[X | Y_1,\ldots,Y_n] - \sum_{i=1}^n E[X | Y_i]\right)^2\right]
\geq E[(X - E[X | Y_1,\ldots,Y_n])^2],
\]

where in the second equality we expanded the square, and used the orthogonality property.

3. Gaussian Random Vector MMSE

Let

\[
\begin{bmatrix} X \\ Y \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\right)
\]

be a Gaussian random vector.

Let

\[
W = \begin{cases} 
1, & \text{if } Y > 0 \\
0, & \text{if } Y = 0 \\
-1, & \text{if } Y < 0 
\end{cases}
\]

be the sign of \(Y\). Find \(E[WX | Y]\).

**Solution:**

First of all notice that since \(W\) is a function of \(Y\) we have that

\[
E[WX | Y] = W E[X | Y].
\]

Now since \(X,Y\) are jointly Gaussian we have that

\[
E[X | Y] = L[X | Y] = 1 + \frac{Y}{2}.
\]

All in all

\[
E[WX | Y] = \begin{cases} 
1 + \frac{Y}{2}, & \text{if } Y > 0 \\
0, & \text{if } Y = 0 \\
-1 - \frac{Y}{2}, & \text{if } Y < 0.
\end{cases}
\]
4. Jointly Gaussian MMSE and Correlation Coefficients

(a) Provide justification for each of the following steps (1 - 5) to prove that the LLSE is equal to the MMSE estimator for jointly Gaussian random variables $X$ and $Y$.

Let $g(X) = L[Y|X]$.

\[
\mathbb{E}[(Y - g(X))X] = 0 \quad (1)
\]
\[
\Rightarrow \text{cov}(Y - g(X), X) = 0 \quad (2)
\]
\[
\Rightarrow Y - g(X) \text{ is independent of } X \quad (3)
\]
\[
\Rightarrow \mathbb{E}[(Y - g(X))f(X)] = 0 \quad \forall f(.) \quad (4)
\]
\[
\Rightarrow g(X) = \mathbb{E}[Y|X] \quad (5)
\]

(b) Let $X, Y, Z$ be jointly Gaussian random variables such that $X$ is conditionally independent of $Z$ given $Y$. Given the correlation coefficients of $(X,Y)$ and $(Y,Z)$ are $\rho_1$ and $\rho_2$, find the correlation coefficient of $X$ and $Z$ in terms of $\rho_1$ and $\rho_2$. For simplicity you may assume they are zero-mean, but the same answer holds even if they are not zero-mean.

\[
\text{cov}(X, Z) = \mathbb{E}[XZ] = \mathbb{E}\left[\mathbb{E}[XZ|Y]\right] = \mathbb{E}\left[\mathbb{E}[X|Y]\mathbb{E}[Z|Y]\right] = \mathbb{E}[L[X|Y]L[Z|Y]]
\]
\[
= \mathbb{E}\left[\frac{\text{cov}(X, Y)}{\sigma_Y^2}Y, \frac{\text{cov}(Y, Z)}{\sigma_Y^2}Y\right] = \frac{\text{cov}(X, Y)}{\sigma_Y^2} \frac{\text{cov}(Y, Z)}{\sigma_Y^2} \mathbb{E}[Y^2] = \frac{\text{cov}(X, Y)\text{cov}(Y, Z)}{\sigma_Y^2}
\]
\[
\rho = \frac{\text{cov}(X, Z)}{\sigma_X \sigma_Z} = \frac{\text{cov}(X, Y)\text{cov}(Y, Z)}{\sigma_X \sigma_Y \sigma_Y \sigma_Z} = \rho_1 \rho_2
\]

(c) Let $X_0, X_1, ..., X_n$ be a jointly gaussian sequence that forms a Markov Chain such that the correlation coefficient of $X_{i-1}$ and $X_i$ is $\rho_i$. Find the correlation coefficient of $X_0$ and $X_n$ in terms of $\rho_1, \rho_2, ..., \rho_n$ (Hint: use induction and part 2).

Solution:

(a) (1) - This is the orthogonality property for the LLSE (the error $Y - g(X)$ must be orthogonal to all linear functions of $X$.)

(2) - $Y - g(X)$ is zero-mean (LLSE is unbiased). This and (1) together make the covariance 0.

(3) - Uncorrelated jointly Gaussian random variables are independent.

(4) - If $Y - g(X)$ is independent of $X$, it is independent of any function of $X$ and since $Y - g(X)$ is 0 mean, it is orthogonal to any function of $X$.

(5) - This is the defining orthogonality property of the MMSE - the error is orthogonal to all functions of $X$.

(b)

\[
\text{cov}(X, Z) = \mathbb{E}[XZ] = \mathbb{E}[\mathbb{E}[XZ|Y]] = \mathbb{E}[\mathbb{E}[X|Y]\mathbb{E}[Z|Y]] = \mathbb{E}[L[X|Y]L[Z|Y]]
\]
\[
= \mathbb{E}\left[\frac{\text{cov}(X, Y)}{\sigma_Y^2}Y, \frac{\text{cov}(Y, Z)}{\sigma_Y^2}Y\right] = \frac{\text{cov}(X, Y)}{\sigma_Y^2} \frac{\text{cov}(Y, Z)}{\sigma_Y^2} \mathbb{E}[Y^2] = \frac{\text{cov}(X, Y)\text{cov}(Y, Z)}{\sigma_Y^2}
\]
\[
\rho = \frac{\text{cov}(X, Z)}{\sigma_X \sigma_Z} = \frac{\text{cov}(X, Y)\text{cov}(Y, Z)}{\sigma_X \sigma_Y \sigma_Y \sigma_Z} = \rho_1 \rho_2
\]

(c) Base Case: Part 2
Assume this is true for $X_0 \ldots X_k$, $2 \leq k \leq n$. Note that since this is a Markov Chain, $X_0, X_{k+1}$ are conditionally independent given $X_k$. So from a) we have $\rho(X_0, X_{k+1}) = \rho(X_0, X_k)\rho_{k+1} = \rho_1 \rho_2 \ldots \rho_{k+1}$
5. Stochastic Linear System MMSE

Let \((V_n, n \in \mathbb{N})\) be i.i.d. \(\mathcal{N}(0, \sigma^2)\) and independent of \(X_0 = \mathcal{N}(0, u^2)\). Let \(|a| < 1\). Define 

\[ X_{n+1} = aX_n + V_n, \quad n \in \mathbb{N}. \]

(a) What is the distribution of \(X_n\), where \(n\) is a positive integer?

(b) Find \(E[X_{n+m} \mid X_n]\) for \(m, n \in \mathbb{N}, m \geq 1\).

(c) Find \(u\) so that the distribution of \(X_n\) is the same for all \(n \in \mathbb{N}\).

Solution:

(a) First, we find \(X_n\) as a function of \(X_0\) and \((V_n)_{n \in \mathbb{N}}\).

\[
\begin{align*}
X_1 &= aX_0 + V_0 \\
X_2 &= aX_1 + V_1 = a^2X_0 + aV_0 + V_1 \\
X_3 &= aX_2 + V_2 = a^3X_0 + a^2V_0 + aV_1 + V_2.
\end{align*}
\]

Thus, if we proceed doing this recursively, we find that

\[ X_n = a^nX_0 + \sum_{i=0}^{n-1} a^iV_{n-1-i}. \]

Since \(X_0\) and \((V_n)_{n \in \mathbb{N}}\) are independent Gaussian random variables, \(X_n\) is also Gaussian, so we need to find the mean and variance. \(X_0\) and \((V_n)_{n \in \mathbb{N}}\) are zero-mean so

\[
E(X_n) = 0.
\]

We know that

\[
\sum_{i=0}^{n-1} a^i = \frac{1 - a^n}{1 - a},
\]

Thus,

\[
\text{var} X_n = a^{2n} \text{var} X_0 + \sum_{i=0}^{n-1} a^{2i} \text{var} V_{n-1-i} = a^{2n}u^2 + \frac{1 - a^{2n}}{1 - a^2}\sigma^2.
\]

Hence,

\[ X_n \sim \mathcal{N}\left(0, a^{2n}u^2 + \frac{1 - a^{2n}}{1 - a^2}\sigma^2\right). \]

(b) Similarly, by a shift of index

\[ X_{n+m} = a^mX_n + \sum_{i=0}^{m-1} a^iV_{n+m-1-i}. \]

Now suppose that we have zero-mean random variables \(X, Y,\) and \(Z\) where \(X = aY + Z\) and \(Y\) and \(Z\) are independent, then

\[ \text{LLSE}[X \mid Y] = aY. \]

(Why?) Now since the random variables are jointly Gaussian, the MMSE is actually linear. Furthermore, \(X_n\) is independent of \(\sum_{i=0}^{m-1} a^iV_{n+m-1-i}\). Thus,

\[ E(X_{n+m} \mid X_n) = a^mX_n. \]
(c) This is equivalent to $X_1$ having the same variance as $X_0$. Thus,
\[ a^2 u^2 + \sigma^2 = u^2. \]
Thus,
\[ u^2 = \frac{\sigma^2}{1 - a^2}. \]

6. Random Walk with Unknown Drift

Consider a random walk with unknown drift. The dynamics are given, for $n \in \mathbb{N}$, as
\[
X_1(n + 1) = X_1(n) + X_2(n) + V(n),
\]
\[
X_2(n + 1) = X_2(n),
\]
\[
Y(n) = X_1(n) + W(n).
\]
Here, $X_1$ represents the position of the particle and $X_2$ represents the velocity of the particle (which is unknown but constant throughout time). $Y$ is the observation. $V$ and $W$ are independent Gaussian noise variables with mean zero and variance $\sigma^2_V$ and $\sigma^2_W$, respectively.

(a) Write down the dynamics of the system in matrix-vector form and write down the Kalman filter recursive equations for this system.

(b) Let $k$ be a positive integer. Compute the prediction $\mathbb{E}(X(n + k) \mid Y^{(n)})$, where $Y^{(n)}$ is the history of the observations $Y_0, \ldots, Y_n$, in terms of the estimate $\hat{X}(n) := \mathbb{E}(X(n) \mid Y^{(n)})$.

(c) Now let $k = 1$ and compute the smoothing estimate $\mathbb{E}(X(n) \mid Y^{(n+1)})$ in terms of the quantities that appear in the Kalman filter equation.

**Hint:** Use the law of total expectation
\[
\mathbb{E}(X(n) \mid Y^{(n+1)}) = \mathbb{E}[\mathbb{E}(X(n) \mid X(n+1), Y^{(n+1)}) \mid Y^{(n+1)}],
\]
as well as the innovation
\[
\tilde{X}(n + 1) := X(n + 1) - L[X(n + 1) \mid Y^{(n)}].
\]

**Solution:**

(a) In matrix form, the dynamics are
\[
\begin{bmatrix}
X_1(n + 1) \\
X_2(n + 1)
\end{bmatrix} = \begin{bmatrix}
1 & 1 \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
X_1(n) \\
X_2(n)
\end{bmatrix} + \begin{bmatrix}
V(n) \\
0
\end{bmatrix},
\]
\[
Y(n) = \begin{bmatrix}
1 & 0
\end{bmatrix}
\begin{bmatrix}
X_1(n) \\
X_2(n)
\end{bmatrix} + W(n).
\]

The Kalman filter equations are
\[
\hat{X}(n) = \begin{bmatrix}
1 & 1 \\
0 & 1
\end{bmatrix} \hat{X}(n - 1) + K_n(Y(n) - \begin{bmatrix}
1 & 1
\end{bmatrix} \hat{X}(n - 1)),
\]
\[ K_n = S_n \begin{bmatrix} 1 \\ 0 \end{bmatrix} \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} S_n \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sigma_W^2 \right)^{-1}, \]

\[ S_n = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \Sigma_{n-1} \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + \begin{bmatrix} \sigma_V^2 & 0 \\ 0 & 0 \end{bmatrix}, \]

\[ \Sigma_n = \left( \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - K_n \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) S_n. \]

(b) First suppose that \( k = 1 \) and note that

\[ \mathbb{E}(X(n + 1) \mid Y^{(n)}) = \mathbb{E}(AX(n) + \tilde{V}(n) \mid Y^{(n)}) \]

and by independence of the noise and linearity of expectation,

\[ \mathbb{E}(X(n + 1) \mid Y^{(n)}) = A \mathbb{E}(X(n) \mid Y^{(n)}) = A\hat{X}(n). \]

The interpretation is quite simple: we take our estimate at time \( n \), \( \hat{X}(n) \), and then move it forwards one time step via the transition dynamics \( A \).

It is then easy to see that

\[ \mathbb{E}(X(n + k) \mid Y^{(n)}) = A^k \hat{X}(n). \]

By computing

\[ A^k = \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \]

then one has

\[ \mathbb{E}(X(n + k) \mid Y^{(n)}) = \begin{bmatrix} \hat{X}_1(n) + k\hat{X}_2(n) \\ \hat{X}_2(n) \end{bmatrix}, \]

that is, your predicted velocity at time \( n + k \) is still \( \hat{X}_2(n) \) (makes sense; the velocity is not changing with time) and your predicted position at time \( n + k \) is \( \hat{X}_1(n) \), plus the velocity \( \hat{X}_2(n) \) added \( k \) times.

(c) The first step is to recognize that

\[ \mathbb{E}(X(n) \mid X(n + 1), Y^{(n+1)}) = \mathbb{E}(X(n) \mid X(n + 1), Y^{(n)}, Y(n + 1)) = \mathbb{E}(X(n) \mid X(n + 1), Y^{(n)}) \]

since \( Y(n + 1) = CX(n + 1) + W(n + 1) \) and \( W(n + 1) \) is independent of everything else, so conditioned on \( X(n + 1), Y(n + 1) \) does not tell you anything new about \( X(n) \). Now, observe that

\[ \mathbb{E}(X(n) \mid X(n + 1), Y^{(n)}) = L[X(n) \mid X(n + 1), Y^{(n)}] = L[X(n) \mid Y^{(n)}] + L[X(n) \mid \tilde{X}(n + 1)] \]

where \( \tilde{X}(n + 1) := X(n + 1) - L[X(n + 1) \mid Y^{(n)}] \) is the innovation. By the previous part, \( L[X(n + 1) \mid Y^{(n)}] = A\hat{X}(n) \). So,

\[ \tilde{X}(n + 1) = X(n + 1) - A\hat{X}(n). \]
Also,
\[
\text{cov}(X(n), \hat{X}(n + 1)) = \text{cov}(X(n), A[X(n) - \hat{X}(n)] + \tilde{V}(n)) \\
= \text{cov}(X(n), X(n) - \hat{X}(n)) A^T \\
= \text{cov}(X(n) - \hat{X}(n)) A^T
\]

since the error \(X(n) - \hat{X}(n)\) is uncorrelated with the estimate \(\hat{X}(n)\). We are in good shape since \(\text{cov}(X(n) - \hat{X}(n)) = \Sigma_n\) by definition. Also, \(\text{cov} \tilde{X}(n + 1) = S_{n+1}\) by definition. Thus,
\[
L[X(n) | \hat{X}(n + 1)] = \Sigma_n A^T S_{n+1}^{-1}(X(n + 1) - A \hat{X}(n))
\]

and
\[
\mathbb{E}(X(n) | Y^{(n+1)}) = \mathbb{E}(\mathbb{E}(X(n) | X(n + 1), Y^{(n+1)}) | Y^{(n+1)}) \\
= \mathbb{E}(\hat{X}(n) + \Sigma_n A^T S_{n+1}^{-1} \hat{X}(n + 1) | Y^{(n+1)}) \\
= \hat{X}(n) + \Sigma_n A^T S_{n+1}^{-1}(\hat{X}(n + 1) - A \hat{X}(n)).
\]

7. [Bonus] Rotationally Invariant Random Variables

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.

You have two independent and identically distributed continuous random variables, with zero mean, such that the joint density is rotation invariant. Show that the random variables have the normal distribution.

**Solution:**

Let \(X\) and \(Y\) be such random variables. Since they are rotation invariant, they should be symmetric, and hence without loss of generality, we can assume that the density looks like:

\[
f_X(x) = a \exp(-h(x^2)), \quad f_Y(y) = a \exp(-h(y^2)).
\]

This is because \(f_X(x)\) must be a function of \(x^2\) only, so if \(f_X(x) = k(x^2)\), then we may write \(f_X(x) = \exp(-\ln(k(x^2)^{-1}))\) and take \(h = \ln \circ k^{-1}\). This requires the density \(f_X\) to be strictly positive almost everywhere; this follows from rotational invariance and independence. The joint density is also isotropic, so we have:

\[
f_{X,Y}(x, y) = A \exp(-g(r^2))
\]

where, \(r^2 = x^2 + y^2\). Now, we know that the random variables are independent, and so \(f_{X,Y}(x, y) = f_X(x)f_Y(y)\). We get \(h(x^2) + h(y^2) = g(x^2 + y^2)\). Plugging \(y = 0\), and assuming \(h(0) = 0\), \(h(x^2) = g(x^2)\) which implies \(h(x) = g(x)\) for all \(x > 0\). Also,

\[
h(x^2) + h(y^2) = g(x^2 + y^2) = h(x^2 + y^2),
\]

\[
h(x) + h(y) = h(x + y)
\]

for all \(x > 0, y > 0\).
\[ h(x) = bx, \quad \text{where } b \text{ is constant.} \]

Now, plugging in, \( f_X(x) = a \exp(-bx^2) \) which is Gaussian.

As an alternative proof, for \( t \in \mathbb{R} \), \( tX + tY \) has the same distribution as \( \sqrt{2}tX \), so \( \varphi(t)^2 = \varphi(\sqrt{2}t) \), where \( \varphi \) is the characteristic function of \( X \). By iterating, \( \varphi(t)^n = \varphi(\sqrt{n}t) \) for all positive integers \( n \). So, letting \( \exp c := \varphi(1) \), for \( t^2 = a/b \), where \( a, b \) are positive integers, then \( \varphi(t) = \varphi(\sqrt{a/b}) = \varphi(1/\sqrt{b})^a \) and \( \varphi(1) = \exp c = \varphi(1/\sqrt{b})^b \), so we have \( \varphi(t) = \exp(ct/b) = \exp(ct^2) \). We have found that the equation \( \varphi(t) = \exp(ct^2) \) holds for all \( t \) such that \( t^2 \in \mathbb{Q} \), and then by continuity of \( \varphi \) the equation must be true everywhere, and we have the characteristic function of a Gaussian.