Matlab Solutions 3

**Starter.** We start by writing
\[
\mathbb{E}((X - \mathbb{E}(X|Y))^2) = \mathbb{E}(X^2) + \mathbb{E}(\mathbb{E}(X|Y)^2) - 2\mathbb{E}(X\mathbb{E}(X|Y))
\]
\[
= \mathbb{E}(X^2) + \mathbb{E}(\mathbb{E}(X|Y)^2) - 2\mathbb{E}(X\mathbb{E}(X|Y)|Y)
\]
\[
= \mathbb{E}(X^2) + \mathbb{E}(\mathbb{E}(X|Y)^2) - 2\mathbb{E}(\mathbb{E}(X|Y)^2)
\]
\[
= \mathbb{E}(X^2) - \mathbb{E}(\mathbb{E}(X|Y)^2).
\]
Additionally, for \(\mathbb{E}(X|Y)\) to be well-defined, it is not necessary that \(Y\) is integrable by definition.

**Exercise 1.**

a) We claim that \(\hat{X} = \mathbb{E}(X|Y) = aY/(1 + a^2)\). To see this, we notice first that \(Y\) is Gaussian with mean 0 and variance \(1 + a^2\) and \(\text{Cov}(X,Y) = a\). Therefore, \((X,Y)\) is bivariate Gaussian and has joint density
\[
p_{X,Y}(x,y) = \frac{1}{2\pi} \exp\left\{ -\frac{1 + a^2}{2} \left( x^2 - \frac{2a}{1 + a^2} xy + \frac{1}{1 + a^2} y^2 \right) \right\}.
\]
[Alternatively, one may compute \(p_{X,Y}(x,y) = p_{Y|X}(y|x) p_X(x) = \ldots\).] A little algebra with the (integration by parts) equality given in the hint then shows that
\[
\psi(y) = \int_{\mathbb{R}} x \frac{p_{X,Y}(x,y)}{p_Y(y)} \, dy = \frac{ay}{1 + a^2},
\]
i.e., \(\hat{X} = \psi(Y) = aY/(1 + a^2)\).

b) Since \(\hat{X} - X = aY/(1 + a^2) - X = a(aX + Z)/(1 + a^2) - X = aZ/(1 + a^2) - X/(1 + a^2)\), we get
\[
\mathbb{E}((\hat{X} - X)^2) = \frac{a^2\mathbb{E}(Z^2)}{(1 + a^2)^2} + \frac{\mathbb{E}(X^2)}{(1 + a^2)^2} = \frac{1}{1 + a^2},
\]
where we have used \(\mathbb{E}(X^2) = \mathbb{E}(Z^2) = 1\).

c) As \(\hat{X}_{\text{ML}} = Y/a - X = Z/a\), we get
\[
\mathbb{E}((\hat{X}_{\text{ML}} - X)^2) = \frac{\mathbb{E}(Z^2)}{a^2} = \frac{1}{a^2} > \frac{1}{1 + a^2} = \mathbb{E}((\hat{X} - X)^2).
\]

d) The random vector \(\mathbf{Y} = (Y_1, \ldots, Y_n)\) is multivariate Gaussian with mean 0 and covariance matrix \(\Sigma\), where \(\Sigma\) has 1 + \(a^2\) as its diagonal elements and has \(a^2\) as its off-diagonal elements. Moreover, \((X,\mathbf{Y})\) is multivariate Gaussian with mean \((0,\mathbf{0})\) and covariance matrix \(\Sigma'\), where \(\Sigma'\) can be obtained similarly by noticing that \(\text{Cov}(X,Y_i) = a\) for every \(i \leq n\). [Again, one may compute alternatively \(p_{X,Y}(x,y) = p_{Y|X}(y|x) p_X(x) = \prod_{i=1}^n p_{Y_i|X}(y_i|x) p_X(x) = \ldots\).] Some algebra then gives that
\[
\psi(y) = \int_{\mathbb{R}^n} x \frac{p_{X,Y}(x,y)}{p_Y(y)} \, dy = \frac{a \sum_{i=1}^n y_i}{1 + na^2}.
\]
Therefore, \( \hat{X}_n = \psi(Y) = a \sum_{i=1}^{n} Y_i/(1 + na^2) \).

e) We have \( X_n - X = a \sum_{i=1}^{n} Y_i/(1 + na^2) - X = a \sum_{i=1}^{n} Z_i/(1 + na^2) + X/(1 + na^2) \). Thus,

\[
\mathbb{E}(X_n - X)^2 = \frac{a^2}{(1 + na^2)^2} \sum_{i=1}^{n} \mathbb{E}(Z_i^2) + \frac{1}{(1 + na^2)^2} \mathbb{E}(X^2) = \frac{1}{1 + na^2}.
\]

Similarly, we have \( \hat{X}_{ML,n} - X = \sum_{i=1}^{n} Z_i/(na) \), so that

\[
\mathbb{E}(X_{ML,n} - X)^2 = \frac{1}{n^2 a^2} \sum_{i=1}^{n} \mathbb{E}(Z_i^2) = \frac{1}{na^2},
\]

which is larger than \( \mathbb{E}(X_n - X)^2 \). Moreover, the ratio \( \mathbb{E}(X_{ML,n} - X)^2/\mathbb{E}(X_n - X)^2 \) is equal to \( (1 + na^2)/(na^2) \) which tends to 1 as \( n \) tends to \( \infty \).

f) We claim that the sequence \( (U_n, n \geq 1) \) converges almost surely to 0. Indeed, we have computed \( \text{Var}(U_n) = 1/(1 + na^2) \) and \( \mathbb{E}(U_n) = 0 \). Chebyshev’s inequality then gives

\[
\mathbb{P}(\{|U_n - 0| > \varepsilon\}) \leq \frac{1}{\varepsilon^2(1 + na^2)}.
\]

Standard techniques seen in class imply that \( (U_n, n \geq 1) \) converges almost surely to 0. See also the code for a numerical confirmation of this fact. Next, we claim that the sequence \( (V_n, n \geq 1) \) converges in distribution to \( \mathcal{N}(0, 1/a^2) \). Indeed, we have

\[
\varphi_{V_n}(t) = \mathbb{E}(e^{itV_n}) = \mathbb{E}\left(e^{it\sqrt{n}(X+a\sum_{i=1}^{n} Z_i)/(1+na^2)}\right)
\]

\[
= \mathbb{E}\left(e^{it\sqrt{n}X/(1+na^2)}\right) \prod_{i=1}^{n} \mathbb{E}\left(e^{it\sqrt{na}Z_i/(1+na^2)}\right)
\]

\[
= e^{-\frac{1}{2} \frac{n-ax}{1+na^2} t^2} \xrightarrow{n \to \infty} e^{-\frac{1}{2} \frac{1}{a^2} t^2},
\]

which is the characteristic function of \( \mathcal{N}(0, 1/a^2) \). By Lévy’s continuity theorem, we get the claim. See also the code for a confirmation of this fact.

**Exercise 2.**

a) First we compute

\[
\mathbb{P}(\{Y \leq y\}) = \mathbb{P}(\{aX + Z \leq y\}) = \sum_{x \in \mathcal{C}} p_x \mathbb{P}(\{Z \leq y - ax\}) = \sum_{x \in \mathcal{C}} p_x \int_{-\infty}^{y-ax} p_Z(z) dz.
\]

Since \( \mathcal{C} \) is countable, we conclude that \( Y \) is continuous with density \( p_Y(y) = \sum_{x \in \mathcal{C}} p_x p_Z(y - ax) \).

b) For any Borel-measurable and bounded function \( g : \mathbb{R} \to \mathbb{R} \), we have

\[
\mathbb{E}(\psi(Y)g(Y)) = \int_{-\infty}^{\infty} \psi(y)g(y)p_Y(y) dy
\]

\[
= \int_{-\infty}^{\infty} \sum_{x \in \mathcal{C}} p_x p_Z(y - ax) g(y) p_Y(y) dy
\]

\[
= \int_{-\infty}^{\infty} \sum_{x \in \mathcal{C}} x g(y) p_x p_Z(y - ax) dy = \mathbb{E}(Xg(Y)).
\]

which confirms the fact that \( \psi(Y) = \mathbb{E}(X|Y) \).
c) In this case, we have \( p_Y(y) = p_Z(y - a)/2 + p_Z(y + a)/2 \). Here, \( p_Z \) is the standard Gaussian density. Therefore, we have
\[
\psi(y) = \frac{p_Z(y - a) - p_Z(y + a)}{p_Z(y - a) + p_Z(y + a)} = \frac{e^{ay} - e^{-ay}}{e^{ay} + e^{-ay}} = \tanh(ay).
\]
So \( \hat{X} = \psi(Y) = \tanh(aY) \).

d) Since \( \mathbb{E}(X^2) = 1 \), the variance
\[
\mathbb{E}((\hat{X} - X)^2) = \mathbb{E}(X^2) - \mathbb{E}(\mathbb{E}(X|Y)^2) = 1 - \mathbb{E}(\tanh(aY)^2)
\]
\[
= 1 - \mathbb{E}(\tanh(a^2X + aZ)^2) = 1 - \mathbb{E}(\tanh(a^2 + aZ)^2) := g(a),
\]
where the last step follows from \( p_Z(z) = p_Z(-z), \forall z \in \mathbb{R} \), and \( \tanh(-x) = -\tanh(x), \forall x \in \mathbb{R} \). See the code for a numerical representation of \( g(a) \). Next, a straightforward computation reveals that the variance
\[
\mathbb{E}((\hat{X}_{\text{ML}} - X)^2) = \mathbb{E}((\text{sign}(a^2X + aZ) - X)^2) = 4\Phi(-|a|) := f(a),
\]
where \( \Phi(x) = \int_{-\infty}^{x} p_Z(z)dz \). Again, see the code for a numerical representation of \( f(a) \), which is clearly above \( g(a) \) (in particular, \( f(0) = 2 > g(0) = 1 \)).

Note also the following:

- \( \Phi(-|a|) = \mathbb{P}(|Z| > |a|) \leq \inf_{s \geq 0} \mathbb{E}(e^{s(Z - |a|)}) = e^{-a^2/2} \).
- It does not hold that \( \mathbb{E}((\hat{X}_{\text{ML}} - X)^2) = \mathbb{E}(X^2) - \mathbb{E}(\hat{X}_{\text{ML}}^2) \) (which is equal to 0 here), because \( \hat{X}_{\text{ML}} \neq \mathbb{E}(X|Y) \).

e) The random vector \((X, Y)\) has “joint pmf/density” \( p_{X,Y}(x, y) = p_x \prod_{i=1}^{n} p_{Z_i}(y_i - ax) \) [see the first part of this exercise for a more precise formulation]. Therefore,
\[
\psi(y) = \prod_{i=1}^{n} p_{Z_i}(y_i - a)/2 - \prod_{i=1}^{n} p_{Z_i}(y_i + a)/2
\]
\[
\quad \quad \quad \quad = \prod_{i=1}^{n} p_{Z_i}(y_i - a)/2 + \prod_{i=1}^{n} p_{Z_i}(y_i + a)/2
\]
\[
= \tanh\left(a \sum_{i=1}^{n} y_i \right).
\]
So \( \hat{X}_n = \psi(Y) = \tanh(a \sum_{i=1}^{n} Y_i) \).

f) With \( Y' = \sum_{i=1}^{n} Y_i \) and \( Z' = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \sim \mathcal{N}(0, 1) \), we have
\[
\mathbb{E}((\hat{X}_n - X)^2) = 1 - \mathbb{E}(\tanh(aY')^2) = 1 - \mathbb{E}(\tanh(na^2 + \sqrt{n}aZ')^2) = g\left(\sqrt{n}a\right).
\]
On the other hand, it is straightforward to compute the variance
\[
\mathbb{E}((\hat{X}_{\text{ML,n}} - X)^2) = \mathbb{E}((\text{sign}(na^2X + \sqrt{n}aZ') - X)^2) = f\left(\sqrt{n}a\right) \leq 4e^{-na^2/2},
\]
which implies that the variances \( \mathbb{E}((\hat{X}_n - X)^2) \) and \( \mathbb{E}((\hat{X}_{\text{ML,n}} - X)^2) \) decay exponentially in \( n \).