Homework #2 Solutions

Please submit the assignment on Gradescope

1. Correlation coefficient. We will prove that $|\rho_{X,Y}| = 1$ if and only if $(X - E(X))$ is a linear function of $(Y - E(Y))$, where $\rho_{X,Y} = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$, using the Schwarz inequality.

a. Prove the following inequality, which is known as the Schwarz inequality.

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

Hint: Use the fact that $E((X + aY)^2) \geq 0$ for every real number $a$.

b. Prove that equality holds if and only if either $Y = cX$ or $X = cY$ for some constant $c$.

c. Show that $|\rho_{X,Y}| = 1$ if and only if $(X - E(X))$ is a linear function of $(Y - E(Y))$ i.e. $(X - E(X)) = c(Y - E(Y))$ or $(Y - E(Y)) = c(X - E(X))$.

d. Use the Schwarz inequality to show that the correlation coefficient $\rho_{X,Y}$ satisfies $|\rho_{X,Y}| \leq 1$.

Solution

a. Consider the following quadratic equation in the parameter $a$:

$$0 = E((X + aY)^2) = E(X^2) + 2aE(XY) + a^2E(Y^2).$$

Since it is the expected value of a non-negative random variable, $E((X + aY)^2) \geq 0$. If $E((X + aY)^2) > 0$ there are two imaginary solutions, while if $E((X + aY)^2) = 0$ there is one real solution. Thus the discriminant must satisfy

$$4(E(XY))^2 - 4E(X^2)E(Y^2) \leq 0,$$

which we can rewrite as

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

Here is another proof. Let

$$a = \pm \sqrt{\frac{E(X^2)}{E(Y^2)}}.$$

Plugging this value into $E((X + aY)^2) \geq 0$ we obtain

$$-E(XY) \leq \sqrt{E(X^2)E(Y^2)}$$ and $E(XY) \leq \sqrt{E(X^2)E(Y^2)}$

Combining the two inequalities or $E(XY)$ yields

$$|E(XY)| \leq \sqrt{E(X^2)E(Y^2)} \Rightarrow (E(XY))^2 \leq E(X^2)E(Y^2),$$

which is what we set out to prove.

b. If $X = cY$ then

$$(E(cY \cdot Y))^2 = c^2(E(Y^2))^2 = E((cY)^2)E(Y^2) = E(X^2)E(Y^2).$$

and similarly for $Y = cX$. Conversely, if $(E(XY))^2 = E(X^2)E(Y^2)$ then the discriminant of the quadratic equation in part (a) is 0. Therefore $E(X + aY)^2 = 0$, which means that $X + aY = 0$ with probability 1. Hence $X = -aY$ with probability 1. Clearly,

$$a = \pm \sqrt{\frac{E(X^2)}{E(Y^2)}}.$$
A similar quadratic equation can be formed with $E((aX + Y)^2) = 0$ to prove that $Y = -aX$ is another possible solution.

c. The square of correlation coefficient is by definition
\[ \rho^2_{X,Y} = \frac{\text{Cov}(X,Y)}{\text{Var}(X)\text{Var}(Y)}. \]

If we define random variables $U = X - E(X)$ and $V = Y - E(Y)$, then
\[ \rho^2_{x,y} = \frac{E(UV)^2}{E(U^2)E(V^2)} \leq 1, \]
where the inequality follows from the Schwarz inequality. Therefore $|\rho_{X,Y}| \leq 1$.

d. Let $U = X - E(X)$ and $V = Y - E(Y)$. $|\rho_{X,Y}| = 1$ if and only if $E(UV)^2 = E(U^2)E(V^2)$. Using what we proved in b., the equality holds if and only if $U$ is a linear function of $V$. Therefore, $|\rho_{X,Y}| = 1$ if and only if $(X - E(X))$ is a linear function of $(Y - E(Y))$.

2. First available teller. A bank has two tellers. The service times for tellers 1 and 2 are independent exponential random variables $X_1 \sim \text{Exp}(\lambda_1)$ and $X_2 \sim \text{Exp}(\lambda_2)$, respectively. You arrive at the bank and find that both tellers are busy but nobody else is waiting to be served. You are served by the first available teller once he/she is free. What is the probability that you are served by the teller 1?

**Solution**

The tellers’ service times are exponentially distributed, hence memoryless. Thus the service time distribution does not depend on my arrival time. The probability that I will be served by the first teller is
\[
P\{X_1 < X_2\} = \int_0^\infty \int_0^\infty \lambda_1 e^{-\lambda_1 x_1} \lambda_2 e^{-\lambda_2 x_2} dx_2 dx_1 = \int_0^\infty \lambda_1 e^{-(\lambda_1 + \lambda_2) x_1} dx_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2}.
\]

In other words, the probability of being served first by teller $i$ is proportional to the teller’s service rate $\lambda_i$.

3. Radar signal detection. The received signal $S$ for a radar channel is 0 if there is no target and a random variable $X \sim \mathcal{N}(0, P)$ if there is a target. Both possibilities occur with equal probability. Thus
\[
S = \begin{cases} 
0 & \text{with probability } \frac{1}{2} \\
X \sim \mathcal{N}(0, P) & \text{with probability } \frac{1}{2}.
\end{cases}
\]

The radar receiver observes $Y = S + Z$, where the noise $Z \sim \mathcal{N}(0, N)$ is independent of $S$. Find the optimal decoder for deciding whether $S = 0$ or $S = X$ and its probability of error. Give your answer in terms of intervals of $y$ and express the boundary points of the intervals in terms of $P$ and $N$.

Hint: You can cast this detection problem in the form discussed in class by defining the signal $\Theta$ to be 0 if $S = 0$ and 1 if $S = X$. 
Solution

To cast this problem as a standard detection problem, we define a random variable $\Theta$ by

$$
\Theta = \begin{cases} 
0 & \text{if } S = 0 \\
1 & \text{if } S = X 
\end{cases}
$$

Then $p_\Theta(0) = p_\Theta(1) = \frac{1}{2}$. The optimal decoder $\hat{\Theta}(\cdot)$ for $\Theta$ uses the MAP rule: i.e., set $\hat{\Theta}(y) = \theta$ where $\theta$ maximizes the conditional pmf $p_{\Theta|Y}(\theta|y)$. By Bayes rule,

$$
p_{\Theta|Y}(\theta|y) = \frac{f_{Y|\Theta}(y|\theta)p_\Theta(\theta)}{f_Y(y)} = \begin{cases} 
1 & \text{if } \frac{f_{Y|\Theta}(y|1)}{f_{Y|\Theta}(y|0)} > 1 \\
0 & \text{otherwise}
\end{cases}
$$

The likelihood ratio can be written

$$
\frac{f_{Y|\Theta}(y|1)}{f_{Y|\Theta}(y|0)} > 1 \iff \sqrt{\frac{N}{P + N}}e^{\frac{y^2}{2(N + P)}} > 1 \iff y^2 > \frac{(P + N)N}{P} \ln \left(\frac{P + N}{N}\right).
$$

Thus the MAP decision rule becomes

$$
\hat{\Theta}(y) = \begin{cases} 
0 & |y| \leq \sqrt{\frac{(P + N)N}{P} \ln \left(\frac{P + N}{N}\right)} \\
1 & \text{otherwise}
\end{cases}
$$

To find the error probability, define $\tau = \sqrt{\frac{(P + N)N}{P} \ln \left(\frac{P + N}{N}\right)}$. Then

$$
P_e = P\{\hat{\Theta}(Y) \neq \Theta\}
= P\{\hat{\Theta}(Y) = 1, \Theta = 0\} + P\{\hat{\Theta}(Y) = 0, \Theta = 1\}
= P\{|Y| \geq \tau, \Theta = 0\} + P\{|Y| < \tau, \Theta = 1\}
= p_\Theta(0) \left(\int_{-\tau}^{-\infty} f_{Y|\Theta}(y|0) dy + \int_{0}^{\infty} f_{Y|\Theta}(y|0) dy\right) + p_\Theta(1) \int_{-\tau}^{\tau} f_{Y|\Theta}(y|1) dy
= \frac{1}{2} \left(2 \int_{-\tau}^{\tau} f_{Y|\Theta}(y|0) dy + \int_{-\tau}^{\tau} f_{Y|\Theta}(y|1) dy\right)
= \frac{1}{2} \left(2Q\left(\sqrt{\frac{\tau}{P + N}}\right) + (1 - 2Q\left(\frac{\tau}{\sqrt{P + N}}\right))\right)
= Q\left(\sqrt{\frac{(P + N)N}{P}} \ln \left(\frac{P + N}{N}\right)\right) + \frac{1}{2} - Q\left(\sqrt{\frac{N}{P}} \ln \left(\frac{P + N}{N}\right)\right).
$$

In Figure 1 on page 4, the pdfs of noise and signal + noise intersect at $\pm \tau$. The decision region for “no signal” is the interval $[-\tau, +\tau]$. The error probability is the average of the probabilities
of the tail of the noise and of the central region of signal + noise. In the example shown in Figure 1, SNR = 4 and $P_e = 0.1780$.

![Figure 1: PDFs of noise ($N = 1$) and radar signal + noise ($P + N = 4$).](image)

4. Ternary signaling. Let the signal $S$ be a random variable defined as follows:

$$S = \begin{cases} -1 & \text{with probability } \frac{1}{3} \\ 0 & \text{with probability } \frac{1}{3} \\ +1 & \text{with probability } \frac{1}{3} \end{cases}$$

The signal is sent over a channel with additive Laplacian noise $Z$, i.e., $Z$ is a Laplacian random variable with pdf

$$f_Z(z) = \frac{\lambda}{2} e^{-\lambda|z|}, \quad -\infty < z < \infty.$$ 

The signal $S$ and the noise $Z$ are assumed to be independent and the channel output is their sum $Y = S + Z$.

a. Find $f_{Y|S}(y|s)$ for $s = -1, 0, +1$. Sketch the conditional pdfs on the same graph.

b. Find the optimal decoding rule $\hat{\Theta}(Y)$ for deciding whether $S$ is $-1, 0$ or $+1$. Give your answer in terms of ranges of values of $Y$.

c. Find the probability of decoding error for $\hat{\Theta}(y)$ in terms of $\lambda$.

Solution

a. We use a trick here that is used several times in the lecture notes. Since $Y = S + Z$ and $Z$ and $S$ are independent, the conditional pdf is

$$f_{Y|S}(y|s) = f_Z(y - s) = \frac{1}{2} \lambda e^{-\lambda|y-s|}.$$
The plots are shown for \( \lambda = 1 \) in Figure 2 on page 5.

b. The optimal decoding rule is MAP: \( \hat{\Theta}(y) = s \) where \( s \) maximizes

\[
p(s|y) = \frac{f(y|s)p(s)}{f(y)}.
\]

Since \( p_S(s) \) is the same for \( s = -1, 0, +1 \), the MAP rule becomes the maximum-likelihood decoding rule: \( \hat{\Theta}(y) = s \) that maximizes \( f(y|s) \). The conditional pdfs are plotted in Figure 2. By inspection, the ML rule reduces to

\[
\hat{\Theta}(y) = \begin{cases} 
-1 & y < -\frac{1}{2} \\
0 & -\frac{1}{2} < y < +\frac{1}{2} \\
+1 & y > +\frac{1}{2}.
\end{cases}
\]
c. Inspection of Figure 2 shows how to calculate the probability of error.

\[
P_e = \sum_i P\{\text{error} \mid i \text{ sent}\} P\{i \text{ sent}\}
\]

\[
= \frac{1}{3} \sum_i P\{\text{error} \mid i \text{ sent}\}
\]

\[
= \frac{1}{3} (1 - P\{-\frac{1}{2} < S + Z < +\frac{1}{2} \mid S = 0\}) + \frac{1}{3} P\{S + Z > -\frac{1}{2} \mid S = -1\} + \frac{1}{3} P\{S + Z < +\frac{1}{2} \mid S = +1\}
\]

\[
= \frac{1}{3} \left(1 - P\{-\frac{1}{2} < Z < \frac{1}{2}\}\right) + \frac{1}{3} P\{Z < -\frac{1}{2}\} + \frac{1}{3} P\{Z > +\frac{1}{2}\}
\]

\[
= \frac{2}{3} \left(P\{Z < -\frac{1}{2}\} + P\{Z > +\frac{1}{2}\}\right)
\]

\[
= \frac{4}{3} P\{Z > +\frac{1}{2}\} \quad \text{(by symmetry)}
\]

\[
= \frac{4}{3} \int_{\frac{1}{2}}^{\infty} \frac{1}{2} \lambda e^{-\lambda z} \, dz = \frac{2}{3} e^{-\frac{1}{2} \lambda}.
\]

5. Function of uniform random variables. Let \(X\) and \(Y\) be two independent \(U[0, 1]\) random variables. Find the probability density function (pdf) of \(Z = (X + Y) \mod 1\) (i.e., \(Z = X + Y\) if \(X + Y \leq 1\) and \(X + Y - 1\) if \(X + Y > 1\)).

**Solution**

First, we find the cdf of \(Z\). For \(z \in [0, 1]\),

\[
F_Z(z) = P\{Z \leq z\}
\]

\[
= P\{Z \leq z, X + Y \leq 1\} + P\{Z \leq z, X + Y > 1\}
\]

\[
= P\{X + Y \leq z\} + P\{1 < X + Y \leq 1 + z\}.
\]

Here, we have used the law of total probability and divided the event \(\{Z \leq z\}\) into two parts, depending on whether it occurs together with the event \(\{X + Y \leq 1\}\) or together with its complement. The way of dividing the event is chosen along the lines of the definition of \(Z\) as given in the problem statement. The events in the last line are shown in Figure 3. Since \(X\) and \(Y\) are independent and uniform, the probability corresponds directly to the size of the shaded areas. Thus

\[
F_Z(z) = \frac{z^2}{2} + \left(\frac{1}{2} - \frac{(1 - z)^2}{2}\right) = z.
\]

Taking the derivative of the cdf, we obtain the pdf \(f_z(z) = 1\) for \(z \in [0, 1]\). Thus, \(Z \sim U[0, 1]\).

6. Iterated expectation. Let \(\Lambda\) and \(X\) be random variables with

\[
\Lambda \sim f_\Lambda(\lambda) = \begin{cases} 
\frac{5}{3} \lambda^{\frac{1}{2}} & 0 \leq \lambda \leq 1 \\
0 & \text{otherwise}
\end{cases}
\]

and \(X|\{\Lambda = \lambda\} \sim \text{Exp}(\lambda)\). Find \(E(X)\).
Figure 3: The shaded regions correspond to the events in question.

Solution

We are given \( X|\Lambda = \lambda \sim \text{Exp}(\lambda) \). Therefore \( \mathbb{E}(X|\Lambda = \lambda) = \frac{1}{\lambda} \).

\[
\mathbb{E}(X) = \mathbb{E}_\Lambda(\mathbb{E}(X|\Lambda = \lambda)) = \mathbb{E}_\Lambda \left( \frac{1}{\lambda} \right) = \int_{-\infty}^{\infty} \frac{1}{\lambda} f_\lambda(\lambda) d\lambda \\
= \int_0^1 \frac{1}{\lambda} \cdot \frac{5}{3} \lambda^{\frac{2}{3}} d\lambda = \frac{5}{3} \int_0^1 \lambda^{-\frac{1}{3}} d\lambda = \frac{5 \lambda^{\frac{2}{3}}}{3} \bigg|_0^1 \frac{1}{2} = \frac{5}{2}.
\]

7. Mean square error estimation. The number of packets arriving per unit time at a node in a communication network is a Poisson random variable \( X \) with rate \( \Lambda \sim \text{Exp}(a) \). Find the MMSE estimate of the rate \( \Lambda \) given the observation \( X \). Your answer should be in terms only of \( X \) and the constant \( a \). Hint: you do not need to evaluate complicated integrals here. Just use integration by parts, i.e., \( \int_\alpha^\beta udv = uv\bigg|_\alpha^\beta - \int_\alpha^\beta vdu \). The final result will look very nice!

Solution

In order to find \( \mathbb{E}(\Lambda|X) \), we need to find \( f_{\Lambda|X}(\lambda|x) \). We apply Bayes rule:

\[
f_{\Lambda|X}(\lambda|x) = \frac{p_{X|\Lambda}(x|\lambda) \mathbb{E}_\Lambda(1)}{\int_0^{\infty} p_{X|\Lambda}(x|\gamma) f_\Lambda(\gamma) d\gamma} f_\Lambda(\lambda) \\
= \frac{\lambda^x e^{-\lambda}}{\int_0^{\infty} \frac{\gamma^x e^{-\gamma}}{\lambda^x a^\gamma} d\gamma} a e^{-a\lambda} \\
= \frac{\lambda^x e^{-\lambda(1+a)}}{\int_0^{\infty} \gamma^x e^{-\gamma(1+a)} d\gamma}.
\]

Instead of trying to evaluate the integral in the denominator, let’s go ahead and find the
conditional expectation:

\[
E(\Lambda \mid X = x) = \int_0^\infty \lambda x e^{-\lambda(1+a)} \left( \frac{\int_0^\infty \gamma^x e^{-\gamma(1+a)} d\gamma}{\int_0^\infty \gamma^xe^{-\gamma(1+a)} d\gamma} \right) d\lambda
\]

\[
= \int_0^\infty \frac{\lambda x e^{-\lambda(1+a)}}{\int_0^\infty \gamma^xe^{-\gamma(1+a)} d\gamma} d\lambda
\]

\[
= \frac{x+1}{\int_0^\infty \gamma^xe^{-\gamma(1+a)} d\gamma} \left( \frac{x+1}{1+a} \right) e^{-(1+a)x} d\gamma
\]

\[
= \frac{x+1}{1+a}.
\]