1. Gaussian Hypothesis Testing

Consider a hypothesis testing problem that if $X = 0$, you observe a sample of $\mathcal{N}(\mu_0, \sigma^2)$, and if $X = 1$, you observe a sample of $\mathcal{N}(\mu_1, \sigma^2)$, where $\mu_0, \mu_1 \in \mathbb{R}$, $\sigma^2 > 0$. Find the Neyman-Pearson test for false alarm $\beta \in (0, 1)$, that is, $\mathbb{P}(\hat{X} = 1 \mid X = 0) \leq \beta$.

Solution:

Let $y$ be the observation. We know that

$$f_i(y) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(y-\mu_i)^2/(2\sigma^2)}, \quad i = 0, 1.$$ 

Thus, the likelihood ratio is

$$L(y) = \frac{f_1(y)}{f_0(y)} = e^{-(y-\mu_1)^2-(y-\mu_0)^2)/(2\sigma^2)} = e^{(\mu_0^2-\mu_1^2+2y(\mu_1-\mu_0))/(2\sigma^2)}.$$ 

Note that when $\mu_1 > \mu_0$, the likelihood ratio $L(y)$ is monotonically increasing in $y$ while when $\mu_1 < \mu_0$, $L(y)$ is monotonically decreasing in $y$.

The Neyman-Pearson decision rule for the continuous case states that the optimal decision rule is determined by

$$\hat{X}(y) = \begin{cases} 0, & L(y) < \lambda, \\ 1, & L(y) \geq \lambda. \end{cases}$$

When the likelihood is monotonic, the decision rule can be written with respect to $y$. When $L(y)$ is monotonically increasing (in this problem, when $\mu_1 > \mu_0$), the decision rule follows

$$\hat{X}(y) = \begin{cases} 0, & y < t, \\ 1, & y \geq t. \end{cases}$$

The signs are switched when $L(y)$ is monotonically decreasing (in this problem, when $\mu_1 < \mu_0$):

$$\hat{X}(y) = \begin{cases} 0, & y > t, \\ 1, & y \leq t. \end{cases}$$
The relationship between $\lambda$ and $t$ can be found as follows.

$$L(y) > \lambda$$

and taking the logarithm of both sides we have

$$y > \frac{\sigma^2}{\mu_1 - \mu_0} \ln \lambda + \frac{\mu_1 + \mu_0}{2} = t.$$ 

when $\mu_1 > \mu_0$. The inequality is the other direction when $\mu_1 < \mu_0$ because we divide by a negative number. This algebra shows why the above decision rules that were written with respect to $y$ are identical to the Neyman-Pearson decision rule written using the likelihood for monotonically increasing and decreasing likelihood functions.

We define the left hand side of the above equation as some threshold $t$. Now we want to find $t$ such that the false alarm is $\beta$ for the two cases.

When $\mu_1 > \mu_0$,

$$\mathbb{P}(\hat{X} = 1 \mid X = 0) = \int_t^\infty f_0(y) \, dy = 1 - \Phi\left(\frac{t - \mu_0}{\sigma}\right) = \beta$$

Thus, $t = \sigma \Phi^{-1}(1 - \beta) + \mu_0$, where $\Phi$ is the CDF of the Gaussian distribution.

When $\mu_1 < \mu_0$, we can follow a similar procedure to find that $\Phi\left(\frac{t - \mu_0}{\sigma}\right) = \beta$, so $t = \sigma \Phi^{-1}(\beta) + \mu_0$.

### 2. BSC Hypothesis Testing

Consider a BSC with some error probability $\epsilon \in [0.1, 0.5)$. Given $n$ inputs and outputs $(x_i, y_i)$ of the BSC, solve a hypothesis problem to detect that $\epsilon > 0.1$ with a probability of false alarm at most equal to 0.05. Assume that $n$ is very large and use the CLT.

**Hint**: The null hypothesis is $\epsilon = 0.1$. The alternate hypothesis is $\epsilon > 0.1$, which is a composite hypothesis (this means that under the alternate hypothesis, the probability distribution of the observation is not completely determined; compare this to a simple hypothesis such as $\epsilon = 0.3$, which does completely determine the probability distribution of the observation). The Neyman-Pearson Lemma we learned in class applies for the case of a simple null hypothesis and a simple alternate hypothesis, so it does not directly apply here.

To fix this, fix some specific $\epsilon' > 0.1$ and use the Neyman-Pearson Lemma to find the optimal hypothesis test for the hypotheses $\epsilon = 0.1$ vs. $\epsilon = \epsilon'$. Then, argue that the optimal decision rule does not depend on the specific choice of $\epsilon'$; thus, the decision rule you derive will be simultaneously optimal for testing $\epsilon = 0.1$ vs. $\epsilon = \epsilon'$ for all $\epsilon' > 0.1$.

**Solution**:

We observe $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$. Let $x$ and $y$ be the vectors of these observations. The likelihood is

$$P(X = x, Y = y \mid \epsilon) = P(X = x \mid \epsilon) \cdot P(Y = y \mid X = x, \epsilon)$$

We can ignore $P(X = x \mid \epsilon)$ since in the final likelihood ratio, it’ll cancel out in the numerator and denominator.

$$\mathbb{P}(Y = y \mid X = x, \epsilon) = \epsilon^{\sum_{i=1}^n 1\{y_i \neq x_i\}} (1 - \epsilon)^{\sum_{i=1}^n 1\{y_i = x_i\}} \propto \left(\frac{\epsilon}{1 - \epsilon}\right)^{\sum_{i=1}^n 1\{y_i \neq x_i\}}.$$
What matters for estimating $\epsilon$ is $t := \sum_{i=1}^{n} 1\{x_i \neq y_i\}$. Therefore we can rewrite the likelihoods as follows. Under the null hypothesis, the likelihood is

$$P(Y = y \mid X = x, \epsilon = 0.1) \propto \left(\frac{0.1}{0.9}\right)^t = \left(\frac{1}{9}\right)^t.$$ 

Fix some $\epsilon' > 0.1$; under the alternate hypothesis $\epsilon = \epsilon'$, then the likelihood is

$$P(Y = y \mid X = x, \epsilon = \epsilon') \propto \left(\frac{\epsilon'}{1-\epsilon'}\right)^t.$$ 

The likelihood ratio is therefore

$$L(t) = \left(\frac{9\epsilon'}{1-\epsilon'}\right)^t.$$ 

The likelihood ratio is a non-decreasing function of $T := \sum_{i=1}^{n} 1\{X_i \neq Y_i\}$, so a threshold on the likelihood ratio corresponds to a threshold on $T$. According to the Neyman-Pearson Lemma, the optimal decision rule is to declare the alternate hypothesis when

$$T > \lambda(\epsilon')$$

where

$$\lambda(\epsilon') = \frac{\epsilon' - 0.1}{\sqrt{0.09n}}.$$

Thus, $\lambda(\epsilon') = 0.1n + 1.65\sqrt{0.09n}$. This does not depend on the choice of $\epsilon'$, so the decision rule is the same for all $\epsilon' > 0.1$ and we are done.

3. Hypothesis Testing for Uniform Distribution

Assume that

- If $X = 0$, then $Y \sim \text{Uniform}[-1, 1]$.
- If $X = 1$, then $Y \sim \text{Uniform}[0, 2]$.

Using the Neyman-Pearson formulation of hypothesis testing, find the optimal randomized decision rule $r : [-1, 2] \to \{0, 1\}$ with respect to the criterion

$$\min_{\text{randomized } r : [-1, 2] \to \{0, 1\}} \mathbb{P}(r(Y) = 0 \mid X = 1)$$

s.t. $\mathbb{P}(r(Y) = 1 \mid X = 0) \leq \beta$,

where $\beta \in [0, 1]$ is a given upper bound on the false positive probability.
Solution:

Here, the likelihood ratio is

\[ \frac{f_{Y|X}(y|1)}{f_{Y|X}(y|0)} = \begin{cases} 1 & 0 \leq y \leq 2 \\ \frac{1}{-1 \leq y \leq 1} & \end{cases} \]

Thus, \( \hat{X} = 1 \) if \( Y > 1 \) and \( \hat{X} = 0 \) if \( Y < 0 \). If \( Y \in [0,1] \) we need randomization, so \( \hat{X} = 1 \) with some probability \( \gamma \). We choose \( \gamma \) such that

\[ P(\hat{X} = 1 | X = 0) = \beta. \]

That is,

\[ \gamma P(Y \in [0,1] | X = 0) = \frac{\gamma}{2} = \beta. \]

Thus, \( \gamma = 2\beta \).

4. Photodetector LLSE

Consider a photodetector in an optical communications system that counts the number of photons arriving during a certain interval. A user conveys information by switching a photon transmitter on or off. Assume that the probability of the transmitter being on is \( p \). If the transmitter is on, the number of photons transmitted over the interval of interest is a Poisson random variable \( \Theta \) with mean \( \lambda \), and if it is off, the number of photons transmitted is 0. Unfortunately, regardless of whether or not the transmitter is on or off, photons may be detected due to “shot noise”. The number \( N \) of detected shot noise photons is a Poisson random variable \( N \) with mean \( \mu \), independent of the transmitted photons. Let \( T \) be the number of transmitted photons and \( D \) be the number of detected photons. Find \( L[T | D] \).

Solution:

\[ L[T | D] = \mathbb{E}[T] + \frac{\text{cov}(T,D)}{\text{var} D} (D - \mathbb{E}[D]) \]

We find each of these terms separately. We can see by the law of total expectation that \( \mathbb{E}[T] = p\lambda \). Now, we have:

\[ \text{cov}(T,D) = \mathbb{E}[(T - \mathbb{E}[T])(D - \mathbb{E}[D])] = \mathbb{E}[(T - \mathbb{E}[T])(T - \mathbb{E}[T] + N - \mathbb{E}[N])] = \mathbb{E}[(T - \mathbb{E}[T])^2] + \mathbb{E}[(T - \mathbb{E}[T])(N - \mathbb{E}[N])] = \text{var} T = p(\lambda^2 + \lambda) - (p\lambda)^2 \]

where the second to last equality follows since \( T \) and \( N \) are independent. We now find:

\[ \text{var} D = \text{var}(T + N) = \text{var} T + \text{var} N = p(\lambda^2 + \lambda) - (p\lambda)^2 + \mu \]

Finally, we have \( \mathbb{E}[D] = \mathbb{E}[T] + \mathbb{E}[N] = p\lambda + \mu \). Putting these together, we have the LLSE (no need to simplify the equation).

5. Gaussian LLSE

The random variables \( X, Y, Z \) are i.i.d. \( \mathcal{N}(0,1) \).
(a) Find $L[X^2 + Y^2 \mid X + Y]$.
(b) Find $L[X + 2Y \mid X + 3Y + 4Z]$.
(c) Find $L[(X + Y)^2 \mid X - Y]$.

Solution:

(a) We note that
$$E[(X^2 + Y^2)(X + Y)] = E[X^3 + X^2Y + XY^2 + Y^3] = 0.$$ Hence,
$$\text{cov}(X^2 + Y^2, X + Y) = 0;$$ so that
$$L[X^2 + Y^2 \mid X + Y] = E[X^2 + Y^2] = 2.$$

(b) We find
$$\text{cov}(X + 2Y, X + 3Y + 4Z) = E[(X + 2Y)(X + 3Y + 4Z)] = 1 + 6 = 7$$
and
$$\text{var}(X + 3Y + 4Z) = 1 + 9 + 16 = 26.$$ Hence,
$$L[X + 2Y \mid X + 3Y + 4Z] = \frac{7}{26}(X + 3Y + 4Z).$$

(c) We observe that $\text{cov}(X + Y, X - Y) = 0$, so that $X + Y$ and $X - Y$ are independent. Hence,
$$L[(X + Y)^2 \mid X - Y] = E[(X + Y)^2] = 2.$$

6. Projections

*The following exercises are from the note on the Hilbert space of random variables. See the notes for some hints.*

(a) Let $\mathcal{H} := \{X : X$ is a real-valued random variable with $E[X^2] < \infty\}$. Prove that $\langle X, Y \rangle := E[XY]$ makes $\mathcal{H}$ into a real inner product space. \(^1\)

(b) Let $U$ be a subspace of a real inner product space $V$ and let $P$ be the projection map onto $U$. Prove that $P$ is a linear transformation.

(c) Suppose that $U$ is finite-dimensional, $n := \dim U$, with basis $\{v_i\}_{i=1}^n$. Suppose that the basis is orthonormal. Show that $Py = \sum_{i=1}^n \langle y, v_i \rangle v_i$. (Note: If we take $U = \mathbb{R}^n$ with the standard inner product, then $P$ can be represented as a matrix in the form $P = \sum_{i=1}^n v_i v_i^\top$.)

Solution:

\(^1\)To be perfectly correct, it is possible for $X \neq 0$ but $E[X^2] = 0$; this occurs if $X = 0$ with probability 1. To fix this, we need to define two random variables $X$ and $Y$ to be *equal* if $P(X = Y) = 1$. In other words, we consider *equivalence classes* of random variables, defined by the relation $X \equiv Y$. With this definition, then if $X \neq 0$ we do indeed have $E[X^2] > 0$. 


(a) To satisfy the definition of a vector space, we must check that $\mathcal{H}$ is closed under vector addition and scalar multiplication, that is, if $X, Y \in \mathcal{H}$ and $c \in \mathbb{R}$, then $X + Y \in \mathcal{H}$ and $cX \in \mathcal{H}$. Since

$$\mathbb{E}[(X + Y)^2] = \mathbb{E}[X^2] + 2\mathbb{E}[XY] + \mathbb{E}[Y^2]$$

and $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$, we must show that $\mathbb{E}[XY] < \infty$, but by the Cauchy-Schwarz Inequality, $\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2] \mathbb{E}[Y^2]} < \infty$. Also,

$$\mathbb{E}[(cX)^2] = c^2 \mathbb{E}[X^2] < \infty.$$

We must also check the following properties of a vector space: for all $X,Y,Z \in \mathcal{H}$ and $c,d \in \mathbb{R}$,

- $X + (Y + Z) = (X + Y) + Z$;
- $X + Y = Y + X$;
- $0 + X = X$;
- $1 \cdot X = X$;
- $c(dX) = (cd)X$;
- $c(X + Y) = cX + cY$;
- $(c + d)X = cX + dX$.

All of the above are familiar properties of functions, so we do not bother to check them. Finally, we have to check the properties of the inner product:

- $\mathbb{E}[XY] = \mathbb{E}[YX]$;
- $\mathbb{E}[(X + cY)Z] = \mathbb{E}[XZ] + c\mathbb{E}[XZ]$;
- $\mathbb{E}[X^2] > 0$ if $X \neq 0$.

The first property is clear; the second property is linearity of expectation. Perhaps surprisingly, the third property is technically challenging, but do not worry about the details.

(b) Let $u, v \in V$ and $c \in \mathbb{R}$. Then, we claim that $P(u + cv) = Pu + cPv$. It suffices to check that $Pu + cPv \in U$ and $u + cv - Pu - cPv \in U^\perp$. Since $Pu \in U$ and $Pv \in U$, then $Pu + cPv \in U$ since $U$ is a subspace. Also, for any $w \in U$, we get

$$\langle w, u + cv - Pu - cPv \rangle = \langle w, u - Pu \rangle + c\langle w, v - Pv \rangle = 0$$

since $u - Pu \in U^\perp$ and $v - Pv \in U^\perp$. Thus, $u + cv - Pu - cPv \in U^\perp$. We therefore have $P(u + cv) = Pu + cPv$ and $P$ is linear.

(c) Since each $v_i \in U$, then $\sum_{i=1}^n \langle y, v_i \rangle v_i \in U$ since $U$ is a subspace. Also, for any $v_j$, we get

$$\langle v_j, y - \sum_{i=1}^n \langle y, v_i \rangle v_i \rangle = \langle v_j, y \rangle - \sum_{i=1}^n \langle y, v_i \rangle \langle v_j, v_i \rangle = \langle v_j, y \rangle - \langle y, v_j \rangle = 0,$$

where we have used the fact that the basis is orthonormal (so $\langle v_i, v_j \rangle = 0$ if $i \neq j$ and $\langle v_j, v_j \rangle = 1$). Since we have $\langle v_j, y - \sum_{i=1}^n \langle y, v_i \rangle v_i \rangle$ for any $v_j$ and $v_1, \ldots, v_n$ is a basis for $U$, then we get $\langle u, y - \sum_{i=1}^n \langle y, v_i \rangle v_i \rangle = 0$ for any $u \in U$, i.e., $y - \sum_{i=1}^n \langle y, v_i \rangle v_i \in U^\perp$. Hence, $Py = \sum_{i=1}^n \langle y, v_i \rangle v_i$. 

6