UC Berkeley
Department of Electrical Engineering and Computer Sciences

EECS 126: PROBABILITY AND RANDOM PROCESSES

Problem Set 12
Fall 2021

1. **Flipping Coins and Hypothesizing**

You flip a coin until you see heads. Let

$$X = \begin{cases} 1 & \text{if the bias of the coin is } q > p. \\ 0 & \text{if the bias of the coin is } p. \end{cases}$$

Find a decision rule $\hat{X}(Y)$ that maximizes $P[\hat{X} = 1 \mid X = 1]$ subject to $P[\hat{X} = 1 \mid X = 0] \leq \beta$ for $\beta \in [0,1]$. Remember to calculate the randomization constant γ.

Solution:

Let Y be the number of flips until we see heads. Write the likelihood ratio.

$$L(y) = \frac{P[Y = y \mid X = 1]}{P[Y = y \mid X = 0]} = \frac{(1-q)^{y-1}q}{(1-p)^{y-1}p},$$

which is strictly decreasing in y since $q > p$. Hence, the hypothesis testing rule is of the form $\hat{X} = 1$ if $Y < \alpha$ for some α. Observe that

$$P[Y < \alpha \mid X = 0] = \sum_{y=1}^{\alpha-1} p(1-p)^{y-1} = 1 - (1-p)^{\alpha-1}.$$

Therefore, we should choose α such that $1 - (1-p)^{\alpha-1} \leq \beta$, i.e.

$$\alpha \leq 1 + \frac{\log(1-\beta)}{\log(1-p)}.$$

Therefore, take $\alpha = \lfloor 1 + \log(1-\beta)/\log(1-p) \rfloor$. For the randomization, let $P[\hat{X} = 1 \mid Y = \alpha] = \gamma$. The probability of false detection is

$$\begin{aligned} P[\hat{X} = 1 \mid X = 0] &= P[Y < \alpha \mid X = 0] + \gamma P[Y = \alpha \mid X = 0] \\ &= 1 - (1-p)^{\alpha-1} + \gamma p(1-p)^{\alpha-1} \leq \beta, \end{aligned}$$

so we take

$$\gamma = \frac{\beta - 1 + (1-p)^{\alpha-1}}{p(1-p)^{\alpha-1}}.$$

Hence, for the values of α and γ described above,

$$\hat{X} = \begin{cases} 1, & Y < \alpha, \\ Z, & Y = \alpha, \\ 0, & Y > \alpha, \end{cases}$$

where Z is 1 with probability γ and 0 otherwise.

2. **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 $\alpha \in (0, 1)$, that is, $P(\hat{X} = 1 \mid X = 0) \le \alpha$.

Solution:

Let y be the observation. We know that

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

Thus, the likelihood ratio is

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

Without loss of generality suppose that $\mu_1 > \mu_0$. Then, solving

$$\frac{f_1(y)}{f_0(y)} > \lambda$$

and taking the logarithm of both sides we have

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

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 α.

$$P(\hat{X} = 1 \mid X = 0) = \int_t^\infty f_0(y)\,\mathrm{d}y = Q\Big(\frac{t - \mu_0}{\sigma}\Big) = \alpha$$

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

3. **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 ϵ'; 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 $x_1, \dots, x_n$ and $y_1, \dots, y_n$. 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.

$$P(Y = y \mid X = x, \epsilon) = \epsilon^{\sum_{i=1}^n \mathbf{1}\{y_i \neq x_i\}}(1-\epsilon)^{\sum_{i=1}^n \mathbf{1}\{y_i = x_i\}} \propto \left(\frac{\epsilon}{1-\epsilon}\right)^{\sum_{i=1}^n \mathbf{1}\{y_i \neq x_i\}}.$$

What matters for estimating ϵ is $t := \sum_{i=1}^n \mathbf{1}\{x_i \neq y_i\}$. 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 $T := \sum_{i=1}^n \mathbf{1}\{X_i \neq Y_i\}$, 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 $\lambda(\epsilon')$ is the threshold that is determined by setting the PFA exactly equal to the constraint, i.e.,

$$\mathrm{PFA} = P\big(T > \lambda(\epsilon') \bigm| \epsilon = 0.1\big) = 0.05.$$

Since n is very large, $T = \sum_{i=1}^n \mathbf{1}\{X_i \neq Y_i\}$ is approximately a normal random variable. Note that without the approximation T is a binomial since input-output pairs of the channel are independent. Now we calculate the following.

$$\begin{aligned} P(X_1 \neq Y_1) &= P(Y_1 = 1 \mid X_1 = 0)P(X_1 = 0) + P(Y_1 = 0 \mid X_1 = 1)P(X_1 = 1) \\ &= \epsilon \end{aligned}$$

Thus, $T \sim \mathcal{N}(n\epsilon, n\epsilon(1-\epsilon))$ and

$$P\big(\mathcal{N}(0.1n, 0.09n) > \lambda(\epsilon')\big) = Q\Big(\frac{\lambda(\epsilon') - 0.1n}{\sqrt{0.09n}}\Big) = Q(1.67) = 0.05.$$

Thus, $\lambda(\epsilon') = 0.1n + 1.67\sqrt{0.09n}$. This does not depend on the choice of ϵ', so the decision rule is the same for all $\epsilon' > 0.1$ and we are done.

4. **Basic Properties of Jointly Gaussian Random Variables**

Let $(X_1, \dots, X_n)$ be a collection of jointly Gaussian random variables. Their joint density is given by (for $x \in \mathbb{R}^n$)

$$f(x) = \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\left(-\frac{1}{2}(x-\mu)^T C^{-1}(x-\mu)\right),$$

where μ is the mean vector and C is the covariance matrix.

(a) Show that $X_1, \ldots, X_n$ are independent if and only if they are pairwise uncorrelated.

(b) Show that any linear combination of these random variables will also be a Gaussian random variable.

Solution:

(a) WLOG suppose our variables are mean-centered (the proof follows through the same, just with more notation). It suffices to show that the densities decompose into a product of individual densities. Note that if they are pairwise uncorrelated, then the matrix C is diagonal, with $(\sigma_{X_i}^2, 1 \leq i \leq n)$ along the diagonal. Then we can write

$$\begin{aligned} f(x) &= \frac{1}{\sqrt{(2\pi)^n \det(C)}} \exp\left(-\frac{1}{2} x^T C^{-1} x\right) \\ &= \frac{1}{\sqrt{(2\pi)^n \prod_{i=1}^n \sigma_{X_i}^2}} \exp\left(-\frac{1}{2} \sum_{i=1}^n \sigma_{X_i}^{-2} x_i^2\right) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_{X_i}^2}} \exp\left(-\frac{1}{2} x_i^2 / \sigma_{X_i}^2\right), \end{aligned}$$

which we see is the product of densities of X_i. Hence pairwise uncorrelated implies independence.

If $X_1, \ldots, X_n$ are independent, then, $\mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)] = \mathbb{E}(X_i - \mu_i)\mathbb{E}(X_j - \mu_j) = 0$. Hence, by definition, they are uncorrelated.

(b) Consider, $Y = \sum_{i=1}^n a_i X_i$ and define $a = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix}^\mathsf{T}$. One needs to show that Y is Gaussian. The characteristic function of Y is

$$\begin{aligned} \phi_Y(u) &= \mathbb{E}\Big[\exp\Big(\mathrm{i}u \sum_{i=1}^n a_i X_i\Big)\Big] = \phi_X(ua_1, \ldots, ua_n) \\ &= \exp\Big(\mathrm{i}u\langle a, \mu\rangle - \frac{1}{2} u^2 a^\mathsf{T} C a\Big). \end{aligned}$$

Therefore, $Y \sim \mathcal{N}(\langle a, \mu\rangle, a^\mathsf{T} C a)$.

5. Independent Gaussians

Let $X = (X, Y)$ be a jointly Gaussian random vector with mean vector $[0, 0]$ and covariance matrix

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

Find a 2×2 matrix U such that $UX = (X', Y')$ where X' and Y' are independent.

Solution: If we were to sketch the contours of X, they would be ovals that are stretched out along the line $y = x$. In particular, we can see that a rotation by $45°$ will do the trick, i.e.

$$U = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

It's easy to check that this works, as

$$\begin{aligned}\mathrm{Cov}(X', Y') &= \mathrm{Cov}\left(\frac{\sqrt{2}}{2}X + \frac{\sqrt{2}}{2}Y, -\frac{\sqrt{2}}{2}X + \frac{\sqrt{2}}{2}Y\right) \\ &= \mathrm{Var}(\frac{\sqrt{2}}{2}X) - \mathrm{Var}(\frac{\sqrt{2}}{2}Y) = 0.\end{aligned}$$

Hence X' and Y' are independent.