ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
School of Computer and Communication Sciences

Handout 11 Principles of Digital Communications
Graded Homework 1 March 24, 2021

PROBLEM 1. Suppose $(a_i : i = 1, \dots, n)$ and $(b_i : i = 1, ..., n)$ are non-negative real numbers. Let $M = \sum_i \max(a_i, b_i)$, $m = \sum_i \min(a_i, b_i)$.

a. Show that

$$\sum_i \sqrt{a_i b_i} \le \sqrt{Mm}.$$

[Hint: use Cauchy-Schwartz inequality.]

One can write $a_i b_i = \min(a_i, b_i) \max(a_i, b_i)$ *and then apply Cauchy Schwartz inequality as*

$$\begin{aligned}\sum_i \sqrt{a_i b_i} &= \sum_i \sqrt{\min(a_i, b_i) \max(a_i, b_i)} \\ &\le \sqrt{\sum_i \sqrt{\min(a_i, b_i)}^2 \sum_j \sqrt{\max(a_j, b_j)}^2} \\ &= \sqrt{mM}\end{aligned}$$

Recall that in a binary hypothesis testing problem with $H \in \{0, 1\}$ and observation Y, the error probability of the optimal rule is given by

$$P(\text{error}) = \sum_y \min(P_H(0)P_{Y|H}(y|0), P_H(1)P_{Y|H}(y|1)).$$

b. Show that, with $Z = \sqrt{P_H(0)P_H(1)} \sum_y \sqrt{P_{Y|H}(y|0)P_{Y|H}(y|1)}$,

$$Z^2 \le P(\text{error}).$$

[Hint: use (a), and upper bound $\max(a_i, b_i)$ by $a_i + b_i$.]

A direct derivation using the hint gives

$$\begin{aligned}Z^2 &= \left(\sum_y \sqrt{P_H(0)P_{Y|H}(y|0) \cdot P_H(1)P_{Y|H}(y|1)}\right)^2 \\ &\le \sum_y \min(P_H(0)P_{Y|H}(y|0), P_H(1)P_{Y|H}(y|1)) \sum_{y'} \max(P_H(0)P_{Y|H}(y'|0), P_H(1)P_{Y|H}(y'|1)) \\ &\le P(\text{error}) \sum_{y'} (P_H(0)P_{Y|H}(y'|0) + P_H(1)P_{Y|H}(y'|1)) \\ &= P(\text{error})\end{aligned}$$

c. Suppose $P_H(0) = P_H(1) = 1/2$, $Y \in \{0,1\}$ with $P_{Y|H}(1|0) = P_{Y|H}(0|1) = \alpha$, $\alpha \in [0, 1/2]$. Evaluate Z and $P(\text{error})$ as function of α.

We can compute

$$\begin{aligned} P(\text{error}) &= \sum_y \min(P_H(0)P_{Y|H}(y|0), P_H(1)P_{Y|H}(y|1)) \\ &= \frac{1}{2}\sum_y \min(\alpha, 1-\alpha) \\ &= \alpha \end{aligned}$$

and

$$\begin{aligned} Z &= \sqrt{P_H(0)P_H(1)}\sum_y \sqrt{P_{Y|H}(y|0)P_{Y|H}(y|1)} \\ &= \frac{1}{2}\sum_y \sqrt{\alpha(1-\alpha)} \\ &= \sqrt{\alpha(1-\alpha)} \end{aligned}$$

d. Continuing with (c), sketch $Z^2/P(\text{error})$ as a function of α. Show that no bound of the form '$\lambda Z^2 \leq P(\text{error})$' with a constant $\lambda > 1$ can hold in general.

A small computation shows that $Z^2/P(\text{error}) = (1-\alpha)$, *Figure 1 shows a sketch of this* It is not possible to say that the bound doesn't hold at $\alpha = 0$, indeed in that case both terms of the bound are 0. Let $\lambda > 1$ and $0 < \alpha < 1 - \frac{1}{\lambda}$, then $\lambda Z^2 = \lambda\alpha(1-\alpha) > \alpha = P(\text{error})$.

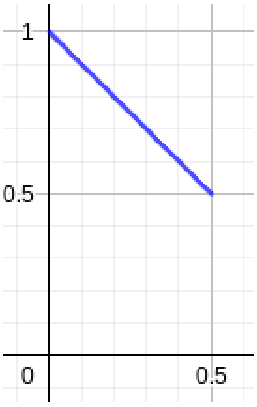

Figure 1: $Z^2/P($error as a function of $\alpha \in [0, 1/2]$

Problem 2. Consider a binary hypothesis testing problem on a biased coin. Let $H = 0$ corresponds to the hypothesis that the coin is biased by $1/2 - \delta$ (i.e., $P(Head \mid H = 0) = 1/2 - \delta$), and $H = 1$ corresponds to the hypothesis that the coin is biased by $1/2 + \delta$, where $0 < \delta \leq 1/2$. Assume that both hypotheses have the same prior.

a. Let $X_1, X_2, ...$ be the outcomes of successive tosses of the coin. Under the optimal decoding rule, compute $P(H \neq \hat{H} | X_1, X_2, \ldots, X_n)$, where $\hat{H}$ is the decision.

Let $\gamma = (1/2 - \delta)/(1/2 + \delta)$, notice that $\gamma \leq 1$. We have

$$\frac{P(H = 0 \mid X_1, \ldots, X_n)}{P(H = 1 | X_1, \ldots, X_n)} = \gamma^{2n_0 - n}$$

where n_0 is the number of heads in $X_1, \ldots, X_n$. This implies that,

$$P(H = 0 | X_1, \ldots, X_n) = \frac{1}{1 + \gamma^{-(2n_0 - n)}} \qquad P(H = 1 | X_1, \ldots, X_n) = \frac{1}{1 + \gamma^{2n_0 - n}}.$$

Under the optimal decoding rule, we always take $\hat{H} = \mathrm{argmax}_i\, P(H = i | X_1, \ldots, X_n)$. Therefore, we have

$$P(H \neq \hat{H} | X_1, \ldots, X_n) = \min \left\{ \frac{1}{1 + \gamma^{-(2n_0 - n)}}, \frac{1}{1 + \gamma^{2n_0 - n}} \right\} = \frac{1}{1 + \gamma^{-|2n_0 - n|}}.$$

We can see that the reliability of tester's decision actually varies depending on the observation. Now consider a different scheme, namely the tester will flip the coin until it observes that |(number of head) – (number of tail)| $= L$. Let us denote the time when it happens as T.

b. Under the optimal decision rule, show that using this new scheme, $P(H \neq \hat{H} \mid T = t, X_1 = x_1, \ldots, X_t = x_t)$ does not depend on the value of t and $x_1, \ldots, x_t$.

First of all, we observe that $P(T|H = 0) = P(T|H = 1)$ by symmetry, i.e., $P(Head|H = 0) = P(Tail|H = 1)$. We have

$$\frac{P(H = 0 | T = t, X_1 = x_1, \ldots, X_T = x_t)}{P(H = 1 | T = t, X_1 = x_1, \ldots, X_T = x_t)} = \gamma^{2n_0 - t}.$$

where now n_0 is the number of head in $x_1, \ldots, x_t$. By the definition of T, we know that if $T = t$ then $X_1 x_1, \ldots, X_t = x_t$ then $|2n_0 - t| = L$. Therefore with similar argument to (a.), we have

$$P(H \neq \hat{H} | T = t, X_1 = x_1, \ldots, X_t = x_t) = \frac{1}{1 + \gamma^{-L}}.$$

Notice that this error probability does not depend on $T = t$ and $X_1 = x_1, \ldots, X_t = x_t$.

Suppose $Z_1, Z_2, ..$ are i.i.d., with $P(Z_1 = -1) = 1/2 - \delta, P(Z_1 = +1) = 1/2 + \delta$. For $n = 0, 1, ...$ let $S_n = Z_1 + ... + Z_n$; note that $S_0 = 0$. Let N be the smallest value of n such that $S_n = 1$.

c. Note that $\mathbb{E}[N \mid Z_1 = 1] = 1$. What is the relationship between $\mathbb{E}[N \mid Z_1 = -1]$ and $E[N]$?

Let $N_{a\to b}$ to be the first n such that $S_n = b$ if $S_0 = a$. We have $E[N|Z_1 = -1] = 1 + N_{-1\to 0} + N_{0\to 1}$. However, we can observe that $N_{-1\to 0} = N_{0\to 1}$, and $N_{0\to 1} = N$ by definition. Hence $E[N|Z_1 = -1] = 1 + 2E[N]$.

d. Show that $\mathbb{E}[N] = 1/2\delta$.

We have,

$$\begin{aligned} E[N] &= E[N|Z_1 = 1]P(Z_1 = 1) + E[N|Z_1 = -1]P(Z_1 = -1) \\ &= \left(\frac{1}{2} + \delta\right) + \left(\frac{1}{2} - \delta\right)(1 + 2E[N]) \\ &= 1 + E[N] - 2\delta E[N]. \end{aligned}$$

Solving for $E[N]$ gives us $E[N] = 1/2\delta$.

For $\ell = 1, 2, ...$ let N_ℓ denote the smallest value of n such that $S_n = \ell$.

e. Show that $\mathbb{E}[N_\ell] = \ell/2\delta$.

[Hint: use d.]

Notice that $N_\ell = N_{0\to 1} + N_{1\to 2} + \cdots + N_{(\ell-1)\to\ell}$. So we have $\mathbb{E}[N_\ell] = \ell\mathbb{E}[n] = \ell/2\delta$.

f. Show that $\mathbb{E}[T \mid H = 0] = \mathbb{E}[T \mid H = 1] \leq L/2\delta$.

[Hint: express T in terms of N_L and N_{-L}.]

We have $\mathbb{E}[T \mid H = 0] = \mathbb{E}[T \mid H = 1]$ by symmety. Take $S_n = 2n_0 - n$, i.e., it is the running sum of the difference between the number of heads and tails that we observed so far. Notice that T can also be written as $\min\{N_L, N_{-L}\}$. Therefore we have $T \leq N_L$, taking the expectation gives us $E[T] \leq E[N_L] = L/2\delta$.

Addendum

Considering its popularity, we have to address this specific form of false argument for point (d.). The argument starts by observing a true statement that $E[S_n|N = n] = 1$. However, it continues by using a false equality, $E[S_n|N = n] = nE[Z]$. We can see that it is false simply by taking $n = 1$, so we have, $E[S_1|N = 1] = 1 \neq E[Z]$.

However, it is true that $E[NE[Z]] = 1$. We can prove it by studying a different random variable $\tilde{Z}_i = Z_i - E[Z]$ and $\tilde{S}_n = \sum_{i=1}^n \tilde{Z}_i$. Notice that this implies that $E[\tilde{S}_{n+1}|\tilde{S}_n] = \tilde{S}_n$. Think of $\tilde{S}_N$ as a payoff of a "fair gamble" if we chose to keep gambling until we observe $S_n = 1$. "Fair gamble" here means that the expectation of the payoff that you get is equal to your initial money. Intuitively, we can see that it implies $E[\tilde{S}_N] = E[\tilde{S}_0]$, i.e., you cannot get advantage on a fair gamble no matter how smart you chose to stop.

This implies that $E[\tilde{S}_N] = 0$. However we also have $E[\tilde{S}_N] = E[S_N - NE[Z]] = 0$. Which implies that $E[NE[Z]] = E[S_N] = 1$.

A more formal argument requires a technical condition whereas we have to ensure that S_N "must always" hit 1, which holds for our case. However verifying this condition requires tools beyond the scope of this course, and you can study it on the Advanced Probability class.

PROBLEM 3. Suppose $Y = [Y_1, ..., Y_n]^T$ is a real random vector with zero mean (i.e., $\mathbb{E}[Y_i] = 0$ for all i) and covariance matrix C (i.e., $C_{ij} = \mathbb{E}[Y_iY_j]$).

a. Show that for any $a = [a_1, ..., a_n]^T$, $\mathbb{E}[(\sum_i a_iY_i)^2] = a^TCa$. Conclude that C is a positive semi-definite matrix. [Recall: a matrix A is said to be positive semi-definite if for all vectors x, $x^TAx \geq 0$. A is said to be positive definite if the equality holds only if $x = 0$.]

Using linearity of expectation we get

$$\begin{aligned} 0 \leq \mathbb{E}&\left[\left(\sum_i a_iY_i\right)^2\right] \\ &= \mathbb{E}\left[\sum_i\sum_j a_iY_ia_jY_j\right] \\ &= \sum_i\sum_j a_i\mathbb{E}\left[Y_iY_j\right]a_j \\ &= \sum_i\sum_j a_iC_{ij}a_j \\ &= a^TCa \end{aligned}$$

which shows that C is positive semi definite.

b. Suppose we take a linear transform of Y to obtain a vector X, i.e., for a matrix $A = [a_{ij}]$, we let $X_i := \sum_j a_{ij}Y_j$. Show that $\mathbb{E}[X] = 0$, and the covariance matrix of X is given by ACA^T.

We again use linearity of expectation

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[AY] \\ &= A \cdot \mathbb{E}[Y] \\ &= 0 \\ \mathbb{E}[XX^T] &= \mathbb{E}[AYY^TA^T] \\ &= A\mathbb{E}[YY^T]A^T \\ &= ACA^T \end{aligned}$$

In the rest of the problem, suppose C is positive definite. Consider the following iterative procedure to define the random vector $Z = [Z_1, ..., Z_n]^T$ as a function of Y.

$$\begin{aligned} W_1 &= Y_1 & Z_1 &= \frac{W_1}{\sqrt{\mathbb{E}[W_1^2]}} \\ W_2 &= Y_2 - \mathbb{E}[Z_1Y_2]Z_1 & Z_2 &= \frac{W_2}{\sqrt{\mathbb{E}[W_2^2]}} \\ W_3 &= Y_3 - \mathbb{E}[Z_1Y_3]Z_1 - \mathbb{E}[Z_2Y_3]Z_2 & Z_3 &= \frac{W_3}{\sqrt{\mathbb{E}[W_3^2]}} \\ &\vdots & &\vdots \\ W_i &= Y_i - \sum_{j=1}^{i-1}\mathbb{E}[Z_jY_i]Z_j & Z_i &= \frac{W_i}{\sqrt{\mathbb{E}[W_i^2]}}. \end{aligned}$$

c. 1. Show that for each i, $\mathbb{E}[W_i^2] > 0$, so the division step in the definition of Z_i is always well-defined.

If we unwrap the definition of W_i, we can express it as a a linear combination of the elements Y_j for $j \leq i$, for instance $W_i = \sum_{j=1}^{i} b_j Y_j$ with $b_i = 1$. Writing $b = [b_1, \dots, b_i, 0, \dots, 0]^T$ we get that $W_i = b^T Y = Y^T b$ and $\mathbb{E}[W_i^2] = \mathbb{E}[b^T Y Y^T b] = b^T C b > 0$ since C is positive definite and $b \neq 0$.

2. Show that $\mathbb{E}[Z_1 W_2] = 0$, and thus $\mathbb{E}[Z_1 Z_2] = 0$.

Using linearity of expectation we get

$$\begin{aligned}
\mathbb{E}[Z_1 W_2] &= \mathbb{E}[Z_1(Y_2 - \mathbb{E}[Z_1 Y_2] Z_1)] \\
&= \mathbb{E}[Z_1 Y_2] - \mathbb{E}[\mathbb{E}[Z_1 Y_2] Z_1^2] \\
&= \mathbb{E}[Z_1 Y_2] - \mathbb{E}[Z_1 Y_2]\mathbb{E}[Z_1^2] \\
&= \mathbb{E}[Z_1 Y_2]\left(1 - \mathbb{E}\left[\frac{W_1^2}{\mathbb{E}[W_1^2]}\right]\right) \\
&= \mathbb{E}[Z_1 Y_2]\left(1 - \frac{\mathbb{E}[W_1^2]}{\mathbb{E}[W_1^2]}\right) \\
&= 0
\end{aligned}$$

and $\mathbb{E}[Z_1 Z_2] = \frac{1}{\sqrt{\mathbb{E}[W_2^2]}}\mathbb{E}[Z_1 W_2] = 0$.

3. Show that $\mathbb{E}[Z_j W_i] = 0$ and $\mathbb{E}[Z_j Z_i] = 0$ for all $j < i$.

We prove this by strong induction on i. The initialisation is the subquestion c.2. Suppose that and for all $j < k < i+1$ we have $\mathbb{E}[Z_j Z_k] = 0$, let $j < i+1$, then

$$\begin{aligned}
\mathbb{E}[Z_j W_{i+1}] &= \mathbb{E}\left[Z_j\left(Y_{i+1} - \sum_{k=1}^{i}\mathbb{E}[Z_k Y_{i+1}] Z_k\right)\right] \\
&= \mathbb{E}[Z_j Y_{i+1}] - \sum_{k=1}^{i}\mathbb{E}[Z_k Y_{i+1}]\mathbb{E}[Z_j Z_k] \\
&= \mathbb{E}[Z_j Y_{i+1}] - \mathbb{E}[Z_j Y_{i+1}]\mathbb{E}[Z_j Z_j] \\
&= 0
\end{aligned}$$

Which implies $\mathbb{E}[Z_j Z_{i+1}] = \frac{1}{\sqrt{1}\sqrt{\mathbb{E}[W_{i+1}^2]}}\mathbb{E}[Z_j W_{i+1}] = 0$ *and finishes the proof.*

d. What is the covariance matrix of Z ?

For any $i \neq j$, we have $\mathbb{E}[Z_i Z_i] = \frac{\mathbb{E}[W_i^2]}{\mathbb{E}[W_i^2]} = 1$ and $\mathbb{E}[Z_i Z_j] = 0$ and so the covariance matrix of Z is the identity matrix. This means that the elements of Z are uncorrelated and have unit variance.

e. The procedure to obtain Z as a function of Y is a linear transformation, let A represent that transformation, i.e., $Z = AY$. What properties does the matrix A have? (e.g., does it have some sort of triangularity? what about its diagonal entries? is it invertible or not?)

The matrix A is lower triangular, it's diagonal entries are $A_{i,i} = \frac{1}{\sqrt{\mathbb{E}[W_i^2]}} > 0$ and so A is invertible.

Problem 4. Consider a binary hypothesis problem with hypothesis $H \in \{0, 1\}$ and observation Y. Let $T = t(Y)$ be a function of the observation.

a. Suppose that for each y and a specific $P_H(0), P_H(1)$, the MAP estimator $H_{\text{MAP}}(y)$ of H from the observation y can be determined from $t(y)$. Can we conclude that T is a sufficient statistic?

   No, this was shown in Homework 4, Problem 2.

b. Suppose further that for every *a priori* distribution $P_H(0), P_H(1)$ on the hypothesis H, $H_{\text{MAP}}(y)$ can be computed from $t(y)$. Show that $p_{Y|H}(y|1)/p_{Y|H}(y|0)$ can be determined from $t(y)$.

   For every value of $p_H(0) \in [0, 1]$ we have a mapping $g_{p_H(0)} : t(y) \to H_{\text{MAP}}(y)$ such that $g_{p_H(0)}(t(y)) = 0$ if $p_{Y|H}(y|1)/p_{Y|H}(y|0) \leq p_H(0)/(1 - p_H(0))$. this means that

$$p_{Y|H}(y|1)/p_{Y|H}(y|0) = \inf_{g_p(t(y))=0} \frac{p}{1-p} = g(t(y))$$

c. Continuing with (b), can we conclude that T is a sufficient statistic?

   Let $h(y) = p_{X|H}(y|0)$, let $g_0(t) = 1$ and $g_1(t) = g(t)$ then

$$p_{Y|H}(y|i) = h(y)g_i(t(y))$$

   the conclusion follows by Fisher Neyman's factorisation theorem.

Consider now a m-ary hypotheses problem; $m \geq 2, H \in \{0, \ldots, m-1\}$.

d. Upon observing $Y = y$, suppose we are also told that H is either i or j, (with $0 \leq i < j < m$). Find the decision rule that minimizes the probability of error; call this rule $H_{opt,i,j}(y)$.

   The optimal decision can be implemented by setting $H_{opt,i,j}(y) = i$ if $p_{Y|H}(y|i)/p_{Y|H}(y|j) \geq p_H(j)/p_H(i)$ and j otherwise.

e. Suppose that for every *a priori* distribution $P_H(0), \ldots, P_H(m-1)$, and each $i < j$, $H_{opt,i,j}(y)$ can be determined from $t(y)$. Can we conclude that T is a sufficient statistic?

   We can consider only the distribution p_H that take support only on two element $i \neq j$, i.e. $p_H(k) = 0$ if $k \notin \{i, j\}$. From question b, this means we can determine $p_{Y|H}(y|i)/p_{Y|H}(y|j)$ from t for all i, j. Let $g_i(t)$ be such that $g_i(t(y)) = p_{Y|H}(y|i)/p_{Y|H}(y|0)$ for all i and let $h(y) = p_{Y|H}(y|0)$, then

$$p_{Y|H}(y|i) = h(y)g_i(t(y))$$

   and so T is a sufficient statistic.

Addendum

We see that are 3 equivalent methods of proving that a statistics $t(y)$ is a sufficient statistic. First we can show that it fulfils Neyman-Fisher factorization; Secondly, we can show that the likelihood ratio can be computed from $t(y)$; Or we can show that it can be used to compute the optimal MAP decision for every prior distribution.

It is really important to show that the MAP rule can be computed for every prior distribution, Just proving that MAP can be computed for some prior is not enough to show that it is a sufficient statistics. This point has tripped many students in problem (e.).